Институт информационных и вычислительных технологий
КН МНВО РК

Казахский Национальный Университет имени аль-Фараби

Казахский Национальный Исследовательский
Технический Университет имени К.И. Сатпаева

Университет Туран

Люблинский технический университет, Польша

МАТЕРИАЛЫ

VIII Международной научно – практической конференции

«Информатика и прикладная математика»,

посвященной памяти 85 д.т.н, профессора Бияшева Р.Г.

26-27 октября 2023 года, Алматы, Казахстан

Алматы 2023 г.

УДК 004(063)
ББК 32.973
И74

Главный редактор:
академик НАН РК, доктор физико-математических наук, профессор
М.Н. Калимолдаев

Ответственные редакторы:
главный ученый секретарь ИИВТ МНВО РК, кандидат PhD ***О.А. Усатова***

И 74 **Информатика и прикладная математика:** Мат. VIII Межд. науч. - практ. конф. (26-27 октября 2023 г). – Алматы, 2023. – 422 с.

ISBN 978-601-332-165-3

В сборнике опубликованы доклады, представленные по 4 секциям от Республики Казахстан, Российской Федерации, Латвии, Польши, Республики Беларусь, Украины, Узбекистана, Японии, Ирана, Малайзии, Кыргызской Республики и других.

Рассмотрены актуальные вопросы в области математики, информатики и управления: математического моделирования сложных систем и бизнес-процессов, исследования и разработки защищенных и интеллектуальных информационных и телекоммуникационных технологий, математической теории управления, технологий искусственного интеллекта.

Материалы сборника предназначены для научных работников, докторантов, магистрантов, а также студентов старших курсов.

УДК 004(063)
ББК 32.973

ISBN 978-601-332-165-3

© Институт информационных и вычислительных технологий МНВО РК, 2023

ПРОГРАММНЫЙ КОМИТЕТ

Почетные председатели программного комитета:

- Мутанов Г.М., академик НАН РК, д.т.н., профессор, Республика Казахстан
- Бегентаев М.М., д.э.н., профессор, Республика Казахстан

Председатель Международного Программного комитета:

- Калимолдаев М.Н., академик НАН РК, д.ф.-м.н., профессор, Республика Казахстан

Сопредседатели Международного Программного комитета:

- Айтжанова Ж. Н., к.э.н., доцент., Республика Казахстан
- Жаутиков Б.А., д.т.н., профессор, Республика
- Разакова Д.И., PhD, ассоц. профессор, Республика Казахстан
- Wójcik Waldemar, professor, Poland

Заместители Председателя Международного Программного комитета:

- Ускенбаева Р.К., д.т.н., профессор, Республика Казахстан
- Урмашев Б.А., к.ф.-м.н., профессор, Республика Казахстан
- Усатова О.А., PhD, Республика Казахстан
- Капалова Н.А., к.т.н., Республика Казахстан
- Зиятбекова Г.З., PhD, Республика Казахстан

Члены международного программного комитета:

- Abdul Razaque, PhD, USA
- Aijaz Shaikh, professor, Finland
- Andrzej Smolarz, professor, Poland
- Dietmar Beyer, PhD, Germany
- Erol Kurt, professor, Turkey
- Hazar Elman, professor, Turkey
- Farajzadeh Ali, professor, Iran.
- Keylan Alimhan, professor, Japan
- Marat Ahmet, professor, Turkey
- Mohamed Othman, professor, Malaysia
- Müslüm Arici, professor, Turkey
- Ravil Muhamediev, professor, Latvia
- Yury Kochetov, professor, Russia
- Айда-Заде К.Р., д.ф.-м.н., профессор, член-кор., НАН Азербайджана
- Барахнин В.Б., д.т.н., профессор, Россия
- Бериков В.Б., д.т.н., профессор, Россия

- Евтушенко Ю.Г., д.ф.-м.н., академик РАН, Россия
- Ельцов И.Н., д.т.н. , профессор, Россия
- Имомназаров Х.Х., д.ф.-м.н., профессор, Россия
- Крак Ю.В., д.ф.-м.н., профессор, член корр., НАН Украины
- Матерухин А.В., д.т.н., профессор, Россия
- Тузиков А.В., д.т.н., профессор, член корр., НАН, Беларусь
- Исмаилов Б.И., д.т.н., профессор, Республика Кыргызстан
- Полуянов А. Н., к.т.н., Россия
- Посыпкин М.А., к.ф.-м.н., член-корреспондент РАН, Россия
- Родионов А.С., д.т.н., профессор, Россия
- Савченко Е.Ю., д.т.н., доцент, Кыргызская Республика
- Торобеков Б. Т., д.т.н., профессор, Кыргызская Республика
- Абдибеков У.С., д.ф.-м.н., профессор, Республика Казахстан
- Амирханова Г.А., PhD, Республика Казахстан
- Ашимов А. А., академик НАН РК, д.т.н., профессор, Республика Казахстан
- Абдильдаева А.А., PhD, ассоц. профессор, Республика Казахстан
- Абдолдина Ф.Н., к.т.н., ассоц. профессор, Республика Казахстан
- Адамов А.А., д.ф.-м.н., профессор, Республика Казахстан
- Аманжолова С.Т., к.т.н., Республика Казахстан
- Амиргалиев Е.Н., д.т.н., профессор, Республика Казахстан
- Алгазы К.Т., PhD, Республика Казахстан
- Алимсеитова Ж.К., к.т.н., ассоц. профессор, Республика Казахстан
- Алипбаев К.А., PhD, Республика Казахстан
- Ахметов И.Р., PhD, Республика Казахстан
- Ахметов Б.С., д.т.н., профессор Республика Казахстан
- Ахмедиярова А.Т., PhD, ассоц. профессор, Республика Казахстан
- Байшемиров Ж.Д., PhD, ассоц. профессор, Республика Казахстан
- Балакаева Г.Т., д.ф.-м.н., Республика Казахстан
- Бегимбаева Е.Е., PhD, Республика Казахстан
- Бедельбаев А.А., к.ф.-м.н., доцент, Республика Казахстан
- Бектемесов М.А., д.ф.-м.н., профессор, Республика Казахстан
- Бектемысова Г.У.,к.т.н., ассоц. профессор, Республика Казахстан
- Бельгибаев Б.А., д.т.н., доцент, Республика Казахстан
- Дарибаев Б.С., PhD, Республика Казахстан
- Дженалиев М.Т., д.ф.-м.н., профессор, Республика Казахстан
- Джомартова Ш..А., профессор, Республика Казахстан
- Жукабаева Т.К., PhD, Республика Казахстан
- Искаков К.Т., д.ф.-м.н., профессор, Республика Казахстан
- Исламгожаев Т.У., PhD, Республика Казахстан
- Калижанова А.У., к.ф.-м.н., доцент, Республика Казахстан
- Карымсакова Н.Т., PhD, Республика Казахстан
- Козбакова А. Х., PhD, ассоц. профессор, Республика Казахстан

- Кошеков К.Т., д.т.н., профессор, член корр., НАН, Республика Казахстан
- Мазаков Т.Ж., д.ф.- м.н., профессор, Республика Казахстан
- Мамырбаев О.Ж., PhD, ассоц. профессор, Республика Казахстан
- Мансурова М.Е. к.ф.-м.н., доцент, Республика Казахстан
- Марченко М.А., д.ф.-м.н., профессор, Россия
- Маринин И.В. , Россия
- Мекебаев Н.О., PhD, Республика Казахстан
- Магзом М.М., PhD, Республика Казахстан
- Мурзабеков З.Н., д.т.н., профессор Республика Казахстан
- Мусабаев Р.Р., к.т.н., Республика Казахстан
- Мусиралиева Ш.Ж., к.ф.-м.н. доцент, Республика Казахстан
- Мустафин С.А., к.т.н., доцент, Республика Казахстан
- Муханова А.А., PhD, ассоц. профессор, Республика Казахстан
- Найзабаева Л.К., д.т.н., профессор, Республика Казахстан
- Нысанбаева С.Е. д.т.н., профессор, Республика Казахстан
- Нурахметов Б.К., д.т.н., профессор, Республика Казахстан
- Омаров Б. С., PhD, Республика Казахстан
- Пак И.Т., д.т.н., профессор, Республика Казахстан
- Пак А.А., к.т.н., профессор, Республика Казахстан
- Сакан К.С., PhD, Республика Казахстан
- Самигулина Г.А., д.т.н., профессор, Республика Казахстан
- Сатыбалдина Д. Ж., к.ф.-м. н, доцент, PhD, Республика Казахстан
- Сатыбалдиева Р.Ж., к.т.н, ассоц. профессор, Республика Казахстан
- Сатымбеков М.Н., PhD, Республика Казахстан
- Сейлова Н.А., к.т.н., ассоц. профессор, Республика Казахстан
- Сембиев О.З., д.т.н., профессор, Республика Казахстан
- Синчев Б.К. , д.т.н., профессор, Республика Казахстан
- Тасболатұлы Н., PhD, Республика Казахстан
- Ташатов Н.Н., к.ф.- м.н., Ph.D, доцент, Республика Казахстан
- Терехов А.Г., к.т.н., Республика Казахстан
- Тұрдалыұлы М., PhD, ассоц. профессор, Республика Казахстан
- Тукеев У.А., д.т.н., профессор, Республика Казахстан
- Тусупова С. А., д.т.н., профессор, Республика Казахстан
- Утегенова А.У., PhD, Республика Казахстан
- Хомпыш А., PhD, Республика Казахстан
- Цеховой А.Ф. д.т.н., профессор, Республика Казахстан
- Юничева Н.Р., к.т.н., доцент, Республика Казахстан
- Якунин К.О., PhD, ассоц. профессор, Республика Казахстан

Ученые секретари конференции:

- Аршидинова М.Т., Республика Казахстан
- Джаксылыкова А.Б., Республика Казахстан
- Тюлепбердинова Г.А., Республика Казахстан

- Шермантаева Ж.У., Республика Казахстан

Организационный комитет

Сопредседатели:

- Урмашев Б.А., к.ф-м.н., профессор, Республика Казахстан
- Усатова О.А., PhD, Республика Казахстан
- Алшимбаева Д.У., PhD, Республика Казахстан

Члены Организационного комитета:

- Аспантаев А.Б., Республика Казахстан
- Байсаринова Ж.А., Республика Казахстан
- Бапышев А.М., Республика Казахстан
- Варенников А.В., Республика Казахстан
- Дюсембаев Д., Республика Казахстан
- Ермолдина Г.Т., Республика Казахстан
- Калиева Г.С., Республика Казахстан
- Лизунов О.А., Республика Казахстан
- Меркебаев А.Г., Республика Казахстан
- Толеу А., Республика Казахстан
- Турарбек А.Т., Республика Казахстан
- Шахмаев Р.А., Республика Казахстан

Место проведения:

Казахский национальный исследовательский технический университет имени К. И. Сатпаева и Институт информационных и вычислительных технологий КН МНВО РК

Республика Казахстан, г. Алматы, улица Сатпаева 22,
тел.: +7 727 -272-37-11, +7 727 -272-39-13,
+7 727-272-33-13, +7 707-376-00-90.

Приветственное слово участникам конференции
Уважаемые участники конференции, гости, коллеги, дорогие студенты!

Рад приветствовать вас в стенах Казахского Национального исследовательского университета им. К.И. Сатпаева, который является одним из лучших технических университетов в Казахстане.

В этом году научный форум «Информатика и прикладная математика» посвящен памяти известного ученого, организатора школы криптографии Республики Казахстан, доктора технических наук, профессора Бияшева Рустема Гакашевича. 11 марта 2023 года Бияшеву Р.Г. исполнилось 85 лет, а ровно через неделю его не стало. В течение всей своей научной деятельности Рустем Гакашевич, как личность, оказывал стимулирующее влияние не только на своих учеников, но и на огромное число ученых, занимающихся проблемами информатики и информационной безопасности в нашей стране.

Также наша научная встреча является частью праздничных мероприятий, посвященных 90-летию Казахского Национального университета им. аль-Фараби. Этот праздник является важным событием не только для университета, но и для всей нашей страны, а также тысяч выпускников, проживающих в десятках странах мира.

Проведение нашей ежегодной Международной научно-практической конференции стало хорошей традицией. С каждым годом растет число участников конференции, повышается качество их выступлений и публикаций. Традиционно в работе конференций участвуют ученые и специалисты из России, США, Украины, Польши, Турции, Малайзии, Ирана и других стран. Их заинтересованное участие придает нашему мероприятию международное измерение. *Целью нашего мероприятия является обмен опытом по новейшим достижениям в области информационных и вычислительных технологий.* Наращивание потенциала отечественной науки, эффективное использование результатов исследований и ускоренное внедрение их в практику - важнейшие приоритеты государственной политики. Особенно важно, что в эту работу включены талантливые студенты и молодые ученые, призванные определять не только настоящее, но и будущее казахстанской науки, всего нашего общества. Нашими партнерами и организаторами этого мероприятия являются КазНУ им. аль-Фараби, КазНИТУ им. Сатпаева, Международный университет информационных технологий, Университет «Туран», Люблинский Технический университет.

Надеюсь, полученные результаты будут полезны всем участникам, что в первую очередь позволит развивать и совершенствовать систему подготовки специалистов для экономики Казахстана, а предложенные рекомендации действительно найдут свое применение в практической деятельности. Уверен, проводимая конференция не станет рядовым событием, пройдет в духе творчества, станет площадкой для обсуждения действительно

актуальных и важных проблем и поможет найти пути их решения. Желаю всем участникам конференции и гостям плодотворных дискуссий и новых достижений!

Я хочу пожелать всем сегодня плодотворной работы и, чтобы идеи, которые высказывались, затем претворялись в жизнь. Это чрезвычайно важно не только для ученых, которые уже состоялись, которые развиваются, но еще более важно для, студентов, магистрантов, докторантов и вообще для научной молодежи.

Выражаю благодарность всем участникам конференции и гостям, которые нашли время, чтобы принять участие в конференции. Желаю плодотворных дискуссий и новых достижений! Желаю всем вам крепкого здоровья, интересной работы и полезных деловых контактов!

Председатель Программного комитета,
академик НАН РК **М.Н. Калимолда**

СЕКЦИЯ 3

Жасанды интеллект технологиялары: Интеллектуалды басқару жүйелері. Сөйлеу технологиялары және есептеуіш лингвистика. Үлгіні тану және кескінді өңдеу. Биоинформатика және биометриялық жүйелер. Адам мен машинаның әрекеттесуі. Машиналық оқыту. Интеллектуалды робототехникалық жүйелер.

Технологии искусственного интеллекта: Интеллектуальные системы управления. Речевые технологии и компьютерная лингвистика. Распознавание образов и обработка изображений. Биоинформатика и биометрические системы. Человеко-машинное взаимодействие. Машинное обучение. Интеллектуальные робототехнические системы.

Artificial intelligence technologies: Intelligent control systems. Speech technologies and computational linguistics. Pattern recognition and image processing. Bioinformatics and biometric systems. Human-machine interaction. Machine learning. Intelligent robotic systems.

TRAINING A MACHINE LEARNING MODEL FOR FOOD PRODUCT IMAGE RECOGNITION AND MODEL PERFORMANCE EVALUATION

Nurzhanuly M., Nugmanova S.A.
e-mail: madinurzhanuly0701@gmail.com, nugm_s@mail.ru
Al-Farabi Kazakh National University, Kazakhstan

Abstract

Food adulteration or food adulteration is a process in which the quality of food is reduced by adding or substituting certain foods. Chemicals harmful to human health include not only the intentional addition or substitution of an ingredient, but also the contamination of food during preparation, storage and distribution. Eating adulterated food can lead to many diseases and premature death. Machine learning can be effectively used in the modern fight against counterfeit products. Its image processing techniques can be used as a key tool in detecting adulterated food products. Results: The classification model was obtained using machine learning; Numerical and graphical results were obtained by comparing several algorithms; The program code is written in the Python programming language for the deep learning algorithms VGG, MobileNet_v2 and InceptionV3.

Introduction

Identification of counterfeit food products is an important task to ensure the safety and quality of food products. Adulteration refers to the addition, removal or substitution of certain ingredients in food products that reduce their nutritional value and jeopardize the health of consumers. Traditional food analysis methods can be labor-intensive and expensive and are not always effective in detecting adulterated foods. On the other hand, machine learning methods can improve the accuracy and speed of identifying counterfeit food products. In this dissertation, we will explore the use of machine learning algorithms for detecting adulterated food products and discuss their advantages and limitations. Food is made from plant or animal materials that are ingested in raw, processed, or semi-processed forms to support various biochemical and physiological functions. In most cases, these are spoiled or falsified products that are harmful to the health of consumers. This includes making false or misleading claims about products for financial gain, and intentionally substituting, adding, misrepresenting or misrepresenting food products, food ingredients or food packaging. A special type of fraud occurs when real ingredients are removed, substituted, or adulterated compounds are added without the buyer's knowledge in order to provide a financial benefit to the seller. Sometimes these nutrients in most of these establishments may be prepared using quality ingredients that satisfy the palate instead of providing a complete nutritious meal. As a result, the health of these consumers deteriorates dramatically. Researchers, government and regulatory agencies are studying the impacts on food quality and safety, as well as the impact on important emerging areas in the food supply chain. Among these various new scientific fields is food adulteration. The term "adulteration" can be interpreted in different ways, including mixing or substituting substances that are superior to harmful ones, or removing some essential components from a particular food product, etc. Also included are any harmful or dangerous ingredients that may

make the food unhealthy. Examples include adding water to milk, animal carcasses to meat products not intended for human consumption, or grain products containing rocks, gravel or animal hair. Consumers often fall victim to fraud or even suffer from illnesses as a result of food adulteration, with death being the worst of them, so knowing the common types of adulterated foods and their health implications is essential for every consumer. We understand that consuming such products does not bring any benefits. We all shouldn't let this happen. Every year, April 7 is celebrated around the world as World Health Day, and according to WHO, its purpose is to raise general awareness of food fraud and encourage and inspire everyone to eat healthy and nutritious foods. In this article, we'll look at how modern machine learning can identify adulterated food. First, we need to classify food images correctly and with high accuracy. This way, when we accept a product, we know how likely it is that it will meet the standards. One of the researchers in this field, food and nutrition researcher Misgana Bunty, studied this topic and wrote a paper titled "Food Adulteration and Some Detection Methods, a Review." But in this article, he only theoretically showed methods for identifying and what kind of disease a person who consumes such products suffers from. We will demonstrate the methods and rationale for detecting adulterated food using machine learning algorithms.

Investigation and identification of counterfeit products

One of the most effective ways to combat food fraud is to use machine learning techniques. These methods can analyze large volumes of data and identify anomalies that may indicate fraud.

Analysis of the chemical composition of products. Machine learning can be used to analyze the chemical composition of a product by processing data about the content of various chemical components. For example, spectral analysis can be used to determine whether a product contains inappropriate ingredients or additives [1].

Texture and visual characteristics analysis: Modern machine learning technologies can analyze the texture and visual characteristics of products. This may include comparing the color, texture, shape and other visual characteristics of the product to reference standards. Analysis of data on the origin of the product. Machine learning can also be used to analyze product origin data. For example, using geolocation and supplier data, you can track the path of delivery of goods from the manufacturer to the store and identify inconsistencies in this process [2].

Using machine learning techniques to detect food fraud offers a number of important benefits. Fast and efficient**:** Machines can analyze large amounts of data much faster and more efficiently than humans.

Accuracy: Machine learning techniques can detect even small deviations, increasing the accuracy of hack detection.

Scalability: These methods can be easily scaled to handle product data across multiple stages of production and distribution.

Reducing health risks. By more accurately and quickly detecting counterfeits, consumer health risks are reduced as harmful additives and non-compliant ingredients can be identified and eliminated [3].

At this point, we are convinced that food and food recognition systems are architectures based on deep convolutional neural networks, now we will look at some algorithms and algorithm design, and a model for effectively applying CNN architecture in food recognition.

CNN is a multi-layer neural network with a unique architecture that uses complex features of the data at each layer to correctly determine the output. CNNs are well suited for perception

tasks. The figure above shows a simple CNN architecture with two convolutional layers and a simple architecture with two pooling layers in the feature extraction layer of the network. The classification layer of the network consists of three layers: the first smoothing layer transforms the multi-feature maps, namely the multi-dimensional array of the final convolution layer, into a one-dimensional array, the second is the dense layer and the prediction layer as the last layer. The output of the prediction layer is typically one-time encoded, that is, the output is a binary vector representation of categorical classes. The activated cell in this vector represents a CNN prediction, such that each class position is represented by a class-defining vector "0" and one vector "1". The CNN architecture is shown in Figure 2 [4].

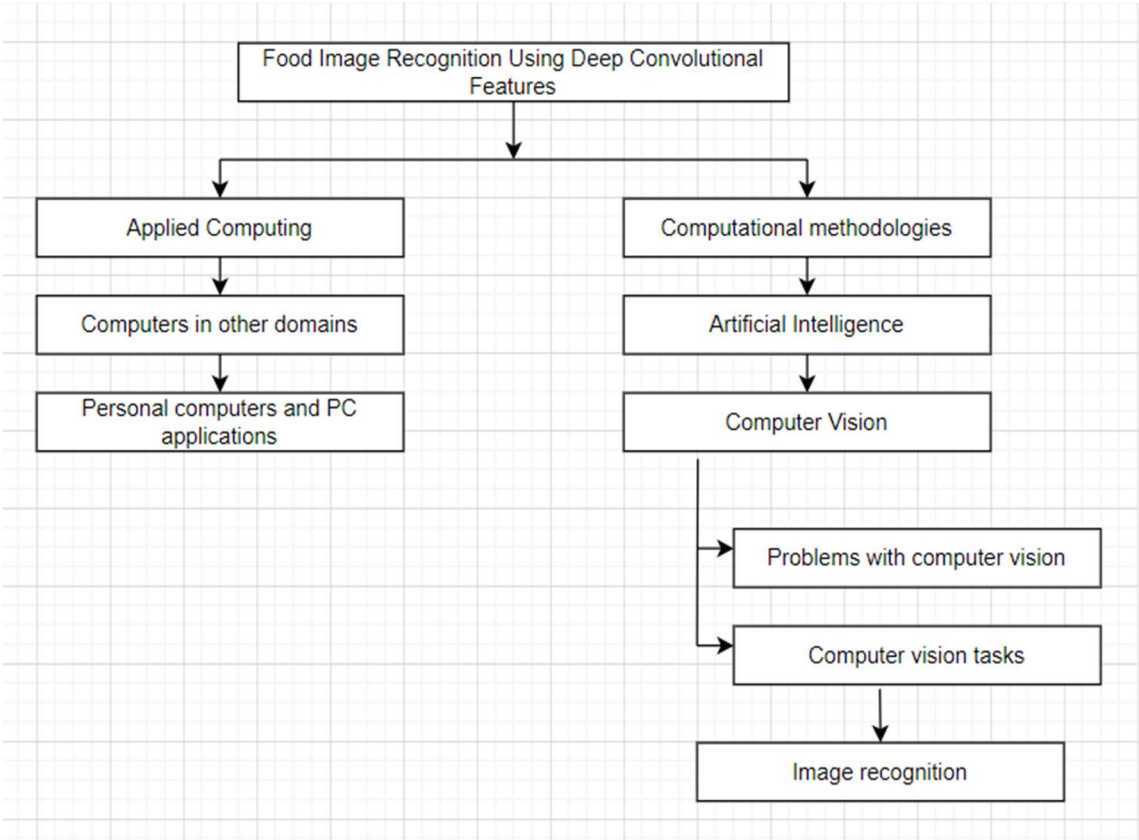


Figure 1 Food image recognition algorithm using deep convolutional functions

CNN is mainly used when there is an unstructured dataset and the model gets information from it. For example, if the task is to guess the name of a picture:

- The CNN takes a color image as a 3D matrix of pixels (three 2D matrices).
- During training, hidden layers allow the CNN to identify unique features by giving them shape
- suitable filters for extracting information from images;
- when the network training process approaches, the CNN can make a prediction and determine whether the image belongs to the class [5].

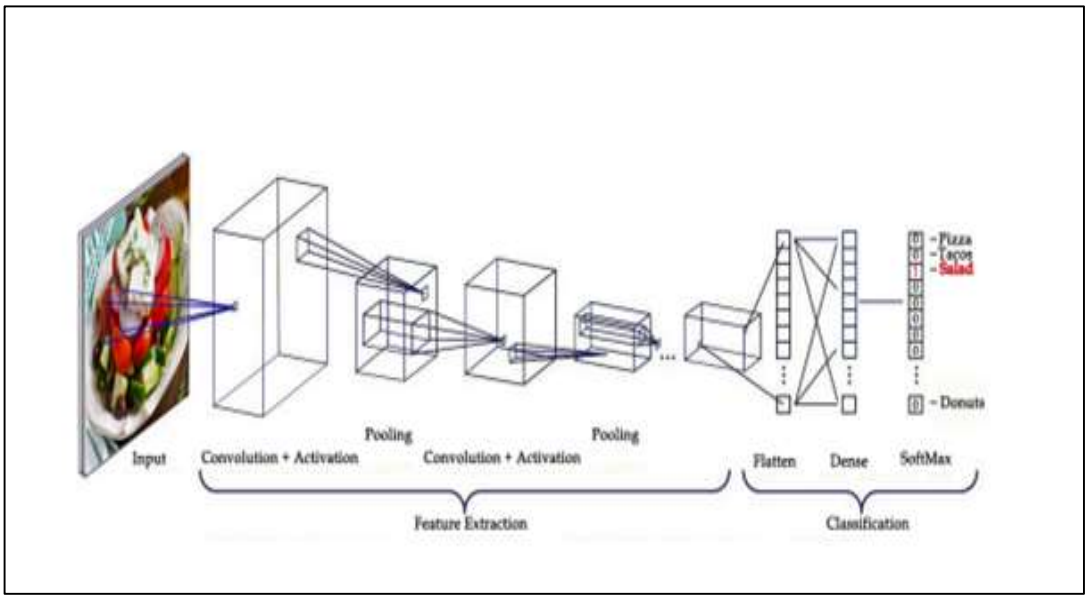


Figure 2 Simple CNN architecture

Training a machine learning model for image recognition

In this section, we present a CNN model and review a research project in Greece aimed at regional recognition of Greek cuisine for mobile tourism applications. This model is a framework based on the building blocks and principles of VGG. A common practice for developing new CNN architectures is to add convolutional layers until the model performs well. After this, parameters corresponding to the hyperparameters of the model are selected, a tuning process to high accuracy is carried out, and the model is created. This process takes a long time and requires a lot of computing resources. For example, the proposed new PureFoodNet model took a full month to develop and was properly configured on a high-performance computing system [5].

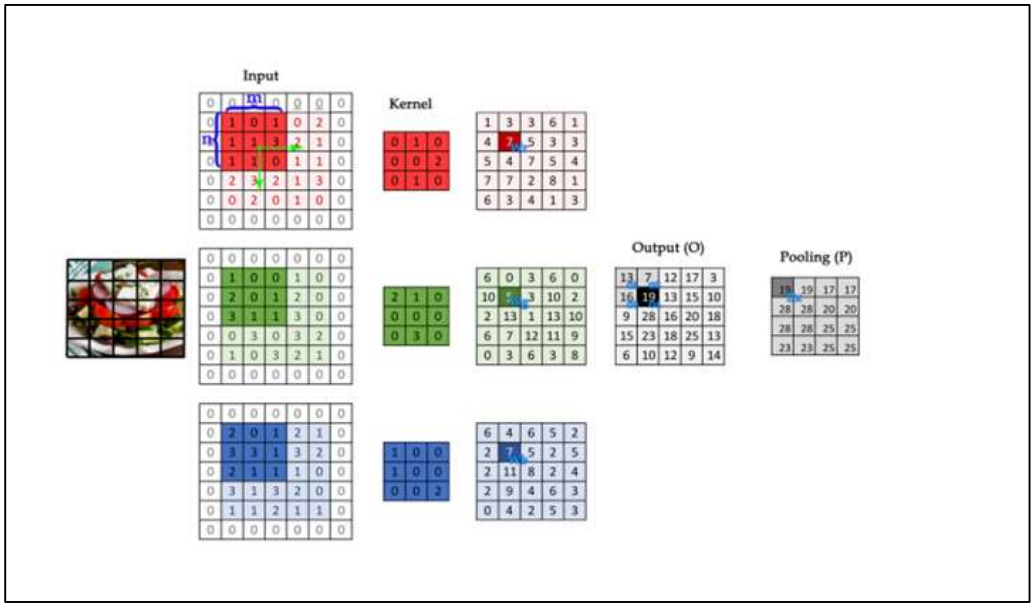


Figure 3 Analysis of the convolution process

Table 1

Comparative analysis of neural network models developed for food fraud detection

Developed neural network models for food adulteration detection	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Count of parameters	Depth	Output Step Time (ms) (CPU)	Output Step Time (ms) (GPU)
Xception	88	79.0%	94.5%	22.9M	81	109.4	8.1
VGG16	528	71.3%	90.1%	138.4M	16	69.5	4.2
VGG19	549	71.3%	90.0%	143.7M	19	84.8	4.4
ResNet50	98	74.9%	92.1%	25.6M	107	58.2	4.6
ResNet50V2	98	76.0%	93.0%	25.6M	103	45.6	4.4
ResNet101	171	76.4%	92.8%	44.7M	209	89.6	5.2
ResNet101V2	171	77.2%	93.8%	44.7M	205	72.7	5.4
InceptionV3	232	76.6%	93.1%	60.4M	311	127.4	6.5
MobileNet	16	70.4%	89.5%	4.3M	55	22.6	3.4
DenseNet201	14	71.3%	90.1%	3.5M	105	25.9	3.8
NASNetMobile	23	74.4%	91.9%	5.3M	389	27.0	6.7
NASNetLarge	343	82.5%	96.0%	88.9M	533	344.5	20.0
EfficientNetV2S	88	83.9%	96.7%	21.6M	-	-	-
EfficientNetV2M	220	85.3%	97.4%	54.4M	-	-	-
EfficientNetV2ML	479	85.7%	97.5%	119.0M	-	-	-
EfficientNetB6	166	84.0%	96.8%	43.3M	360	958.1	40.4
EfficientNetB7	256	84.3%	97.0%	66.7M	438	1578.9	61.6
ConvNeXtX-Large	1310	86.7%	-	350.1M	-	-	-

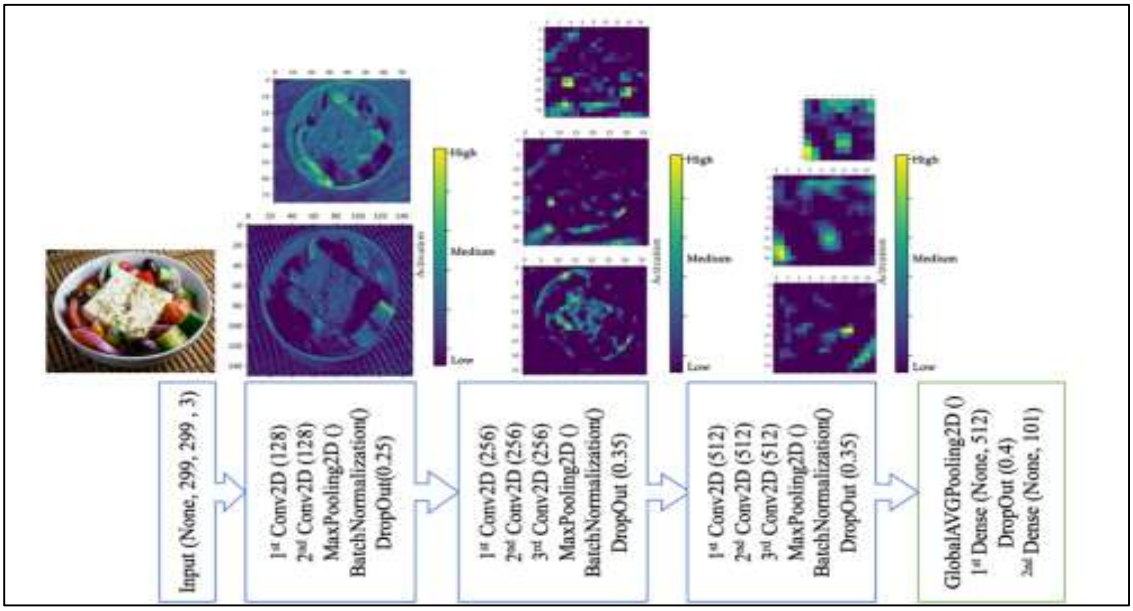


Figure 4 Simple CNN architecture for food recognition, PureFoodNet.

The PureFoodNet architecture consists of three transformation blocks and a classification block. The first block contains two convolutional layers, each containing 128 filters. The second block has three convolutional layers of 256 filters each, and the third block has three convolutional layers of 512 filters each. The final block or "classification layers," commonly referred to as "top layers," consist of a dense layer containing 512 neurons and a prediction layer. Built to match the 101 classes (food categories) of the Food101 dataset. An initialization method was also used to obtain maximum accuracy of PureFoodNet. A simple architecture is shown in Figure 4.

Work Outline Results: Our data consists of images such as: pudding, sandwich, french fries, pizza, soup, juice, caprese salad, etc. The data size used in the model is 10100 and we used data from 101 class. Based on the open technology of digital imaging and computer vision, "Using Machine Learning to Detect Fake Food," a paper on this topic, through in-depth exploration of the required datasets and experiments, can provide a sustainable solution to the pressing problem of fruits. and adulteration of vegetables and food products, and also protects consumers from many diseases and prevents the consumption of toxic additives that lead to premature death [7].

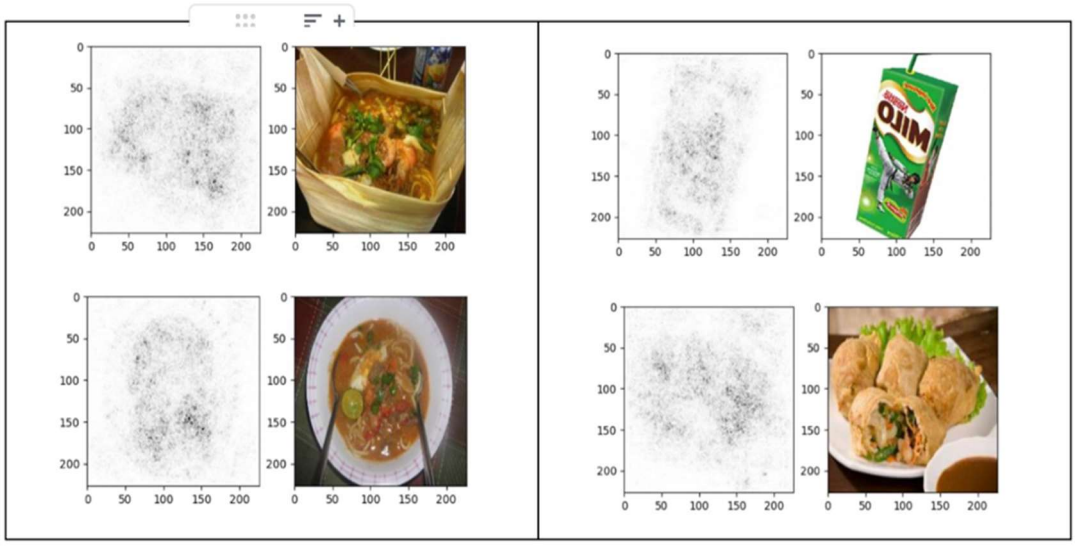


Figure 5 Some steps in food recognition

```
image_df['Label'].value_counts()

pork_chop        100
bread_pudding    100
club_sandwich    100
french_fries     100
beef_tartare     100
                ...
creme_brulee     100
hummus           100
clam_chowder     100
caprese_salad    100
foie_gras        100
Name: Label, Length: 101, dtype: int64
```

Figure 6 Image classification classes

```
Epoch 1/100
177/177 [==============================] - 72s 401ms/step - loss: 1.4190 - accuracy: 0.6243 - val_loss: 2.5008 - va
l_accuracy: 0.4059
Epoch 2/100
177/177 [==============================] - 76s 427ms/step - loss: 1.0923 - accuracy: 0.7056 - val_loss: 2.4580 - va
l_accuracy: 0.4243
Epoch 3/100
177/177 [==============================] - 77s 438ms/step - loss: 0.8376 - accuracy: 0.7707 - val_loss: 2.6106 - va
l_accuracy: 0.4024
Epoch 4/100
177/177 [==============================] - 78s 442ms/step - loss: 0.6321 - accuracy: 0.8373 - val_loss: 2.7404 - va
l_accuracy: 0.3960
Epoch 5/100
177/177 [==============================] - 79s 446ms/step - loss: 0.4582 - accuracy: 0.8913 - val_loss: 2.7505 - va
l_accuracy: 0.4144
```

Figure 7 Model training and evaluation

The label extraction process is shown in Figure 5. A related paper based on digital imaging and open source computer vision technology, “Using Machine Learning to Detect Fake Food,” through in-depth exploration of the required datasets and experiments, can provide a sustainable solution to the pressing problem. adulteration of fruits, vegetables and food products, and also protects consumers from many diseases and prevents the consumption of toxic additives that lead to premature death. And the main classes shown in Figure 6 have the ability to work with multiple adulteration methods are very important for studying the composition and quality of food products, since they provide more descriptors and signs of adulteration, which contributes to better classification [8]. Figure 7 shows the result during model training. Analytical chemists, IT professionals, and the general public are leading the research and development of food authentication technologies based on their methodological expertise. However, food authentication is the most important interdisciplinary field, including instrumentation, biology, computer science, mathematics and statistics, agriculture and food technology, and is important for everyone. Machine learning-based image and text recognition and classification could be used as a key technology in the fight against food contamination. In food production, to reduce the cost of the product, low-quality and non-standard products are used, which are very dangerous for the human body. During this thesis, the optimal way to identify unsuitable and harmful products for the human body using machine learning was demonstrated [9].

Machine learning algorithm configuration

The second alternative to deep learning-based food recognition methods is the recognition process, which is implemented using transfer learning methodology based on the use of popular machine learning models. This method uses existing machine learning models. For example, if a pre-trained PureFoodNet model is used, the top layer should be removed and only the convolutional layers should be retained, and the weights should also be retained. New top layers should be added according to the needs of the new problem, and the number of outputs in the prediction layer should be the same as the number of classes obtained from classification. Many iterations will be required to reconfigure the parameters and achieve the desired accuracy [10].

Due to the ease of use of transfer learning, very few people are currently developing new CNNs and learning models from scratch. Also, as mentioned earlier, developing a new model is a complex and time-consuming process. Instead, it is more convenient to reuse an already well-trained model. There is a large dataset such as ImageNet which contains over 14 million images and over 20,000 images in total. Also, given that ImageNet has many classes with

images depicting foods or edibles in general. Thus, a pre-trained model is expected in the ImageNet dataset, which is very efficient and easy to tune to improve the accuracy of the model in food recognition.

Moreover, in this section, we will present a comparative study of the above deep learning methods. As an image-based food recognition model, we will test the PureFoodNet architecture and Food101 dataset against other popular pre-trained models, and deep learning platforms will be tested against each other to identify their strengths and weaknesses [11].

Table 2.

Results during model evaluation

	Precision	Recall	f1-score
Rottenbanana	0.99	0.98	0.99
Freshoranges	0.99	0.84	0.91
Rottenoranges	0.84	0.95	0.89
Freshbanana	1.00	0.95	0.97
Rottenapples	0.96	0.93	0.95
freshapples	0.89	1.00	0.94

First of all, PureFoodNet has a much weaker architecture than off-the-shelf models. The models compared were the VGG16, InceptionV3, ResNet50, and Inception ResNetV2 architectures. Counts. The experiment model is written in Python. These comparative studies are carried out in several directions. This is because the resulting models cannot be tested only in terms of accuracy. In addition to accuracy, it is important not to forget about model evaluation metrics such as f-score and Roc curve. In this case, it is worth paying attention to whether the model is used during training or at the testing stage in terms of time and computing device resources.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (1)$$

$$recall = \frac{TP}{TP+FN}, \quad (2)$$

$$specificity = \frac{TN}{FP+TN}, \quad (3)$$

$$F_1 score = \frac{Precision*recall}{Precision+reca} * 2, \quad (4)$$

All these formulas were used to estimate the main parameters of the model. With their help, we can evaluate how accurately the model predicts counterfeit products that destroy the human body [12].

Conclusion

While machine learning has great potential to make a significant difference in preventing adulteration in the food industry, it may struggle to address new challenges, such as the emergence of new types of food adulteration, through the manipulation of algorithms and data.

However, we can reduce the risk of food adulteration through machine learning. This will help create an open and secure food ecosystem while ensuring the safety and well-being of consumers. As a result, the image recognition process was implemented using VGG, MobileNet_v2 and InceptionV3 algorithms. The above algorithms were tested on Food-101 database and fruit image databases and the results were obtained. Comparing the above algorithms in terms of efficiency and in terms of image recognition, a conclusion was drawn. During the operation of the model, numerical and graphical results were obtained.

References

1. Yoshiyuki Kawano, Keiji Yanai. Food image recognition with deep convolutional features // Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014. - P. 589–593.
2. Krizhevsky A., Sutskever I., Hinton G. Imagenet classification with deep convolutional neural networks // In Advances in Neural Information Processing Systems 25, 2020. – P. 1106–1114.
3. Girshick R. B., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation // In Computer Vision and Pattern Recognition, 2014.
4. Atalyurek U.V.C., Aykanat C., Uc¸ar B. Two-dimensional sparse matrix partitioning: Models, methods, and a recipe // SIAM J. Sci. Comput. - 32(2):656–683, 2020.
5. Arora S., Bhaskara A., Ge R., Ma T. Provable bounds for learning some deep representations. CoRR, abs/1310.6343, 2014.
6. Erhan D., Szegedy C., Toshev A., Anguelov D. Scalable object detection using deep neural networks // CVPR, 2014.
7. van de Sande K. E.A., Uijlings J. R.R., Gevers T., ASmeulders. W.M. Segmentation as selective search for object recognition // In Proceedings of the 2011 International Conference on Computer Vision, ICCV ’11, 2011. – P. 1879–1886.
8. Hinton G. E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R. Improving neural networks by preventing co-adaptation of feature detectors // CoRR, abs/1207.0580, 2021.
9. Farinella G.M., Moltisanti M., Battiato S. Classifying Food Images Represented as Bag of Textons // In International Conference on Image Processing, 2014. - P. 5212—-5216.
10. Krizhevsky A., Sutskever I., Hinton G. Imagenet classification with deep convolutional neural networks // In Advances in Neural Information Processing Systems 25, 2012. – P. 1106–1114.
11. LeCun Y., Boser B., Denker J.S., Henderson D., Howard R.E., Hubbard W., L. D. Jackel. Backpropagation applied to handwritten zip code recognition // Neural Comput. 1(4):541–551, Dec. 2019.
12. Toshev A., Szegedy C. Deeppose: Human pose estimation via deep neural networks // CoRR, abs/1312.4659, 2013.

ЧАТ-БОТЫ, ГОЛОСОВЫЕ ПОМОЩНИКИ И УМНЫЕ КОЛОНКИ: ОБЗОР И ПЕРСПЕКТИВЫ

Баймиров К.М., Мергенгали Е.К.
e-mail: kuanysh23@gmail.com, estony.9999@gmail.com
Казахский национальный университет имени аль-Фараби, Казахстан

Аннотация. *В последние годы широкое внедрение технологий искусственного интеллекта (ИИ) и обработки естественного языка (NLP) привело к разработке различных диалоговых интерфейсов, в том числе чат-ботов, голосовых помощников и умных колонок. Эти интеллектуальные системы произвели революцию в том, как мы взаимодействуем с компьютерами, и стали неотъемлемой частью нашей повседневной жизни. В этой статье представлен обзор чат-ботов, голосовых помощников и умных колонок, обсуждаются лежащие в их основе технологии, приложения и возможные направления развития. Мы изучаем достижения в NLP и ИИ, благодаря которым эти диалоговые агенты становятся все более функциональными и удобными для пользователя. Мы рассматриваем проблемы и этические соображения, связанные с внедрением этих технологий, и выделяем потенциальные направления исследований для дальнейшего продвижения в этой области.*

Введение

Мотивация разработки чат-ботов, голосовых помощников и умных колонок связана с необходимостью создания более интуитивно понятных и естественных способов взаимодействия людей с машинами. Традиционные человеко-компьютерные интерфейсы, такие как графические пользовательские интерфейсы (GUI) и интерфейсы командной строки, часто требуют от пользователей изучения определенных команд или навигации по сложным меню. Это создает барьер между людьми и машинами, ограничивая доступность и участие пользователей.

Диалоговые интерфейсы, с другой стороны, используют достижения в области искусственного интеллекта и NLP, чтобы обеспечить более человеческое взаимодействие. Понимая естественный язык и реагируя в разговорной манере, эти технологии устраняют разрыв между людьми и машинами, делая технологии более доступными и удобными.

Преимущества и возможности

Улучшенный пользовательский интерфейс. Чат-боты, голосовые помощники и умные колонки обеспечивают пользователям более плавное и интуитивно понятное взаимодействие. Они позволяют пользователям общаться с машинами более естественным образом, подобно тому, как они взаимодействуют с другими людьми. Это приводит к повышению вовлеченности пользователей, удовлетворенности и эффективности. С появлением смартфонов и интеллектуальных колонок разговорные интерфейсы стали легко доступны на различных устройствах и в различных средах. Пользователи могут получать доступ к информации, управлять устройствами и выполнять задачи с помощью голоса, что устраняет необходимость в физическом взаимодействии или визуальном внимании. Эта доступность особенно полезна для

людей с ограниченными возможностями или ограниченной подвижностью. Чат-боты, голосовые помощники и умные колонки способны понимать и запоминать пользовательские предпочтения, контекст и историю. Это позволяет им предоставлять персональные рекомендации, индивидуальную помощь и индивидуальный подход. Они могут действовать как виртуальные личные помощники, помогая пользователям с задачами, напоминаниями, планированием и многим другим. Диалоговые интерфейсы позволяют автоматизировать и оптимизировать процессы, предоставляя быстрые и эффективные ответы на запросы и запросы пользователей. Они могут справляться с рутинными и повторяющимися задачами, высвобождая человеческие ресурсы, чтобы сосредоточиться на более сложных и полезных действиях. Голосовые помощники и умные колонки служат центральными узлами для управления устройствами умного дома и экосистемами Интернета вещей (IoT). Пользователи могут использовать голосовые команды для управления освещением, термостатами, бытовой техникой и системами безопасности, создавая подключенную и интеллектуальную среду обитания. Чат-боты все чаще используются в сфере обслуживания и поддержки клиентов, что позволяет компаниям обрабатывать запросы клиентов, оказывать круглосуточную помощь и сокращать время отклика. Голосовые помощники также интегрируются в различные отрасли, включая здравоохранение, автомобилестроение и розничную торговлю, для повышения качества обслуживания клиентов и оптимизации операций. Разработка чат-ботов, голосовых помощников и умных колонок способствовала развитию ИИ, NLP и смежных областей. Исследователи постоянно изучают новые методы и алгоритмы для улучшения понимания языка, моделирования контекста, управления диалогами и взаимодействия с пользователем, внедряя инновации в более широкую область разговорного ИИ.

Таким образом, чат-боты, голосовые помощники и умные колонки играют жизненно важную роль в преобразовании взаимодействия человека с компьютером. Они улучшают взаимодействие с пользователем, обеспечивают повсеместный доступ, персонализируют помощь, автоматизируют задачи, интегрируются с IoT и стимулируют исследования и инновации. Поскольку эти технологии продолжают развиваться, ожидается, что их значение в различных областях и их влияние на общество значительно возрастут. Чат-боты, голосовые помощники и умные колонки – все это диалоговые интерфейсы, в которых для взаимодействия с пользователями используются технологии искусственного интеллекта и NLP. Несмотря на то, что они имеют некоторое сходство, существуют явные различия в их функциях и дизайне:

Различия между чат-ботами, голосовыми помощниками и умными колонками

Чат-боты – это компьютерные программы, предназначенные для имитации человеческого разговора с помощью текстовых или голосовых взаимодействий. Чат-боты в основном работают через текстовые интерфейсы, такие как приложения для обмена сообщениями, веб-сайты или платформы социальных сетей. В некоторых случаях они также могут поддерживать голосовое взаимодействие. Чат-боты обычно используются в сфере обслуживания клиентов, электронной коммерции, виртуальных помощниках и системах поиска информации. Чат-боты, как правило, ориентированы на задачи, ориентированы на предоставление конкретной информации, ответы на запросы или выполнение предопределенных действий. Пользователи обычно инициируют общение с чат-ботами, отправляя сообщения или запросы. Чат-боты могут использовать подходы на основе правил или более продвинутые методы, такие как алгоритмы

машинного обучения, для понимания и генерации ответов. Примеры: чат-боты службы поддержки, виртуальные помощники по покупкам, чат-боты для изучения языка.

Голосовые помощники – это программные приложения на базе искусственного интеллекта, которые используют распознавание и синтез голоса для обеспечения голосового взаимодействия с пользователями. Голосовые помощники предназначены для голосового взаимодействия и могут понимать разговорный язык и реагировать на него. Они обычно используются в смартфонах, интеллектуальных колонках и других устройствах с голосовой поддержкой для различных задач, включая установку напоминаний, воспроизведение музыки, ответы на вопросы и управление устройствами «умный дом». Голосовые помощники универсальны и могут обрабатывать как целевые, так и общие разговорные запросы. Пользователи обычно инициируют взаимодействие с голосовыми помощниками, произнося голосовые команды или запросы.

Голосовые помощники полагаются на передовые методы NLP, системы распознавания речи и управления диалогами для обработки и ответа на запросы пользователей. Примеры: Amazon Alexa, Яндекс Алиса, Apple Siri, Google Assistant, Microsoft Cortana.

Умные колонки – это аппаратные устройства, оснащенные встроенными голосовыми помощниками, разработанными специально для голосового взаимодействия в режиме громкой связи. Умные колонки представляют собой специальное устройство для голосового взаимодействия и часто включают в себя микрофоны, динамики и варианты подключения. Умные колонки в основном используются в качестве узлов домашней автоматизации, позволяя пользователям управлять устройствами умного дома, воспроизводить музыку, получать сводки о погоде и получать доступ к целому ряду услуг на основе голосового помощника. Умные колонки ориентированы на обеспечение удобного голосового доступа к различным сервисам и управление подключенными устройствами. Пользователи могут инициировать взаимодействие со смарт-динамиками, используя слова пробуждения (например, «Привет, Siri», «Alexa», «Алиса»), за которыми следуют голосовые команды или запросы. Умные колонки сочетают в себе аппаратные компоненты для обработки звука, такие как формирование луча и шумоподавление, с программным обеспечением голосового помощника для обеспечения голосового взаимодействия. Примеры: Amazon Echo (Alexa), Яндекс, Google Home (Google Assistant), Apple HomePod (Siri).

В то время как чат-боты, голосовые помощники и умные колонки призваны облегчить взаимодействие человека и машины, их особый дизайн, варианты использования и режимы взаимодействия отличают их друг от друга. Чат-боты часто основаны на тексте, ориентированы на задачи и используются на платформах обмена сообщениями, в то время как голосовые помощники ориентированы на голосовое взаимодействие и интегрированы в различные устройства. Умные колонки — это специализированные аппаратные устройства, оснащенные встроенными голосовыми помощниками, которые в основном используются для голосового управления и домашней автоматизации.

История чат-ботов

История чат-ботов начинается с середины 20 века, когда были разработаны первые элементарные системы, из которых наиболее известными являются: ELIZA (1966) [1], основанный на правилах чат-бот, предназначенный для имитации роджерианского психотерапевта, и PARRY (1972), который имитировал человека с параноидальной шизофренией. Эти ранние чат-боты использовали простое сопоставление с образцом и предопределенные ответы. Затем на рынке появились чат-боты, основанные на правилах

- предопределенных правилах и шаблонах для определения намерений пользователей и генерирования соответствующих ответов. Сравнение между наиболее популярными чат-ботами на основе правил Dialogflow и IBM Ватсон показывается в [2]. Появление методов искусственного интеллекта и машинного обучения произвело революцию в возможностях чат-ботов. Вместо использования предопределенных правил чат-боты на основе ИИ стали учиться на данных и адаптировать свои ответы на основе взаимодействия с пользователем. Недавние достижения в NLP - встраивание слов [3], языковые модели (GPT-4) [4] и анализ настроений [5] - значительно повысили производительность чат-ботов. Эти методы позволяют чат-ботам понимать и генерировать человекоподобный текст, интерпретировать намерения пользователей, участвовать в более содержательных беседах. Появление умных колонок и виртуальных помощников расширило возможности чат-ботов для поддержки голосового взаимодействия [6]. Чат-боты, интегрированные с голосовыми интерфейсами, позволяют пользователям общаться на естественном языке, повышая удобство и доступность.

Чат-боты широко используются в обслуживании клиентов, обеспечивая автоматизированную и мгновенную поддержку клиентов [7]. Платформы электронной коммерции используют чат-боты для рекомендаций продуктов и отслеживания заказов. Чат-боты также облегчают персонализированное взаимодействие в таких секторах, как здравоохранение, финансы и образование, предлагая информацию, советы и помощь. Достижения в области разговорного ИИ позволили чат-ботам вести более сложные и контекстно-зависимые разговоры [8]. Чат-боты могут вести историю диалогов, запоминать предпочтения пользователей и давать персональные рекомендации. Они также могут поддерживать многосторонние разговоры, понимать сложные вопросы и отвечать на них. Современные чат-боты часто используют гибридные подходы, сочетая системы, основанные на правилах, с методами искусственного интеллекта [9]. Компоненты на основе правил обрабатывают определенные знания предметной области и предопределенные потоки, а модели ИИ улучшают способность чат-бота генерировать динамические и контекстно-зависимые ответы. Чат-боты обеспечивают более разговорный и удобный интерфейс по сравнению с традиционными графическими пользовательскими интерфейсами (GUI), снижая когнитивную нагрузку и повышая вовлеченность пользователей [10], могут работать круглосуточно, предоставляя мгновенные ответы и поддержку, тем самым повышая удовлетворенность клиентов и сокращая время отклика, могут одновременно вести несколько разговоров, что позволяет организациям эффективно обрабатывать большие объемы взаимодействий с пользователями [11], могут автоматизировать повторяющиеся задачи и поддержку клиентов, снижая операционные расходы для бизнеса.

С помощью методов NLP и искусственного интеллекта чат-боты могут изучать предпочтения пользователей и персонализировать их ответы, что приводит к более адаптированным и релевантным взаимодействиям. Чат-боты прошли долгий путь от простых систем, основанных на правилах, до сложных диалоговых агентов, основанных на искусственном интеллекте и NLP. Их эволюция значительно улучшила взаимодействие человека с компьютером, предлагая улучшенный пользовательский опыт, масштабируемость, экономию средств и персонализацию. Благодаря постоянным достижениям в области искусственного интеллекта и NLP чат-боты готовы продолжать трансформировать то, как мы взаимодействуем с технологиями и получаем доступ к информации и услугам.

Разработка чат-ботов для умных колонок

Разработка чат-ботов для умных колонок включает в себя сочетание подходов и методов, обеспечивающих понимание естественного языка, распознавание речи и управление диалогами. Вот некоторые из ключевых подходов и методов, используемых при разработке чат-ботов для умных колонок.

Распознавание именованных объектов (NER): идентификация и извлечение именованных объектов, таких как имена, местоположения или даты, из пользовательского ввода [14].

Анализ настроений - анализ настроений или эмоций, выраженных в пользовательском вводе, для предоставления соответствующих ответов.

Распознавание речи имеет решающее значение для обеспечения голосового взаимодействия с умными колонками. Методы, используемые для распознавания речи, включают: акустическое моделирование, моделирование языка, методы машинного обучения и глубокого обучения.

Акустическое моделирование – это создание статистических моделей, которые сопоставляют характеристики звука с фонемами или словами [15].

Моделирование языка - разработка языковых моделей, которые предсказывают вероятность последовательностей слов для повышения точности и улучшения распознавания [16]. Обнаружение ключевых слов: определение определенных ключевых слов или пробуждающих слов, которые вызывают активацию чат-бота.

Для расширения возможностей чат-ботов используются методы машинного обучения и глубокого обучения. Эти методы включают в себя:

- Распознавание намерений с помощью использование алгоритмов обучения с учителем, таких как машины опорных векторов (SVM) [17], случайные леса [18], или моделей глубокого обучения, таких как рекуррентные нейронные сети (RNN) или модели Transformer [19], для классификации намерений пользователя.
- Управление диалогами - использование таких методов, как системы, основанные на правилах, автоматы с конечным числом состояний или обучение с подкреплением, для управления потоком разговоров и обработки многооборотных диалогов [20].
- Генерация ответов - использование таких подходов, как модели «последовательность к последовательности» [21], рекуррентные нейронные сети (RNN) или модели преобразования для создания контекстуально релевантных и естественно звучащих ответов.
- Для обеспечения точных и информативных ответов, чат-боты часто полагаются на базы знаний и методы генерации языка. К ним относятся:
- графики знаний, построение графиков знаний, которые представляют структурированную информацию и взаимосвязи, позволяющие извлекать соответствующие факты или данные;
- поиск информации, использование алгоритмов или методов поиска, таких как TF-IDF (частота терминов, обратная частоте документа), для извлечения соответствующей информации из больших корпусов или баз данных;
- генерация языка на основе шаблонов или правил - использование предопределенных шаблонов или правил для генерации ответов на основе пользовательского ввода;

- Персонализация и контекстная осведомленность. Для улучшения взаимодействия с пользователем чат-боты для умных колонок используют методы персонализации и учета контекста, в том числе:
- профилирование пользователей: сбор и хранение пользовательских предпочтений, истории и контекстной информации для адаптации ответов, и рекомендаций;
- управление контекстом: отслеживание истории разговоров, чтобы понимать запросы пользователей и соответствующим образом реагировать на них с учетом контекста предыдущих взаимодействий.
- Интеграция с API и сервисами, чат-боты Smart Speaker часто интегрируются с внешними API и службами для предоставления дополнительных функций, таких как обновления погоды, поиск новостей или управление домашней автоматикой. Методы интеграции включают в себя:
- Интеграция API - подключение к внешним API для получения информации или выполнения определенных задач.
- Web Scraping - извлечение данных с веб-сайтов для предоставления актуальной информации или создания динамических ответов.
- Интеграция с Интернетом вещей (IoT): подключение к устройствам умного дома или платформам IoT для управления подключенными устройствами и выполнения задач домашней автоматизации.

Эти подходы и методы составляют основу для разработки чат-ботов для умных колонок. Конкретная комбинация и реализация зависят от требований, сложности и варианта использования чат-бота.

Проблемы и будущие направления

Чат-боты для умных колонок добились значительного прогресса, но они по-прежнему сталкиваются с рядом проблем и ограничений. Вот некоторые из основных проблем и способы их решения. На точность распознавания речи могут влиять такие факторы, как фоновый шум, акценты и вариации речевых паттернов. Постоянное совершенствование технологии автоматического распознавания речи (ASR) направлено на повышение точности и надежности. Такие методы, как глубокое обучение и акустическое моделирование, помогают обучать модели на больших и разнообразных наборах данных, что приводит к повышению эффективности распознавания речи.

Понимание естественного языка. Понимание и точная интерпретация намерений и контекста пользователя на естественном языке могут быть сложными, особенно с двусмысленными или диалоговыми запросами. Методы обработки естественного языка (NLP), включая распознавание намерений, распознавание именованных объектов и анализ настроений, используются для улучшения понимания чат-ботом вводимых пользователем данных. Модели машинного обучения и языковые модели помогают фиксировать и анализировать нюансы естественного языка.

Поддержание контекста и управление многоходовыми разговорами для предоставления последовательных и актуальных ответов может быть сложной задачей.

Такие методы, как отслеживание состояния диалога, управление памятью и обучение с подкреплением, используются для того, чтобы чат-боты могли поддерживать контекст и эффективно управлять потоком диалога. Гибридные подходы, сочетающие системы на основе правил и модели машинного обучения, помогают обрабатывать сложные разговоры и давать соответствующие ответы.

Чат-ботам требуется доступ к точным и актуальным знаниям и данным для предоставления актуальной информации. Однако получение и поддержание всеобъемлющей базы знаний может оказаться сложной задачей.

Решение: чат-боты интегрированы с внешними источниками данных, API и сервисами для доступа к информации в режиме реального времени. Графы знаний и методы поиска информации используются для хранения и извлечения соответствующих данных. Механизмы непрерывного обучения и обновления реализованы для поддержания базы знаний в актуальном состоянии.

Персонализация взаимодействий на основе пользовательских предпочтений, истории и контекста требует сбора и использования пользовательских данных с соблюдением требований конфиденциальности. Методы профилирования пользователей позволяют чат-ботам персонализировать ответы на основе пользовательских данных, таких как прошлые взаимодействия, предпочтения и демографические данные. Меры конфиденциальности и протоколы защиты данных реализованы для обеспечения безопасности и конфиденциальности пользовательской информации. Понимание эмоций и выражений пользователей и адекватная реакция на них могут быть сложными для чат-ботов. Достижения в области анализа настроений и аффективных вычислений позволяют чат-ботам распознавать эмоции пользователей и реагировать на них. Такие методы, как анализ настроений, анализ тона и моделирование эмпатии, используются для повышения эмоционального интеллекта и обеспечения более чутких и адекватных ответов. Обеспечение совместимости и бесшовной интеграции с различными платформами интеллектуальных колонок, операционными системами и внешними API-интерфейсами может оказаться сложной задачей. Разработчики сосредоточены на создании чат-ботов, совместимых с популярными платформами умных колонок, такими как Amazon Alexa, Google Assistant или Apple Siri. Стандартизированные протоколы и API используются для облегчения интеграции с внешними службами и устройствами.

Чат-боты должны придерживаться этических принципов, уважать конфиденциальность пользователей и соблюдать правовые нормы, такие как законы о защите данных и требования согласия пользователей. Разработчики и организации отдают приоритет этическим соображениям, внедряют меры конфиденциальности и соблюдают требования законодательства. Прозрачное использование данных, механизмы согласия пользователей и четкие политики конфиденциальности используются для обеспечения этического и правового соответствия.

Эти проблемы и ограничения решаются с помощью постоянных исследований и разработок с использованием достижений в области технологий искусственного интеллекта, NLP и распознавания речи. Ожидается, что по мере развития технологии чат-ботов эти проблемы будут смягчаться, что приведет к созданию более сложных и функциональных чат-ботов для умных колонок.

Выводы

В статье освещаются значительные достижения в технологии чат-ботов для умных колонок. К ним относятся улучшения точности распознавания речи, понимания естественного языка, понимания контекста и управления диалогами. Машинное обучение и методы глубокого обучения сыграли решающую роль в расширении возможностей чат-ботов, обеспечивая более точные и контекстно-зависимые ответы.

В статье представлен широкий спектр приложений для чат-ботов на умных колонках. Они широко используются в виртуальной помощи, поддержке клиентов, домашней автоматизации, развлечениях и СМИ, образовании, здравоохранении,

финансах, путешествиях и личной продуктивности. Чат-боты облегчают персонализированный опыт, предлагают информацию в режиме реального времени и упрощают различные задачи для пользователей в разных областях. Было выявлено несколько проблем, связанных с чат-ботами для умных колонок. Эти задачи включают повышение точности распознавания речи в различных средах, улучшение понимания естественного языка и понимания контекста, получение и поддержание точных баз знаний, обеспечение конфиденциальности и безопасности данных, а также работу с эмоциональным интеллектом. Решение этих проблем требует постоянных исследований и разработок. Чат-боты на умных колонках могут значительно улучшить взаимодействие с пользователем и обеспечить большее удобство. Пользователи могут взаимодействовать с умными колонками посредством разговоров на естественном языке, упрощая задачи и без труда получая доступ к информации. Чат-боты обеспечивают автоматизацию и снижают потребность в человеческом вмешательстве в различные задачи и услуги. Это приводит к повышению эффективности, сокращению времени отклика и экономии средств для предприятий и организаций.

В будущем ожидается, что постоянные улучшения в технологиях искусственного интеллекта, NLP и распознавания речи будут способствовать дальнейшему совершенствованию возможностей чат-ботов, обеспечивая более естественное и интеллектуальное взаимодействие. Интеграция с новыми технологиями, такими как дополненная реальность (AR) и Интернет вещей (IoT), расширит область применения и возможности чат-ботов на умных колонках.

В статье освещаются достижения, приложения и проблемы в разработке чат-ботов для умных колонок. Чат-боты предлагают улучшенный пользовательский интерфейс, автоматизацию, персонализацию и доступность. Однако необходимо решить проблемы, связанные с распознаванием речи, пониманием естественного языка, пониманием контекста и конфиденциальностью. Потенциальное влияние чат-ботов на экосистемы интеллектуальных колонок является значительным, и будущие перспективы указывают на дальнейшие улучшения и расширение приложений для чат-ботов.

Литература

1. Wiezenbaum, J. ELIZA - A computer Program For the Study of Natural Language Communication Between Man and Machine // Communications of the ACM 9(1), 1966. – P.36-45, http://dx.doi.org/10.1145/365153.365168
2. Sandeep, A. Thorat, Vishakha, Jadhav. A Review on Implementation Issues of Rule-based Chatbot Systems // Proceedings of the International Conference on Innovative Computing & Communications (ICICC), 2020.
3. Qilu Jiao, Shunyao Zhang. A Brief Survey of Word Embedding and Its Recent Development, 2021 // IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). DOI: 10.1109/IAEAC50856.2021.9390956.
4. OpenAI Adrien Ecoffet. GPT-4 Technical Report. https://doi.org/10.48550/arXiv.2303.08774.
5. Wagner da Silva Maciel Sodré, Julio Cesar Duarte. Chatbot Optimization using Sentiment Analysis and Timeline Navigation // Journal of Theoretical and Applied Informatics, DOI: https://doi.org/10.22456/2175-2745.125825.
6. Francis Rakotomalala and others. Voice User Interface: Literature Review, Challenges and Future Directions. DOI:10.52846/stccj.2021.1.2.26 // System theory, control and computing journal, 2021. - Vol. 1, No. 2. – P. 65-89.

7. Eleni Adamopoulou, Lefteris Moussiades. Chatbots: History, technology, and applications // Machine Learning with Applicationsю - Volume 2, 2020. https://doi.org/10.1016/j.mlwa.2020.100006.
8. Mohit Jain and others. Convey: Exploring the Use of a Context View for Chatbots // CHI Conference, 2018. - DOI:10.1145/3173574.3174042.
9. Julio Villena. Hybrid Approach Combining Machine Learning and a Rule-Based Expert System for Text Categorization // Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference, 2011, USA.
10. Daniëlle Duijst. Can we Improve the User Experience of Chatbots with Personalisation?, Thesis for: MSc Information Studies (Human-Centered Multimedia), DOI:10.13140/RG.2.2.36112.92165.
11. Alistair Barros. Caling up chatbots for corporate service delivery systems // Communications of the ACM, 2021. - Volume 64, Issue 826. - pp 88–97, https://doi.org/10.1145/3446912
12. Jonathan Webster, Chunyu Kit. Tokenization as the initial phase in NLP // Proceedings of the 14th conference on Computational linguistics. Vol. 4, DOI:10.3115/992424.992434.
13. Hongwei Li and others. Part-of-Speech Tagging with Rule-Based Data Preprocessing and Transformer. Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, China, 2021.
14. Rachna Jain and others. Named Entity Recognition in English Text. Journal of Physics: Conference Series 1712 (2020) 012013, IOP Publishing, doi:10.1088/1742-6596/1712/1/012013.
15. S. Bhatt, A. Jain, A. Dev. CICD acoustic modeling based on Monophone and Triphone for HINDI Speech Recognition. // International Conference on Artificial Intelligence and Speech Technology (AIST2019) 2019.
16. Chengwei Wei and others. An Overview on Language Models: Recent Developments and Outlook, 2023. - https://doi.org/10.48550/arXiv.2303.05759.
17. Marcelo Mendoza, Juan Zamora. Identifying the Intent of a User Query Using Support Vector Machines. 2009, DOI:10.1007/978-3-642-03784-9_13 // String Processing and Information Retrieval, 16th International Symposium, SPIRE 2009.
18. Breiman, L. Random Forests. 2001. Machine Learning 45(1):5-32, DOI:10.1023/A:1010950718922.
19. Abir Rahali, Moulay A. Akhloufi. End-to-End Transformer-Based Models in Textual-Based NLP. Perception, Robotics, and Intelligent Machines Research Group (PRIME) // Department of Computer Science, Université de Moncton, Moncton, Canada, AI 2023, 4(1), 54-110; https://doi.org/10.3390/ai4010004.
20. Jinjie Ni. Recent advances in deep learning based dialogue systems: a systematic survey // Artificial Intelligence Review, 2023. 56:3055–3155, https://doi.org/10.1007/s10462-022-10248-8.
21. Shansan Gong. DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models // Computation and Language (cs.CL); Feb 2023, https://doi.org/10.48550/arXiv.2210.08933.

APPLICATION OF STATISTICAL MODELS FOR FORECASTING ASSETS OF OIL AND GAS COMPANIES

Teshebayev N.M., Akhmetova A.M.
e-mail: nurdauletteshebaev@gmail.com, ardak_66@mail.ru
Al-Farabi Kazakh National University, Kazakhstan

Abstract. *Investment activity plays an important and strategic role in the development of companies and ensuring their stability in the market. This activity is associated with spending on asset acquisition, technological progress, expansion of production, research and development, as well as many other areas that together form the basis for the long-term success of the organization. This thesis will present a study on the use of machine learning methods to predict important parameters in business, such as investments, their impact on the financial condition of the company, as well as methods of effective management of investment resources in order to ensure sustainable development and prosperity of the enterprise.*

Introduction

The last five years have been characterized by the expansion and development of the scope of intelligent information systems in the financial sectors.The use of intelligent information computing systems allows you to reduce the economic stability of the company, the human factor, automate the process of buying and selling shares, and make the right decisions taking into account the factors affecting the value of shares. At the present time, when competition in the market is becoming more intense, investment management is becoming a key tool for creating value and ensuring the growth of the company.

Over the past decade, there has been increased attention to research in the field of analysis of investment activity of companies. Researchers use various methods and tools to analyze time series. In recent years, with the development of computer technology and machine learning methods, new and more powerful methods of time series analysis have become available. For the analysis of time series, various factors have recently been taken into account, depending on the specific field of application and the purpose of the study. In order to understand price dynamics and predict future movements in financial markets, when analyzing financial time series, factors such as prices and trading volumes (changes in prices and volumes to identify trends, volatility and other patterns), technical indicators (analyzing price charts and identifying signals), fundamental indicators (profitability of companies, dividends payments, profit and loss statements, financial condition and other data to assess the fundamental value of assets), geopolitical events (political changes, conflicts, trade disputes and other geopolitical events can have a significant impact on financial markets), news and events in the world (crises, court decisions), alternative data (data from social media, text data analysis).

To predict the time series, I chose the data of the company KazTransOil, covering the period from 2014 to the current date. This time interval allows us to more fully investigate the long-term dynamics and trends in the company's activities, as well as to identify the impact of various factors on its results. Analyzing this data and making forecasts will allow you to better understand the future development of the company and make more informed decisions in investment or strategic areas.

The relevance of the problem of forecasting financial assets for domestic and global financial markets is also confirmed by an extensive list of scientific papers carried out by domestic and foreign researchers. The following works published in recent years should be noted: Adebiyi A Ariyo, Adewumi O Adewumi, and Charles K Ayo. “Stock price prediction using the ARIMA model”,2014; H. Hewamalage, C. Bergmeir and K. Bandara. “Recurrent neural networks for time series forecasting: Current status and future directions”,2020; David MQ Nelson, Adriano CM Pereira, and Renato A de Oliveira. “Stock market’s price movement prediction with LSTM neural networks”,2017; J. C. B. Gamboa. “Deep learning for time-series analysis”,2017; T. Fischer and C. Krauss. “Deep learning with long short-term memory networks for financial market predictions”,2018; Lee J. , Koh H, Choe H.J. “Learning to trade in financial time series using high-frequency through wavelet transformation and deep reinforcement learning”, 2021; Ryo Akita, Akira Yoshihara, Takashi Matsubara, and Kuniaki Uehara. “Deep learning for stock prediction using numerical and textual information”,2016;

Despite a large number of scientific publications in this field, there are still unresolved problems associated with improving the accuracy of financial asset forecasting models.

2. Methods

Modern methods of forecasting time series are based mainly on the principle of forecasting the future based on historical data. A feature of the stock analysis indicators is the presence of multidirectional trends, cyclical and seasonal fluctuations that impose certain requirements on the choice of appropriate models and methods. The research focuses on classical methods of time series analysis: autoregressions and exponential smoothing models, machine learning and deep learning algorithms.

2.1 Autoregressive Approach

One of the fundamental methods used to analyze and predict time series is the autoregression model (AR)[1].

Autoregression is a statistical model based on the idea that the current value of a time series depends on its previous values. This means that a time series is treated as a sequence where each value depends on its predecessors. This approach allows you to take into account trends, cyclical patterns and other time dependencies, which makes autoregression a powerful tool for analyzing and predicting time series.

One of the powerful and widely used methods for this purpose is the "Autoregressive Moving Average" model (ARMA - Autoregressive Moving Average). This model combines two key components - autoregression (AR) and moving average (MA), which makes it a powerful tool for analyzing and predicting time series[2].

When modeling, we use the moving average, which is indicated below:

$$\hat{y}_t = \frac{1}{k} \sum_{n=0}^{k-1} y_{t-n} \quad (1)$$

Where, $\hat{y}_t$ is the smoothed value at the current time step; k number of periods (time steps) used to calculate the moving average; y_{t-n} value of the time series at the time step $t - n$, where n varies from 0 to $k - 1$; $\frac{1}{k}$ coefficient used to average the sum values of the time series[3].

Table 1. The results of the study depend on the size of the window.

$n/_{metrics}$	MAE	MSE	RMSE
5	11.6479931	343.5694157	18.53562558
10	19.865883	863.9208276	29.39253013
15	26.0491268	1375.031511	37.08141733
20	31.4700759	1869.492092	43.23762357
25	36.3518939	2377.061135	48.75511393

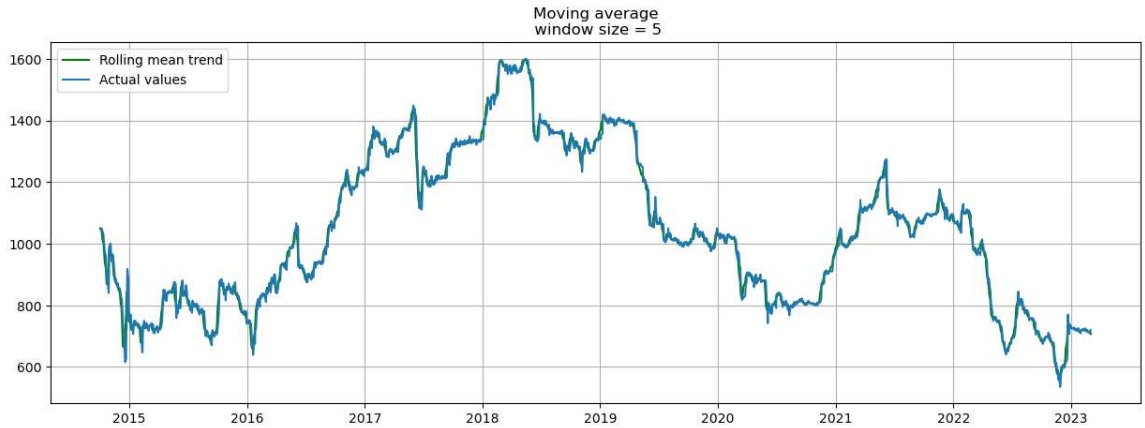


Figure 1. A time series smoothing graph with a window size of 5.

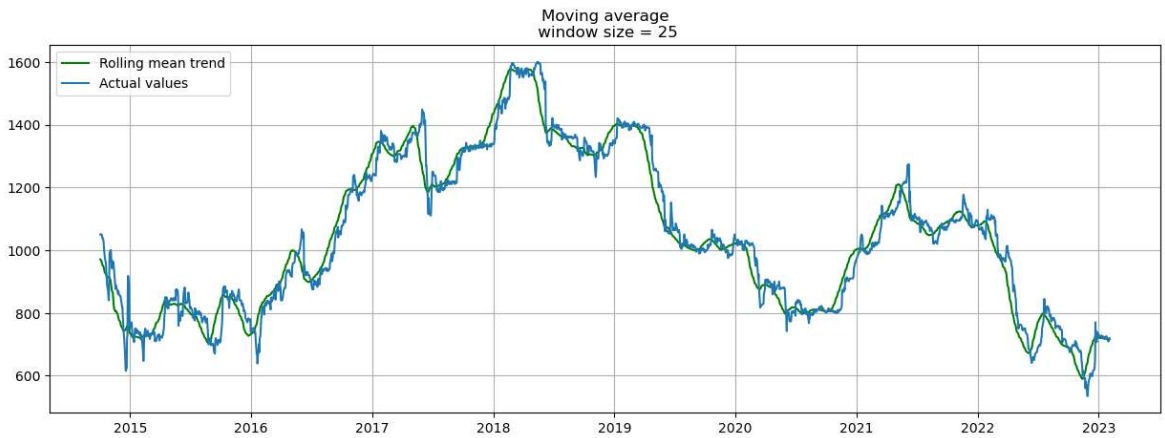


Figure 2. A time series smoothing graph with a window size of 25.

2.2 Exponential Smoothing Approach

Exponential smoothing is a statistical method of time series analysis that is used to predict future values based on historical data. It differs from the moving average (MA) in that it takes into account the weights assigned to different time periods. Newer data is assigned more weight, which makes the method more sensitive to changes and trends in the data [4].

Table 2

The results of the study depending on the exponential smoothing parameter.

$\alpha/_{metrics}$	MAE	MSE	RMSE
0.05	101.03710	15184.01544	123.22343
0.1	37.0173760	2329.95191	48.269575
0.15	24.6873177	1147.38220	33.87303
0.2	18.8477701	714.98775	26.73925
0.25	15.1941138	714.98775	22.155587
0,3	12.6410337	354.20128	18.820236
0.35	10.7002498	262.894825	16.214031
0.4	9.15509337	198.278010	14.081122
0.45	7.87241379	150.72900	12.277174
0.5	6.78337814	114.764847	10.712835

The basic idea of exponential smoothing is that the current value of a time series depends on its previous values, and this dependence decreases exponentially with increasing distance in time. The exponential smoothing formula looks like this:

$$\hat{y}_t = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{y}_{t-1} \quad (2)$$

Where ,y_t - сглаженное значение на текущем временном шаге; α (альфа) - smoothing parameter that determines the weight of the current value of the time series (y_t) and the weight of the previous smoothed value ($\hat{y}_{t-1}$); $\hat{y}_t$- value of the time series at the current time step.

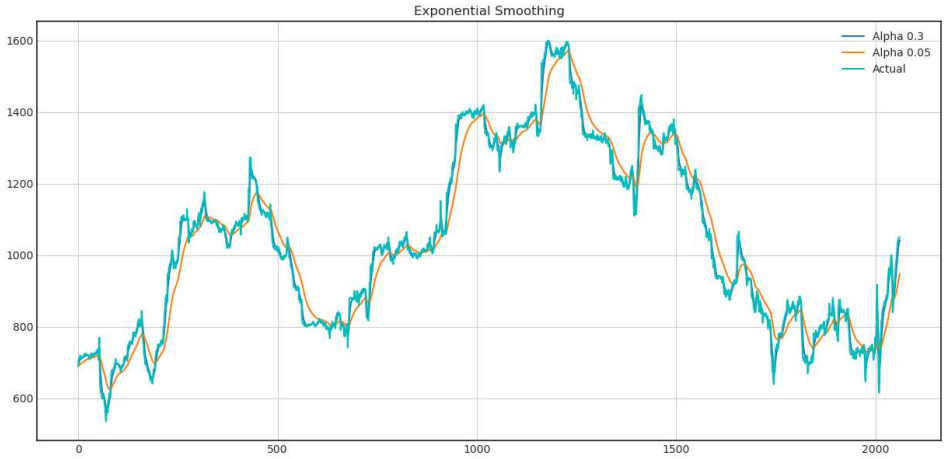


Figure3. Time series smoothing graph at values α=[0.3,0.05] .

2.2 Double Exponential Smoothing Approach

For more accurate and reliable forecasts, it is often necessary to take into account complex time dependencies, such as trend and seasonality. In such cases, the double exponential smoothing method is used.Double exponential smoothing, also known as the Holt-Winters method, is an extension of exponential smoothing that takes into account both trend and seasonality in time series[5].

Double exponential smoothing includes two components: exponential smoothing of the level (ESL - Exponential Smoothing of Level) and exponential smoothing of the trend (ETS - Exponential Smoothing of Trend).

Exponential Level Smoothing (ESL): This component takes into account changes in the level (the level is the average value in the data) and uses exponential smoothing to calculate weights for new values. This allows you to track and account for level changes in the data.

Double exponential smoothing includes two components:

$$l_x = \alpha \cdot y_x + (1 - \alpha) \cdot (l_{x-1} + b_{x-1}) \quad (3)$$

Где, l_x - smoothed level value at the current time step; α (alpha) - level smoothing parameter that determines the weight of the current value of the time series (y_x) and the weight of the previous smoothed level value ($l_{x-1} + b_{x-1}$); y_x- the value of the time series a row at the current time step.

Exponential Trend Smoothing (ETS): This component takes into account trend changes in the data, i.e. changes in the slope of the time series. It also uses exponential smoothing to calculate weights for the trend, which allows you to account for changes in the trend.

$$b_x = \beta \cdot (l_x - l_{x-1}) + (1 - \beta) \cdot b_{x-1} \quad (4)$$

where, b_x - smoothed trend value at the current time step; β (betta) - is the trend smoothing parameter, which determines the weight of the current level change ($l_x - l_{x-1}$) and the weight of the previous smoothed trend value (b_{x-1}).

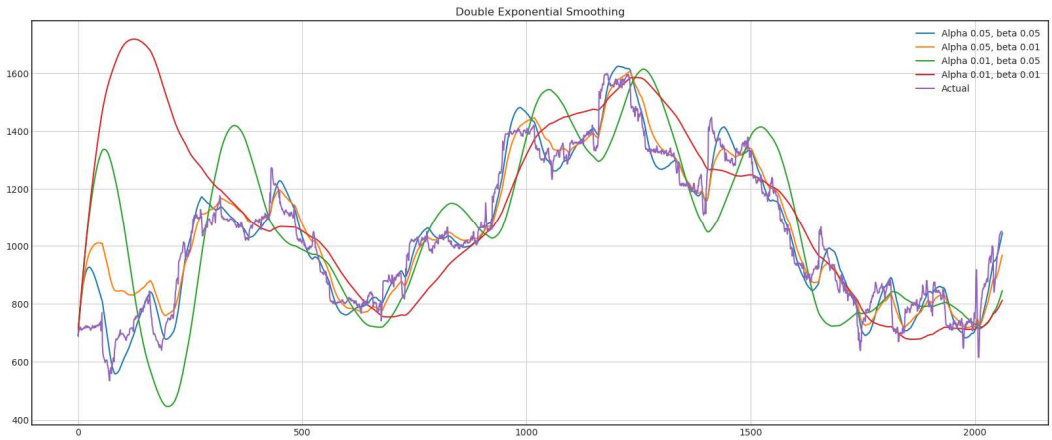


Figure 4. Time series smoothing graph at values α=[0.01,0.05]

Table 3. Results of the study of double exponential smoothing.

α/β	0.01	0.05
0.01	MSE: 87549.8069 RMSE: 295.888166 MAE: 171.95909657	MSE: 37158.8344 RMSE: 192.76626 MAE: 143.77284
0.05	MSE: 5986.8339093 RMSE: 77.374633500 MAE: 52.01418260	MSE: 3231.979846 RMSE: 56.8505043 MAE: 43.02253898

Conclusion

In this article, we investigated various methods of time series analysis, such as autoregression, exponential smoothing and double exponential smoothing, to identify the best approaches to time series forecasting. We analyzed the data and evaluated the effectiveness of each method based on the quality assessment metrics of the model mean absolute error (MAE): 11.6479931, mean square error (MSE): 343.5694157, root of mean square error (RMSE): 18.53562558. For the exponential smoothing model, the indicators were as follows, MAE: 6.78337814, MSE: 114.764847, RMSE: 10.712835.

Our results show that the exponential smoothing model demonstrates the best indicators of prediction accuracy compared to other methods. It has the smallest mean absolute error (MAE) and the smallest root of the mean quadratic error (RMSE), which makes it the preferred choice for predicting time series in this context. However, it should be noted that the choice of method depends on the specific characteristics and features of the data, as well as on the goals and requirements of the study. In some cases, an autoregression or double exponential smoothing model may be more appropriate, especially if the data has pronounced trends or seasonal components. The next stage of the research is the application of classical machine and deep learning methods for time series forecasting [5-13]. In general, the choice of the optimal method of time series analysis should be reasonable and based on the specification of the task and the characteristics of the data to ensure the best accuracy and quality of forecasting.

References

1. Adebiyi A Ariyo, Adewumi O Adewumi, and Charles K Ayo. Stock price prediction using the ARIMA model // In 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, IEEE, 2014. – P. 106–112.
2. Hyndman R., Koehler A., Ord K., Snyder, R. Forecasting with Exponential Smoothing // Springer Berlin Heidelberg, 2008.
3. Ayodele Ariyo Adebiyi, Aderemi Oluyinka Adewumi, and Charles Korede Ayo. Comparison of ARIMA and artificial neural networks models for stock price prediction // Journal of Applied Mathematics, 2014.
4. David MQ Nelson, Adriano CM Pereira, and Renato A de Oliveira. Stock market's price movement prediction with LSTM neural networks // In 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, 2017. – P. 1419-1426.
5. Gamboa, J.C.B. Deep learning for time-series analysis, CoRR abs/1701.01887 2017.
6. Hewamalage H., Bergmeir C., Bandara, K. Recurrent neural networks for time series forecasting: Current status and future directions // Int. Journal of Forecasting, 2020.
7. Tian Y., Pan, L. Predicting short-term traffic flow by long short-term memory recurrent neural network // Proceedings IEEE ISC2 2015 , pp. 153–158.

8. Fischer, T., Krauss, C. Deep learning with long short-term memory networks for financial market predictions // European Journal of Operational Research 270(2), 2018. – P. 654–669.
9. S. Smyl, A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting // Int. J. Forecast. 36(1), 2020. P. 75-85.
10. Deng, Y., Bao, F., Kong, Y., Ren, Z., Dai, Q. Deep direct reinforcement learning for financial signal representation and trading. // IEEE Transactions on Neural Networks and Learning Systems, 2017. – P. 653-664
11. Y. Wang, W. Liao and Y. Chang, Gated recurrent unit network-based short-term photovoltaic forecasting // Energies 11, 2018.
12. Lee J. , Koh H, Choe H.J. Learning to trade in financial time series using high-frequency through wavelet transformation and deep reinforcement learning. // APPLIED INTELLIGENCE, 2021 – P. 6202-6223.
13. Sagheer and M. Kotb, Time series forecasting of petroleum production using deep LSTM recurrent networks // Neurocomputing 323, 2019.

AUTOMATED HATE SPEECH CLASSIFICATION THROUGH "EMOTIONAL" COMMENTS OF NETWORK USERS

Toktarova A.B[1]., Omarov B.S[2]., Azhibekova Zh. Zh[3]., Yeshenkozhaev N.Sh[3].
E-mail: toktar.aigerim@list.ru

[1]*Khoja Akhmet Yassawi International Kazakh – Turkish University, Turkistan, Kazakhstan*
[2]*Al- Farabi Kazakh National University, Almaty, Kazakhstan*
[3]*Asfendiyarov Kazakh National Medical University, Almaty, Kazakhstan*

Abstract. *The purpose of this article is to analyze articles on cyberbullying of children and adolescents and to develop a framework for working on one's own project that is specifically relevant to Kazakh social media. The page offers illustrations of related works by researchers from different nations, including the United States, England, and other European nations. The examination of previous works on the subject was done using theoretical methods including analysis, synthesis, and empirical comparison and experimentation. This work was intended to analyze the issue in the Kazakh space. In a nutshell, we discuss the problem of building a parser and gathering data for teaching machine learning and deep learning algorithms that could discover and block sentences featuring a humiliating slope in real time.*

Keywords. *hate speech, cyberbullying, classification, machine learning, deep learning, social networks, and social media*

Introduction

Thanks to the Internet, access to and distribution of information is happening at an unprecedented speed. In addition, social media allows making a person famous through online content and it is a good medium for spreading opinions, some of which, are harmful to the Internet user or contain insults and discrimination [1]. In this regard, several scientific researchers have investigated the difficulty of automatically detecting this phenomenon [1], and they have found that there are several classes of actions for solving comparable problems (e.g. aggression [2] or hateful/offensive comments).

To address this issue, one tactic that can be used is the use of key patterns or phrases. Combining traditional artificial intelligence methods with goal-oriented activities is an idea that is gaining popularity. Kwok and Wang employed the Bag of Words (BoW) technique in conjunction with a Naive Bayes classifier to produce the desired outcomes [3].

However, other researchers (such as those that used SVM, naive Bayes, and logistic regression) have employed advanced feature representation techniques for machine learning algorithms [4]. It was determined that the high number of false positives was the reason for this. The experimenters suggested employing an SVM with a linear kernel and a variety of discrimination and non-discrimination classifications (for example, blame, insult, etc.).

To identify what slang phrases are, numerous scientific studies have been undertaken and published. The vast majority of studies on offensive terms have used English-language data because pertinent corpora are easily accessible. However, several studies have been carried out recently to identify hate speech in the Dutch, Indonesian, and Kazakh languages.

When there is a chance to identify offensive or insulting content, such as on discussion forum pages on contentious issues, keyword-based methods have occasionally shown to be more likely to uncover unpleasant language from individuals or from hostile web pages.

In other words, it done while taking into account the restrictions related to gathering information from a varied collection of people. To gather comments that contained rude and disparaging language, the researchers employed several different strategies. Depending on the source, inappropriate language messages may be present. [6]

Materials and methodology

Creating a corpus of Kazakh-language texts to train and test machine learning algorithms particularly for recognizing language-based comments on online resources. The research will initially construct a corpus of texts that do not display any indicators of hostility as well as texts that use hostile language. Extremism (sharing extremely negative opinions of political and religious orientation), radicalism (verbal criticism of the current political environment), cyberbullying (insulting or harassing a person on the Internet), discrimination (restriction of human rights through excessively obscene words), offensive language, and radicalism are all examples of this.

A variety of greetings, the usage of animal names in reference to animals, a string of remarks and messages devoid of any hint of intimidation or prejudice, numerous advertising spam, etc. are among the texts that fall under the category of texts without any indication of hatred.

At the start of the investigation, online databases were searched for documents with profane language. A volunteer annotator, or online searcher with a specific goal, gathered a string of offensive comments in the Kazakh language from the Instagram accounts of bloggers, singers, and other well-known individuals with a Kazakh-language audience. These comments contained texts with offensive words in the Kazakh language. The produced database, however, is a series of texts with just one, two, or three words per text. For example, the not long phrases "cattle", "like a pig", "you are donkey" were not enough to train the machine learning system.

For example, the meaning of the word "cattle" can be used in different contexts, it is difficult in machine learning to include or not include the word "cattle" in the sentences "cattle feed was delivered to the barn" and "you are a cattle person". Therefore, 10,224 sentences with obscene content consisting of three or more words were collected from Kazakh language communities and accounts on Facebook and Youtube platforms using online parsing and a volunteer annotator.

In the developed corpus, a document with obscene language texts consists of 3 columns: the serial number of the message or comment, the message or comment itself, and the attribute "1" indicating that the message or comment is obscene, as shown in Fig. 1. In the future, it is planned to further fill the corpus with obscene language texts.

data hate speech -kazakh 10k.xlsx

	A	B
10027	Қалмақ шалды пір тұтып жүргендер кезінде шалдың арқасында билікке, байлыққа қол жеткізгендер. Олардың : „Бізге егемендікті алып берген Назарбаев" дегенін естігенде есіңнен танып қала жаздайсың. Құлдық сана оңайшылықпен өзгермейді екен.	1
10028	Бұл адамның дінге ешқандай қатысы жоқ, мұсылмандар қауымдастығы деген қоғамды өзі ашып алған замозванец, немен айналысатынын ешкім білмейді, әйтеуір ұсталған екен оңбаған, тексеру керек қай жердің жансызы екенін	1
10029	Мына малдарға заң жоқпа күннен күнге өршіп баратыр,билік қайда қарап отыр бұларды ату жазасына берсе көрер едік ойларына келгенді жасағандарын.Сәбилерге өмір сүруге мүмкіндік бермеген қайюандар.	1
10030	Кішкентай қыздың өмірін құртқан педофилдерге ату жазасын беру керек.Неге Заң қатаң болмайды.Ату жазасын берсін.Өлтіру керек педофилдерді.Сонда педофил дегендер болмайды.	1
10031	Обалай ату керек ондаймағұлынды	1
10032	Жанболат бостандыққа шықсын . Назарбаевпен Тоқаев қамалсын !	1
10033	Шахмат деген ақыл ойдың ойыны емеспе, оған бәлендей жатықтырушыныңда қажеті шамалы сияқты, бейшара мен Мәскеуде тұрам дегенді дәреже көріп жүргенде, қазіргі саяси жағдайда Мәскеуіңе итте саруға жиіркенеді, орыстың қазаққа деген көзқарасыда жақсы емес, скинхэдтер былай тұрсын жәй орыстар оңаша кезігіп қалса айыратыны айдан анық, далдұлың шығып жезөкшелікке түсіпкетпесең болар әйтеуір.	1
10034	Кілең ұрылар отыр билікте,яғни бұндай қаражатты қайдан алды???..... Назарбаевтың ұры жүйесі қарапайым халықтың жүйкесін жұқартты	1
10035	Керісінше бұл сатқын қазақтың төлқұжатын өткізсін. Дені сау болса, алдымен неге қашып кетті? Қара басын қорғаштап қашты. Одан әрі құрысын. Қазақты ойламаған, Отанын сатқындардың жазасы ауыр	1
10036	Қазақттар сігіп жіберер ашуна тисек деген шығар 😂	1
10037	Провакатор группа сиактыгой бул тупой админ нацист	1
10038	Сенин ебальнигинен пил тепкен мени бузау тепсе	1
10039	Өңкей еріншек, жалқаулар ғой мына тізім жасаған прокурорлар. Жөнді тізім жарты жылда жасауға да жарымады	1
10040	Токаев қыра салыпты гой. Құдайынан корыкпаган бале	1

Fig 1. Example of profanity-laced comments in Kazakh language

Word cloud: Word clouds were used to visualize the data. User posts with potentially discriminatory or hateful ideas are highlighted separately in Fig 2. As we can see, hateful posts often use words like "slut", "bragging slave", and "should be shot", which directly indicate that they are intended to belittle, insult or threaten users.

a) word cloud of bad words

b)word cloud of neutral words

Fig 2. Word clouds for hate speech neutral words

The analysis of hate speech language texts in the Kazakh language showed that they have their own characteristics:

- Replacing the letters of the Kazakh language with Cyrillic letters
- Replacing the letters of the Kazakh letters with Latin letters
- Disparagement using similes or metaphors
- The use of insulting words based on the place of residence
- Spelling mistakes or writing using a symbol instead of a letter
- Creating new “Kazakh” words from the Russian language using additional suffixes

The database of obscene words includes content texts of the following types:

• Depending on the body shape of a person, based on the height or shortness of the person, there are expressions such as "short" and "tall giraffe man".

- By age (ageism) - it can be observed that it is often used to discriminate against the elderly and small children. For example, there are expressions such as "old woman", "old crow", "stupid cripple", and "look at your troubles like a crow".
- Caste discrimination - restricting or dividing the rights of people of a certain place of residence, or social groups based on their origin or small population, for example, you can find expressions such as "Russians", "idiots" for northerners, "hedgehogs" for southerners.
- Discrimination based on profession or field of work, disparagement of the owner of any profession, for example, singers are called "jumping goats" or doctors are called "they don't have and if you don't give them money", as well as phrases with the meaning of "they're idiots because they bought their diploma".
- Aylism or disablism - stigmatization of people based on their disability or physical characteristics can be attributed, for example, the group of words "handicapped", "disabled", and "monster" can be attributed to this section.
- Racism - discrimination of people based on ethnic origin and race. Racists try to prove that a person is different from others based on his appearance, color, behavior, and manners, and to prove that he is different by discrimination, for example, sentences consisting of the phrases "sart", "kafir", "pig", "arapkul" can be met.
- In discrimination based on religious beliefs, a person puts his religion above others, that is, he tries to degrade and humiliate those who follow other religions, as can be seen from the expressions "Islamists", "Arabs", "infidels".
- Sexism - separating people based on their gender and preferring one over the other. Can be linked to gender inequality. Sexism is prejudice against women and men based on their sex, not accepting their power and rights as equal, for example, phrases with offensive words such as "gays", "lesbians", and "you are bitch women".
- Homophobia or heterosexism is the denial of rights and the humiliation of people because of their sexual orientation.
- Discrimination and prejudice against someone who can only speak one language or cannot speak any language are both based on spoken language. [7-8]

Research results. As a result of the above work, a) text data was collected from the social network "Instagram" pages and groups of Kazakh language users on social networks "Youtube" and "Facebook" thanks to online parsing and annotator; b) manual classification of texts into two classes: "hate speech" and "neutral texts"; c) a corpus was created for further training of the model; d) analysis of the corpus was carried out using the distribution of hate speech and texts and word clouds, and the features of hate speech texts were determined.

Using deep learning to detect hate speech comments.

Deep learning is ideal for natural language processing (NLP) tasks such as emotion analysis, text classification, and automatic translation. We use the following deep learning techniques for text classification: Convolutional Neural Network (CNN) and LSTM Recurrent Neural Network. Deep learning is ideal for natural language processing (NLP) tasks such as sentiment analysis, text classification, and machine translation. For text classification, we use the following deep learning methods: Convolutional Neural Network (CNN) and LSTM Recurrent Neural Network. [9]

The following algorithm was used to filter the collected database:

- Converting all characters to one case and removing extra characters,
- Remove stop words,
- Hatching and lemmatization;
- Showing lexemation model (distribution of text into words - lexemes) and n-gram model of words (number of possible words in a lexeme).

Neural networks can only learn to find patterns in numerical data, so before feeding the text as input to the neural network, we converted each word to a numerical value. This process is called word encoding or tokenization.

Word embedding was used for tokenization, a technique that represents words as dense word vectors (also called word embedding). This word "input" means that it packs a lot of information into a small amount. Their purpose is to show semantic meaning in geometric space. This geometric space is called embedding space. This shows words that are similar in meaning close to the cell space, such as numbers or colors [10].

```
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_seq
uences
from keras.preprocessing import text, sequence

tokenizer = Tokenizer(num_words=20000)
tokenizer.fit_on_texts(list(X_train))

X_train = tokenizer.texts_to_sequences(X_train)
X_test  = tokenizer.texts_to_sequences(X_test)

X_train = sequence.pad_sequences(X_train, maxlen
=200)
X_test  = sequence.pad_sequences(X_test,  maxlen
=200)

print('X_train shape:', X_train.shape)
print('X_test shape: ', X_test.shape)
```

Two parameters are used in the tokenizer: num_words, which is responsible for setting the size of the dictionary, and pad_sequence(), which fills the sequence of words with zero, which parameter is used to solve the problem of words of different lengths in the text sequence. Parameter num_words is also added and is responsible for setting dictionary size. The word_count value is set to 2000. One of the problems of the research is that each text sequence is often different in word length. By adding the maxlen parameter we manage to specify how long sequences should be. This will truncate chains that exceed this number[11,13].

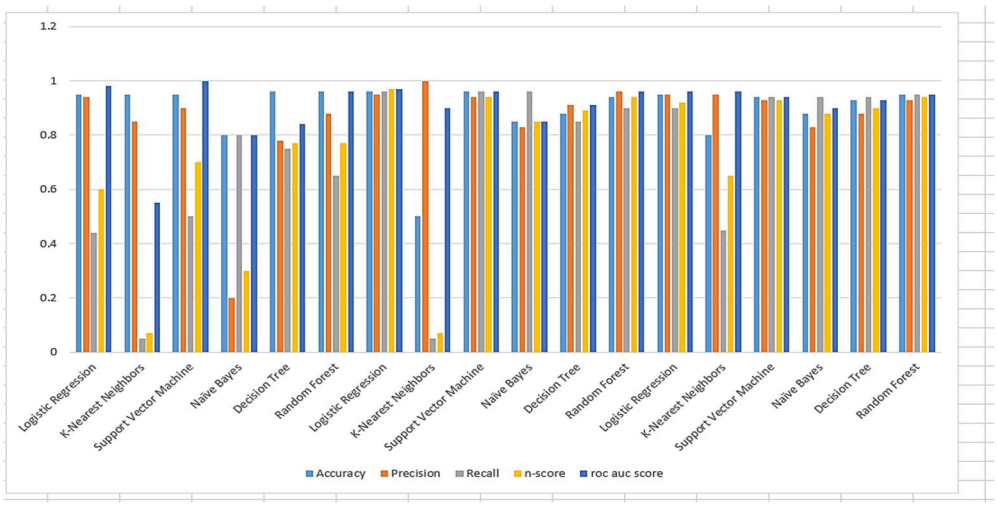


Fig3. Comparison of results of classification algorithms

Applying machine learning to detect hate speech in text. This section compares the outcomes of employing multiple machine learning algorithms to categorize hate speech using various feature combinations. The most popular techniques for creating and training classifiers in contemporary research include the ones listed below: k-nearest neighbors, logistic regression, naive Bayes, decision tree, random forest, and support vector machine in Fig3.

Table 1 displays the findings from a comparison of various approaches employing various features. The table illustrates how adding more functions to a method improves its performance overall. This observation attests to the usefulness and potency of the learned traits. The contribution of each attribute, however, varies widely, showing variations in the outcomes of different approaches.

Conduct comparisons between the algorithms:

The analysis and presentation of the authors' findings on this topic was all fairly strong [12-13]. The most often used metrics to assess the algorithms were:

$$precision = \frac{TP + FP}{TP + FP} \tag{1}$$

The ratio of true positives to all predictions is known as *precision* in line (1). It merely denotes the proportion of positive classifications the model properly detected.

$$recall = \frac{TP}{TP + FN} \tag{2}$$

Recall in line (2) the ratio between the total number of accurate forecasts and the sample space's accurate observations. It refers to the percentage of actual positives that were successfully identified.

$$F1 = \frac{2\,precision * recall}{precision + recall} \tag{3}$$

F stands for the F-measure *or F1-score* which in line (3) , which is used to evaluate how well the model performs when the class distribution is unbalanced. F1-score is a more intelligent metric to test a model since imbalanced class distribution happens in the majority of real-life text classification jobs.

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{4}$$

The ratio of a correct forecast to all observations is called *accuracy* which in line (4).In various and unbalanced data sets, accuracy is not always the best option; as a result, other evaluation metrics, such as the F1-score, may be taken into account

When determining if the model operates in a meaningful manner, those four markers are the most frequently utilized. [14-17]

Table 1.

Comparison of different methods using different features

Methods	properties	Acc.	Prec.	Recall	F1	AUC
	Statistical Features + TF-IDF	0.8204	0.2423	0.7593	0.3673	0.8622
	Statistical Features + TF-IDF + POS	0.8412	0.2512	0.6625	0.3643	0.8263
	Statistical Features + TF-IDF + POS + LIWC	0.1065	0.0641	0.8834	0.1196	0.5357
	Statistical Features + TF-IDF	0.9444	0.9529	0.201	0.332	0.6472
Decision tree	Statistical Features + TF-IDF + POS	0.9444	0.8969	0.2159	0.348	0.6395
	Statistical Features + TF-IDF + POS + LIWC	0.9444	0.8812	0.2208	0.3532	0.6274
	Statistical Features + TF-IDF	0.9368	1.0	0.0794	0.1471	0.9179
RF	Statistical Features + TF-IDF + POS	0.9369	1.0	0.0819	0.1514	0.9151
	Statistical Features + TF-IDF + POS + LIWC	0.9364	1.0	0.0744	0.1386	0.914
	Statistical Features + TF-IDF	0.9335	0.8421	0.0397	0.0758	0.5847
KNN	Statistical Features + TF-IDF + POS	0.9354	0.8158	0.0769	0.1406	0.6105
	Statistical Features + TF-IDF + POS + LIWC	0.9351	0.7037	0.0943	0.1663	0.701
	Statistical Features + TF-IDF	0.9681	0.8942	0.6079	0.7238	0.9739
naive Bayes	Statistical Features + TF-IDF + POS	0.9625	0.806	0.598	0.6866	0.9687
	Statistical Features + TF-IDF + POS + LIWC	0.9543	0.7304	0.531	0.6149	0.9599
	Statistical Features + TF-IDF	0.9601	0.9568	0.4392	0.602	0.9759
logistic regression	Statistical Features + TF-IDF + POS	0.9598	0.9418	0.4417	0.6014	0.9759
	Statistical Features + TF-IDF + POS + LIWC	0.9409	0.6647	0.2804	0.3944	0.9336

discussions

Thus, a text recognition method related to hate speech in social networks has been proposed and explained in this study. It involves several steps, including cleaning the data using various techniques, data extraction, n-gram detection, text tokenization, stop word removal, and the resulting preprocessed texts.

The language Kazakh sentence corpus is then split into the training dataset and the test set. 10224 sentences make up the dataset that is being presented. 2556 (or 25%) are for model verification, while 7683 (or 75% of them) are for model training.

Conclusion

Therefore, assessing Online material to determine the presence or absence of a destructive orientation plays a crucial role in protecting the information security of the Internet information space. To look for various forms of damaging information, multiple methods should be employed. Using a variety of intelligent data processing techniques in concert could be a solution to this issue, increasing the effectiveness of the system that is currently being

developed. The proposed combined methodology can be used to identify damaging information in a promising and complementary way. The application of this method enables you to reduce the number of calculations by combining them, increase the visibility and interconnectedness of the outcomes of the analysis of specific problems, and thereby increase the efficiency of the analysis of dangerous content by increasing the validity and efficiency of management decisions.

The system's capacity to automatically learn new knowledge, which consists of the dictionary's self-replenishment, is another element of the proposed technique. The research led to the acquisition of the data, whose analysis allowed us to draw the conclusion that the proposed method can be used as a foundation for the creation of a sophisticated automated system for recognizing damaging information.

.

References

1. Gröndahl, T., Pajola, L., Juuti, M., Conti, M., & Asokan, N. (2018, January). All you need is" love" evading hate speech detection. In Proceedings of the 11th ACM workshop on artificial intelligence and security (pp. 2-12).
1. Rohmawati, U. A. N., Sihwi, S. W., & Cahyani, D. E. (2018, November). SEMAR: An interface for Indonesian hate speech detection using machine learning. In 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) (pp. 646-651). IEEE.
2. TOKTAROVA, A., IZTAEV, Z., KOZHABEKOVA, P., SUIEUOVA, N., OPONDO, R. O., KERIMBEKOV, M., & ZHUNISBEKOVA, Z. (2022). AUTOMATED HATE SPEECH CLASSIFICATION USING EMOTION ANALYSIS IN SOCIAL MEDIA USER GENERATED TEXTS. Journal of Theoretical and Applied Information Technology, 100(22).
3. Omar, A., Mahmoud, T. M., & Abd-El-Hafeez, T. (2020, April). Comparative performance of machine learning and deep learning algorithms for Arabic hate speech detection in osns. In The International Conference on Artificial Intelligence and Computer Vision (pp. 247-257). Springer, Cham.
4. Arango, A., Pérez, J., & Poblete, B. (2020). Hate speech detection is not as easy as you may think: A closer look at model validation (extended version). Information Systems, 101584.
5. Sultan, D., Suliman, A., Toktarova, A., Omarov, B., Mamikov, S., & Beissenova, G. (2021, January). Cyberbullying Detection and Prevention: Data Mining in Social Media. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 338-342). IEEE.
6. Nayel, H. A., & Shashirekha, H. L. (2019, December). DEEP at HASOC2019: A Machine Learning Framework for Hate Speech and Offensive Language Detection. In FIRE (Working Notes) (pp. 336-343).
7. Sultan, D., Toktarova, A., Zhumadillayeva, A., Aldeshov, S., Mussiraliyeva, S., Beissenova, G., ... & Imanbayeva, A. (2023). Cyberbullying-related hate speech detection using shallow-to-deep learning. CMC-COMPUTERS MATERIALS & CONTINUA, 74(1), 2115-2131.
8. Ibrohim, M. O., & Budi, I. (2019, August). Multi-label hate speech and abusive language detection in Indonesian twitter. In Proceedings of the Third Workshop on Abusive Language Online (pp. 46-57).
9. Toktarova, A., Beissenova, G., Kozhabekova, P., Makhanova, Z., Tulegenova, B., Rakhymbek, N., ... & Azhibekova, Z. (2021). Automatic offensive language detection in online user generated contents. Journal of Theoretical and Applied Information Technology, 99(9), 2054-2067.

10. K. Kumari, J. P. Singh, Y. K. Dwivedi and N. P. Rana, “Bilingual cyber-aggression detection on social media using LSTM autoencoder,” Soft Computing, vol. 25, no. 14, pp. 8999–9012, 2021
11. Agrawal, S., & Awekar, A. (2018, March). Deep learning for detecting cyberbullying across multiple social media platforms. In *Advances in Information Retrieval: 40th European Conference on IR Research, ECIR 2018, Grenoble, France, March 26-29, 2018, Proceedings* (pp. 141-153). Cham: Springer International Publishing.
12. Kim, K. H., Cho, Y. H., & Bae, J. A. (2020). Exploratory Study on Countering Internet Hate Speech: Focusing on Case Study of Exposure to Internet Hate Speech and Experts' in-depth Interview. *The Journal of the Korea Contents Association*, *20*(2), 499-510.
13. Biere, S., Bhulai, S., & Analytics, M. B. (2018). Hate speech detection using natural language processing techniques. *Master Business AnalyticsDepartment of Mathematics Faculty of Science*.
14. Barron, J. A. (2019). Internet Access, Hate Speech and the First Amendment. *First Amend. L. Rev.*, *18*, 1.
15. Claussen, V. (2018). Fighting hate speech and fake news. The Network Enforcement Act (NetzDG) in Germany in the context of European legislation. *Media Laws*, *3*(3), 110-136.
16. Izquierdo, A., Laforgue, N., & Abril, D. Hate speech: a systematic review of scientific production and educational considerations. *Fuentes, 2022, vol. 24, num. 2, p. 222-233*.

РАЗРАБОТКА СИСТЕМЫ АВТОМАТИЧЕСКОГО СЛЕЖЕНИЯ ЗА ПОЛОЖЕНИЕМ СОЛНЦА

Исаева Г. Қ., Карымсакова Н. Т.
e-mail: gauhar.isaeva18@gmail.com, nkarymsakova1@gmail.com
Казахский национальный университет имени аль-Фараби, Казахстан

Аннотация. *Эта статья представляет собой разработку системы, которая позволяет точно определять положение Солнца и управлять движением солнечного трекера для максимальной эффективности использования солнечной энергии. Важность данной темы в современном мире не может быть преувеличена, так как солнечная энергия является одним из наиболее перспективных источников возобновляемой энергии.*

Также были рассмотрены принципы работы солнечных батарей, которые позволяют преобразовывать солнечную энергию в электрическую, и сенсоров для определения положения Солнца. Также были описаны принципы работы солнечных элементов и их роли в системе. Для эффективной работы солнечного трекера необходимо точно определять положение Солнца на небе, и для этого используются специальные сенсоры, которые обеспечивают высокую точность определения.

Ключевые слова: *система слежения, солнечный трекер, автоматическое управление, робототехника.*

Введение. Солнечная энергия-один из самых перспективных источников энергии в современном мире. Он относится к категории возобновляемых источников энергии, что означает, что его использование не приводит к истощению природных ресурсов и не наносит вреда окружающей среде.

Солнечная энергия имеет много преимуществ перед другими источниками энергии. Это бесплатно, безопасно, экологически безопасно и доступно во многих регионах мира.

В современном мире солнечная энергия используется для производства электроэнергии, нагрева воды и помещений, а также для обогрева и освещения открытых пространств. Солнечные панели устанавливаются на крышах зданий, улицах, пустырях и других местах для получения солнечной энергии.

Системы наблюдения за Солнцем. Для получения солнечной энергии нам необходима система наблюдения за Солнцем. Солнечная энергия, астрономия, исследования космоса-вот некоторые области, в которых используются системы наблюдения за солнцем. В этом обзоре мы рассмотрели несколько из них.

• Механические трекеры основаны на принципах механики и используются для отслеживания движения Солнца по небу. Они обладают низкой надежностью, высокой точностью и низкими затратами на техническое обслуживание, но могут быть дорогими в производстве и иметь ограниченный диапазон.

• Фотоэлементты-это электронный прибор, который преобразует энергию фотонов в электрическую энергию и в системе они используются для отслеживания движения

Солнца по небу. Они могут быть основаны на одном или нескольких фотоэлементах, точность которых может быть ограничена такими факторами, как облачность.

• Системы мониторинга программного обеспечения используют математические модели и алгоритмы для отслеживания движения Солнца по небу. Они обладают высокой точностью и масштабируемостью, но требуют значительной вычислительной мощности и могут быть трудными в использовании.

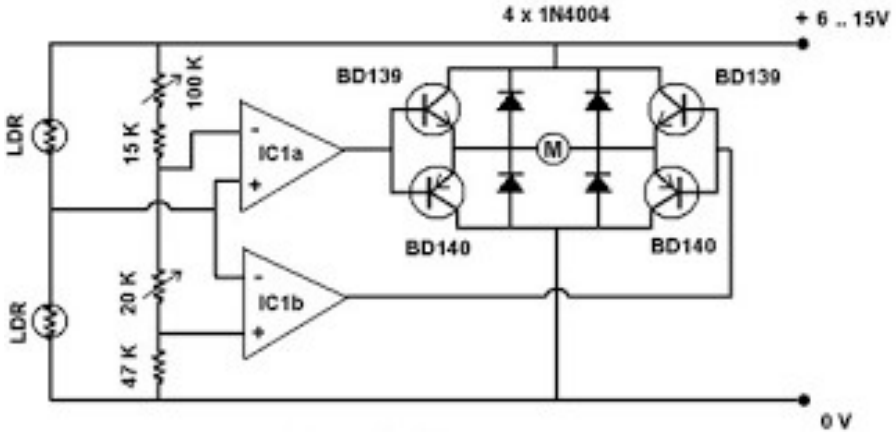


Рис.1 Схема устройства контроля положения Солнца

На рис.1 представлена распространенная схема устройства контроля положения Солнца.[3]

Элементы в схеме:

1. Фоторезисторы (LDR и LDR'): это полупроводниковые устройства, сопротивление которых варьируется в зависимости от интенсивности падающего на них света. В этой схеме фоторезисторы действуют как датчики, регистрирующие изменения интенсивности света.

2. Операционный усилитель IC1 (ou): это электронный инструмент, используемый для усиления, фильтрации и обработки сигналов. В этой схеме ОС используется для усиления и обработки сигналов от фоторезисторов ldr и ldr.

3. Компаратор (IC2): это электронное устройство, которое сравнивает два входных сигнала и выдает на выходе один из двух уровней напряжения, в зависимости от того, большой или маленький сигнал от другого. В этой схеме компаратор используется для сравнения двух выходных сигналов ou IC1.

4. Резисторы: это элементы, которые ограничивают ток в цепи и создают определенные уровни напряжения. В этой схеме резисторы используются для установления пределов значений напряжения на выходе ОС, а также для регулирования точки равновесия и чувствительности устройства.

5. Подстроечные резисторы: это специальные резисторы, которые можно отрегулировать до определенного значения сопротивления с помощью специального винта. В этой схеме тримпоты используются для регулировки точки равновесия и чувствительности устройства. Подстроечный двигатель 20K позволяет регулировать диапазон изменения напряжения, с которого запускается двигатель. Тримпот 100K позволяет регулировать точку равновесия, то есть значение напряжения, при котором двигатель не работает.

6. Двигатель: это электродвигатель, который начинает вращаться при поступлении сигнала от компаратора IC2. В этой схеме двигатель используется для обеспечения движения устройства при изменении положения источника света.

Исходя из расмотренных вариантов системы наблюдения в данной работе мы будем использовать солнечные трекеры.

Солнечные трекеры – это устройства, которые автоматически устанавливают солнечные панели в оптимальное положение, обеспечивая максимальное освещение на протяжении всего дня. Эти устройства используют различные технологии и методы для отслеживания движения Солнца и подстройки положения панелей. [3]

Основные компоненты трекера:

- Датчик положения Солнца-это устройство, которое определяет точное положение Солнца на небе. Обычно это фотоэлемент, который использует свет от солнца для определения его положения.
- Двигатель – это устройство, используемое для поворота солнечных панелей или других устройств в направлении Солнца. Он может работать на электрической или механической основе.
- Контроллер – это устройство, которое управляет работой трекера, управляя двигателем, используя информацию, полученную от датчика положения Солнца. Он может быть запрограммирован или работать при определенных условиях.

Трекеры бывают разных типов и дизайнов, но все они работают по принципу отслеживания Солнца и максимальной ориентации устройств на него.

Типы трекеров:

- одноосные трекеры – отслеживают движение Солнца по одной оси и всегда направляют солнечный модуль к Солнцу.
- двухосные трекеры – отслеживают движение Солнца по двум осям и поворачивают солнечный модуль так, чтобы он всегда был направлен к Солнцу и находился в правильной плоскости.
- и полярные трекеры – отслеживают движение Солнца относительно Земли и поворачивают солнечный модуль так, чтобы он всегда был направлен к Солнцу и находился в правильной плоскости, а также изменял угол наклона, чтобы учесть изменение высоты солнца в разное время года.

В качестве объекта исследовании рассмотрим одноосный трекер, также добавим к нему качающийся эффект. [1]

Одноосный трекер с качающимся эффектом. Одноосный трекер с эффектом качания – это устройство, которое отслеживает положение Солнца в небе и вращает солнечные панели таким образом, чтобы они всегда были направлены прямо к Солнцу.

В отличие от статических солнечных панелей, которые установлены в фиксированное положение и могут быть направлены только на юг в странах Северного полушария или на север в странах Южного полушария, трекеры отслеживают движение Солнца по небу в течение дня, его угол относительно горизонта постоянно меняется, что приводит к изменению угла падения солнечного света на солнечные панели.

Расчеты координат трекера. Для начала сделаем расчеты координат, чтобы определить нахождение Солнца на небе. [1]

Для определения положения Солнца в городе Алматы можно использовать следующие координатные расчеты:

Расчет широты и долготы города Алматы:

Широта: 43.238949° С. Ш. Долгота: 76.889709° В. Д.

Расчет угла наклона оси Земли:

$$\varepsilon = 23.45° \cdot sin(360/365 \cdot (284 + n)) \quad (1)$$

где ε – угол наклона земной оси в данном месте и в определенное время;

n – количество дней с начала года.

Используя (1) формулу 25 ноября 2023 года дата года равна 298, поэтому:

$$\varepsilon = 23.45° \cdot sin(360/365 \cdot (284 + 298)) \approx -0.48894°$$

Расчет угла отклонения орбиты Солнца:

$$\delta = arcsin(sin(\varepsilon) \cdot sin(360/365 \cdot (n - 81))) \quad (2)$$

где δ – угол отклонения орбиты Солнца в данном месте и в определенное время.

Расчет угла отклонения орбиты Солнца с использованием (2) формулы для 25 ноября:

$$\delta = arcsin(sin(\varepsilon) \cdot sin(360/365 \cdot (298 - 81))) \approx -0.50358°$$

Расчет времени восхода и захода солнца:

$$t = 12 \cdot (1 - 4/60 \cdot (76.8512 - lng) - equation_of_time/60) \quad (3)$$

где t – время восхода или захода солнца в определенном месте и в определенное время;

lng – долгота земли;

equation_of_time – уравнение времени.

Расчет времени восхода и захода солнца

$$t = 12 \cdot \left(1 - \frac{4}{60} \cdot (76.8512 - lng) - \frac{equation_of_time}{60}\right) \approx 12 \text{ часов}$$

Расчет высоты солнца над горизонтом:

$$h = arcsin(sin(\varphi) \cdot sin(\delta) + cos(\varphi) \cdot cos(\delta) * cos(15 \cdot (t - 12))) \quad (4)$$

где h – высота солнца над горизонтом в определенном месте и в определенное время;

φ – широта места;

δ – угол отклонения солнечного света;

ω – часовой угол солнца.

Вычисляем высоту солнца над горизонтом и устанавливаем значения.Тогда мы можем вывести значение.

$$h = \arcsin(\sin(43.238949°) \cdot \sin(-0.50358°) + \cos(43.238949°) \cdot \cos(-0.50358°) \cdot cos(15 \cdot (12 - 12))) \approx -0.50485° h \approx -0.50485°$$

Расчет солнечного Азимута:

$$Az = acos((sin(\delta) - sin(\varphi) \cdot cos(h))/(cos(\varphi) \cdot sin(h))) \quad (5)$$

где Az – Азимут солнца в определенном месте и в определенное время.

$$Az = arccos\left(\frac{\sin(-0.50358°) - \sin(43.238949°) \cdot \cos(-0.50485}{\cos(43.238949°) \cdot \sin(-0.50485°)}\right) \approx 179.71817°$$

Эти вычисления можно выполнять с помощью программы Python, которая позволяет автоматизировать процесс и получать точные данные в режиме реального времени.

Часовой угол дня можно рассчитать по формуле:

$$\omega = 15 \cdot (\text{местное время} - 12) \quad (6)$$

Где местное время находится в часах относительно местного стандартного времени. Местное стандартное время для города Алматы соответствует часовому поясу GMT+6.

Таким образом, для определения высоты солнца над горизонтом 25 ноября 2023 года в городе Алматы необходимо рассчитать часовой угол Солнца по формуле:

$$\omega = 15 \cdot (\text{местное время} - 12) = 15 \cdot (12 - 12) = 0°$$

Эти вычисления используются в коде преобразования солнечной энергии в электрическую с помощью программы Python.

Экспериментальный анализ с помощью программы Python показан в виде таблицы 1. и алгоритм принципа работы на рис. 2

Таблица 1

Результаты экспериментов в среде Python

Время суток	Угол наклона	Азимут	Энергия
6:00	42.22	97.99	26.92
9:00	47.34	114.22	324.12
12:00	49.34	128.69	466.27
15:00	47.34	243.77	327.50
18:00	42.22	262.00	23.21

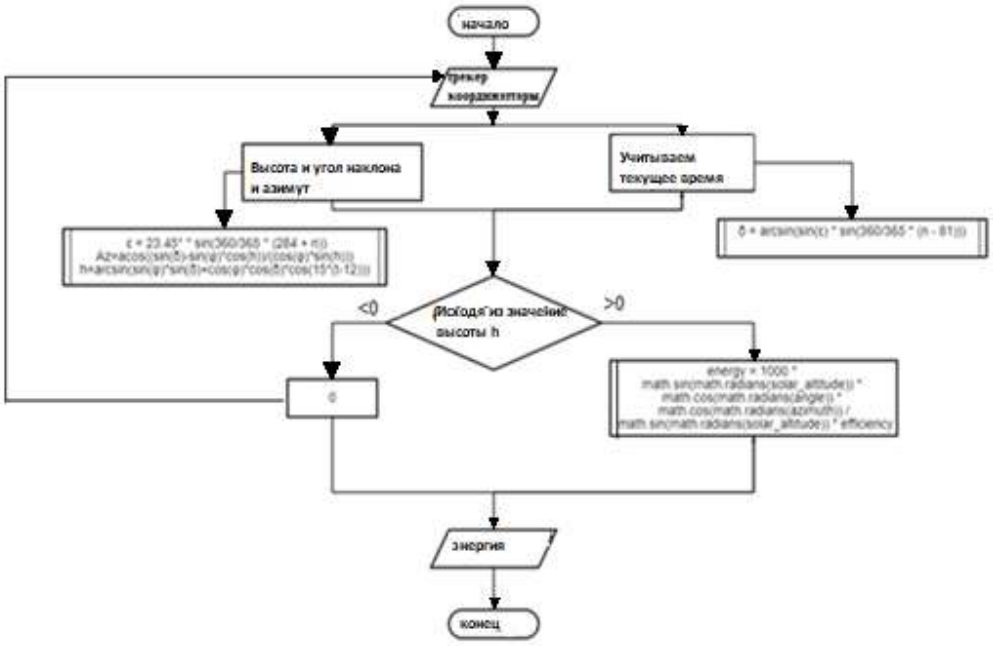


Рис. 2 Алгоритм принципа работы на программе Python

Схема одноосного солнечного трекера на основе программы Tinkercad показана на рис. 3.Используемые компоненты установлены в таблице 2.

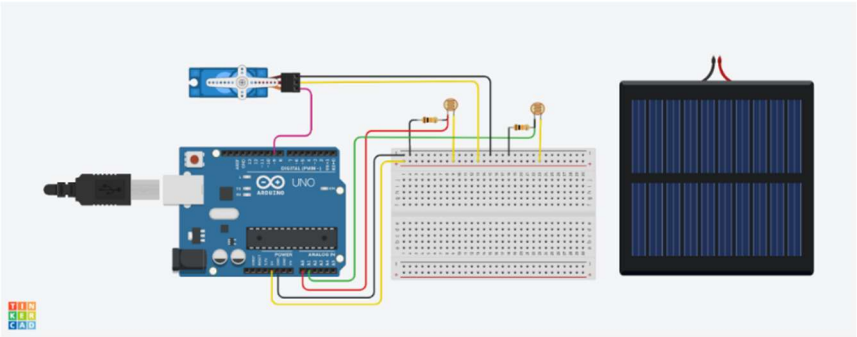


Рис. 3 Схема подключения одноосного солнечного трекера

Таблица 2.

Список компонентов

Название	Количество	Компонент
Servo1	1	Позиционный сервопривод
U1	1	Arduino Uno
R1 R2	2	10 кОм Резистор
R3 R4	2	Фоторезистор
SC1	2	5V,100 mA Солнечная батарея

Программный код представлен на рис. 4 и блок-схема данной работы на рис. 4.[2]

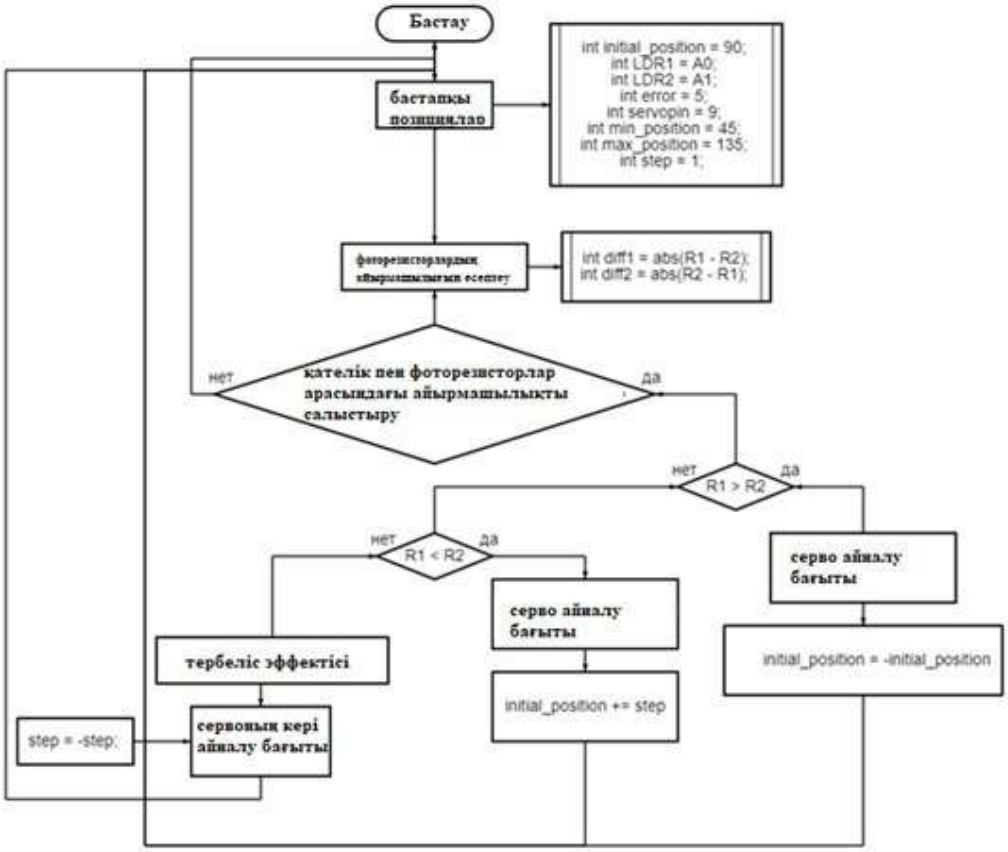


Рис. 4 Блок-схема работы трекера

Заключение. В ходе нашего исследования были исследованы и рассмотрены ключевые аспекты солнечных трекеров и их компонентов, а также принципы их работы.

Интеграция программных компонентов, основанных на языке Python, в систему управления является критически важным шагом для оптимизации работы солнечных трекеров и повышения их эффективности.

Для более наглядного моделирования и тестирования одноосного солнечного трекера с качающимся эффектом мы воспользовались программой Tinkercad. Этот инструмент позволил создать трехмерную модель трекера с учетом его основных компонентов и качающегося эффекта. Это важный шаг для оптимизации работы солнечных трекеров перед физической реализацией.

Наше исследование подчеркивает важность солнечных трекеров в сфере альтернативной энергетики, а также их потенциал для повышения эффективности использования солнечной энергии. Благодаря интеграции программных компонентов и тщательному моделированию, мы можем достичь более устойчивого и экологически чистого производства энергии, соответствуя требованиям современного мира.

Литература

1. Tilt and azimuth angles in solar energy applications – a review / A. Z. Hafeza, A. Solimana, K. A. El-Metwallya, I. M. Ismaila // Renewable and Sustainable Energy Reviews. – 2017. – V. 77. – P. 147–168.
2. Ревич Ю. В. Р32 Практическое программирование микроконтроллеров Atmel AVR на

языке ассемблера. – 2-е изд., испр. – СПб.: БХВ-Петербург, 2011. — 352 с.: ил. – (Электроника)

3. ГОСТ Р 57229–2016. Системы фотоэлектрические. Устройства слежения за Солнцем. Технические условия. – М.: Стандарт информ, 2016. – 64 с.
4. С. У. Исмаилов Тіректі-айналмалы механизмді күн батареясының жұмысын басқаруға арналған Heliostat_and_aktuator_01 бағдарламалық қамтамасыздандырудың мүмкін-діктеріне шолу жасау Вестник КазНТУ №3, Алматы, 2015. стр. 222–226 ISBN 1680-9211

РАЗРАБОТКА КОГНИТИВНОГО ЧЕЛОВЕКО-МАШИННОГО ИНТЕРФЕЙСА ДЛЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ СЛОЖНЫМИ ОБЪЕКТАМИ И ДИАГНОСТИКИ СОСТОЯНИЯ ОБОРУДОВАНИЯ

[1]Самигулина Г. А., [2]Самигулина З. И.
[1]e-mail: galinasamigulina@mail.ru
[2]e-mail: zarinasamigulina@mail.ru
[1]Институт информационных и вычислительных технологий КН МН РК, Казахстан
[2]Казахстанско-Британский Технический Университет

Аннотация. *Исследования посвящены актуальной проблеме интеллектуализации промышленного производства на основе создания высокоэффективного когнитивного человеко-машинного интерфейса (HMI) для управления и диагностики оборудования в нефтегазовой отрасли. Разработаны теоретические основы по дизайну когнитивного HMI дисплея для эффективного управления сложными объектами и диагностики промышленного оборудования на основе унифицированной искусственной иммунной системы (УИИС) с целью снижения когнитивной нагрузки на оператора. Результаты моделирования получены на реальных данных нефтеперерабатывающего предприятия ТОО «ТенгизШевроил» с установки D-301, предназначенной для очистки газа высокого давления. Пример когнитивной мнемосхемы реализован с помощью программного продукта Vijeo Designer (Schneider Electric).*

Стремительное развитие в последние годы промышленного искусственного интеллекта, новейших информационных технологий и цифровизация производства предъявляют особые повышенные требования к вопросам человеко-машинного взаимодействия и снижения когнитивной нагрузки (внутренней и внешней) на оператора, работающего на производстве. Когнитивная нагрузка — это количество информации, которое может обработать краткосрочная память человека. Актуально и остро стоят вопросы эффективного восприятия человеком огромного потока информации, который продуцируются современной микропроцессорной техникой и его способности оперативно принимать решения для предотвращения и ликвидации аварийных ситуаций. Внутренняя когнитивная нагрузка зависит от сложности информации и опыта оператора, внешняя когнитивная нагрузка возникает за счёт представления и формы подачи данных. Большое значение имеет способ представления информации и создание адаптивных интерфейсов. Например, исследования [1] посвящены разработке когнитивной модели пользователя, основанной на психофизиологических характеристиках при проектировании графических интерфейсов для улучшения человеко-машинного взаимодействия в информационных системах. Проведённые эксперименты с моделью продемонстрировали снижение избыточности интерфейса и повышение эффективности в работе информационной системы. В статье [2] рассматриваются вопросы создания сложных адаптивных интерфейсов на основе использования психологических характеристик оператора для повышения когнитивной гибкости и уменьшения негативных последствий

при длительной напряженной работе, а также повышения эффективности и безопасности. С помощью оценок машинного обучения и с учётом состояния оператора осуществляется адаптация дизайна интерфейса путем изменения информации, модальности представления, значимости стимулов и др. Обсуждается актуальность и перспективность данного направления для будущего развития эффективного человеко-машинного взаимодействия.

Однако, особенно трудная задача стоит по снижению внутренней когнитивной нагрузки на человека. Это связано, прежде всего, с необходимостью наличия высокой квалификации оператора в области, в которой он работает. Подготовка опытного оператора занимает около двух лет и не всегда на производстве это условие выполняется. Оператор, не имеющий соответствующей подготовки, испытывает сильную внутреннюю когнитивную нагрузку. Эта особенность требует более тщательного подхода при проектировании когнитивных HMI с целью максимального упрощения интерфейса, но с сохранением максимальной информативности. Более того, применение интеллектуальных подходов для управления сложными современными автоматизированными производственными линиями и разработка когнитивных мнемосхем на их основе также вносит свои особенности при проектировании HMI.

Предлагаемые исследования являются продолжением большого цикла работ по разработке когнитивных мнемосхем с помощью распределённой системы управления Experion PKS фирмы Honeywell для нефтегазовой отрасли на основе подхода искусственных иммунных систем. Актуальность данного вопроса связана с тем, что по статистике более пятидесяти процентов аварийных ситуаций на промышленном производстве происходит из-за «человеческого фактора». Длительная работа с мнемосхемами для управления технологическими процессами вызывает утомляемость и потерю концентрации внимания оператора. Рассматривались вопросы снижения *внешней* когнитивной нагрузки на пользователей [3] за счёт подачи и формы представления информации (расположения на дисплее монитора, цветопередачи и т.д.).

В данной работе *постановка задачи формулируется следующим образом*: необходимо разработать когнитивный HMI для интеллектуальной системы управления и диагностики оборудования на основе применения унифицированной искусственной иммунной системы с целью снижения *внутренней* когнитивной нагрузки на оператора и повышения эффективности производственного процесса для объектов нефтегазовой отрасли.

Результаты моделирования когнитивного HMI дисплея получены на реальных данных нефтеперерабатывающего предприятии ТОО «ТенгизШевроил» с установки D-301, предназначенной для очистки газа высокого давления. Пример когнитивной мнемосхемы реализован с помощью программного продукта Vijeo Designer Basic для оборудования компании Schneider Electric [4].

Структурная схема взаимодействия оборудования и программного обеспечения для управления сложными объектами представлена на рисунке 1. Контроль за процессом осуществляется с помощью программируемого логического контроллера (ПЛК), например Modicon M580/340/241 в зависимости от сложности задач. Программа для управления контроллером создается в пакетах EcoStructure Control Expert или EcoStructure Machine Expert. Контроллер под управлением данных программ взаимодействует с панелью оператора, например серии Harmony или Magelis.

Рассмотрим пример создания когнитивного дисплея для контура регулирования колонной D-301. Создано три диалоговых окна: «Процесс», «ПИД-регулятор», «Диагностика». На рисунке 2 представлен классический HMI дисплей (вкладка «Процесс») для

управления технологическим процессом на Установке 300/800 по очистке газа высокого давления D-301. Стиль оформления мнемосхемы выбран «по умолчанию». Мнемосхема состоит из объекта управления, размещённого с левой стороны, кнопочных элементов и индикаторов с правой стороны. Кнопки перемещения между экранами представлены в верхней части диалогового окна. Реализована кнопка «Выбор типа мнемосхемы», позволяющая изменить ее *внешнюю* когнитивную нагрузку на оператора в рамках исследований [3].

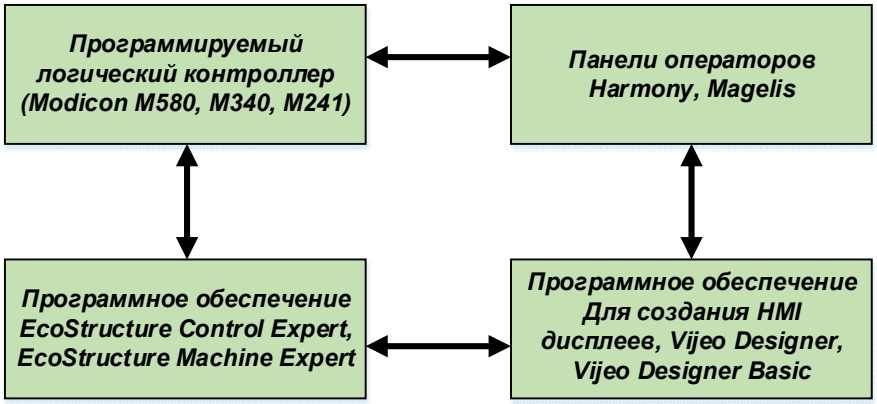


Рис. 1 Структура взаимодействия программно-аппаратных составляющих системы управления технологическим процессом

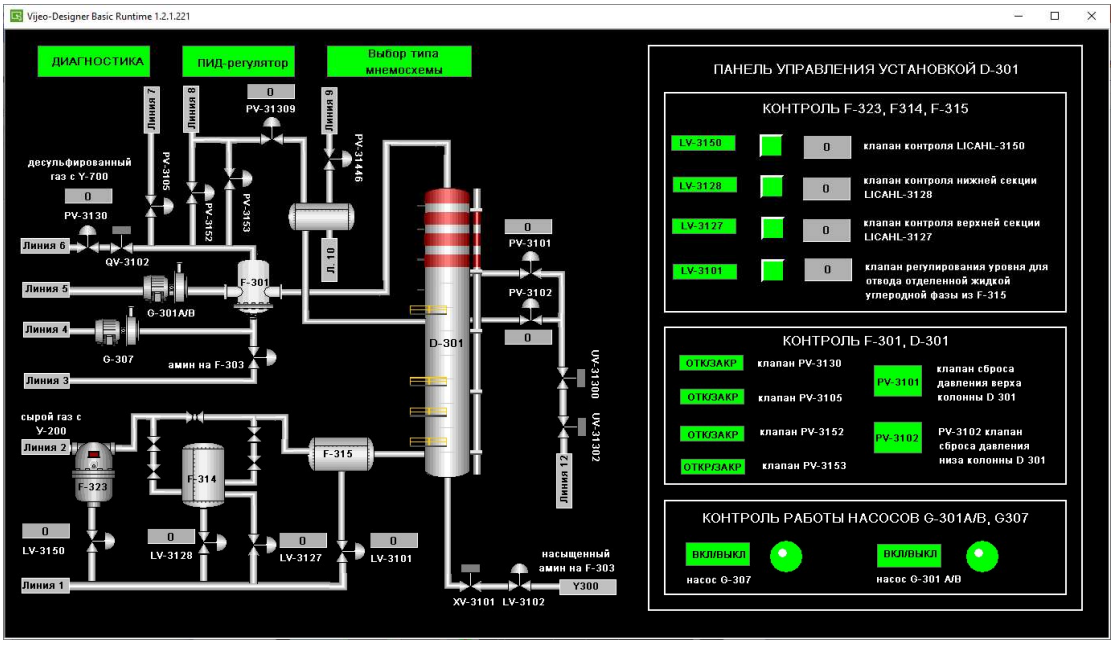


Рис. 2 Классический HMI дисплей для управления процессом по очистке газа высокого давления на установке D-301. Вкладка «Процесс»

Следующая вкладка «ПИД-регулятор» содержит с левой стороны график для отображения динамики объекта и работу ПИД-регулятора (Рисунок 3). Позволяет менять положения клапанов, а также значения коэффициентов *Ki, Kp, Kd*, в случае необходимости. Данный дисплей выполнен в классическом стиле.

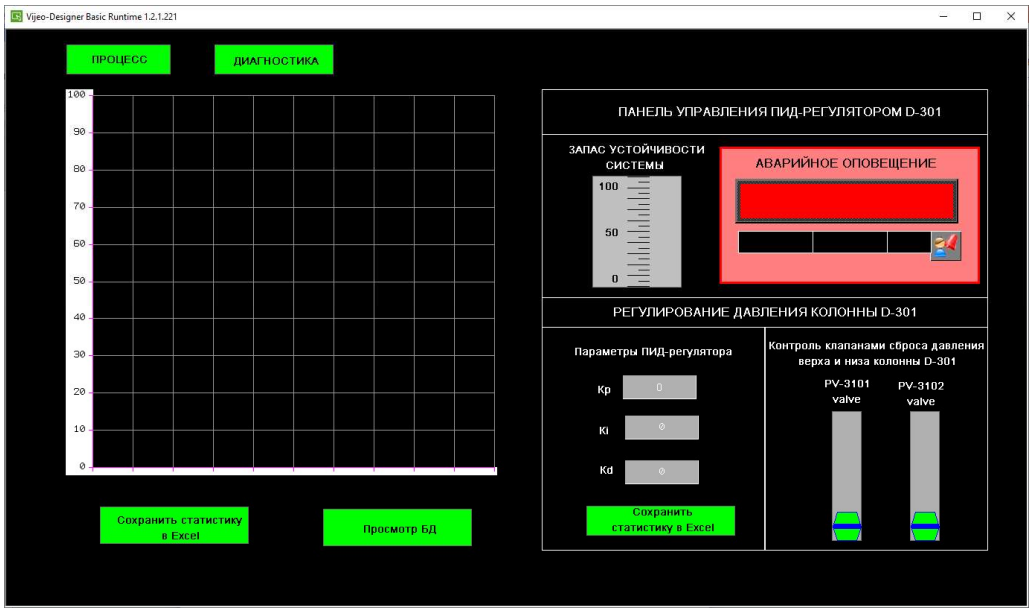


Рис. 3 Классический HMI дисплей ПИД регулятора установки D-301.
Вкладка «ПИД-регулятор»

Самой сложной для отображения вкладкой является дисплей «Диагностика». Модель интеллектуальной диагностики [5] на базе метода FMEA (Failure Mode and Effects Analysis), унифицированной искусственной иммунной системы и принципов иммунологического гомеостаза (FMEA+UAIS+Homeostasis) является избыточной. На рисунке 4 показана полная версия мнемосхемы в стиле Vijeo Designer Basic «по умолчанию».

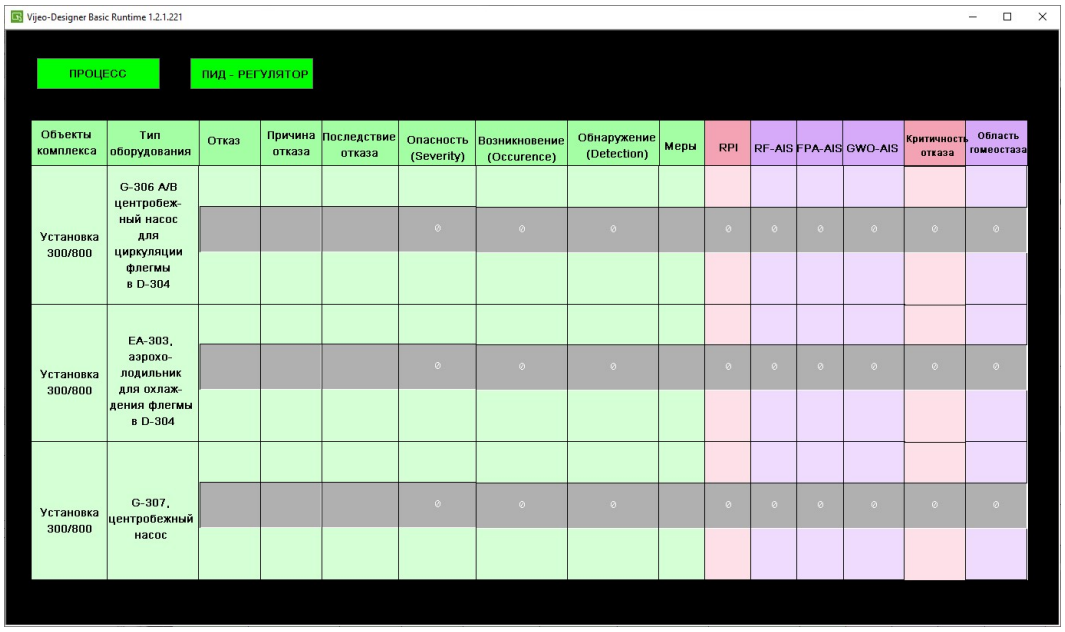


Рис. 4 Классический HMI дисплей ПИД регулятора установки D-301.
Вкладка «ПИД-регулятор»

Данное представление для оператора является затруднительным. Отслеживание аварийных сегментов работы процесса невозможно из-за большого количества параметров диагностики. Статистика загруженности дисплея представлена в таблице 1.

Таблица 1

Статистика Vijeo Designer Basic по загрузке системы

Загрузка памяти	Общий объем памяти	Использованный объем памяти	Доступный объем памяти
Статическая память с произвольным доступом (Static random access memory, SRAM).	128 KB	41	87 KB
Динамическая память с произвольным доступом (Dynamic random access memory, DRAM).	88 834 KB (90 966 016 байт)	вес проекта 788 KB	Использовано 29%

Согласно статистике, представленной в таблице 1 видно, что всего три диалоговых окна загрузили операционные возможности дисплея на 29%, тогда как для полноценного управления технологической линией требуются дополнительные элементы. В таблице 2 представлена статистика использования переменных и элементов для создания HMI дисплея.

Использовано в рассматриваемом примере: переменных типа «bool» – 34, переменных типа «real» – 20, переменных типа «integer» – 20, переменных типа «string» – 30.

Таблица 2

Статистика используемых элементов

Имя переменных	Тип	Источник	Начальное значение/адрес	Описание
CurPanelID	Integer	Internal	0	показывает текущие открытие панели.
Integer01	Integer	PLC	DM100	Pump1 Value
Integer02	Integer	PLC	DM101	Pump2 Value
Discrete01	Discrete	PLC	DM200:00	Pump1 Switch

Таким образом, актуально снижение *внутренней* когнитивной нагрузки на оператора за счёт снижения избыточности элементов и расстановки значков оповещения, для своевременной сигнализации о неисправности агрегатов.

На рисунке 5 представлен когнитивный дисплей для вкладки «Диагностика». Стиль дисплея выбран «по умолчанию».

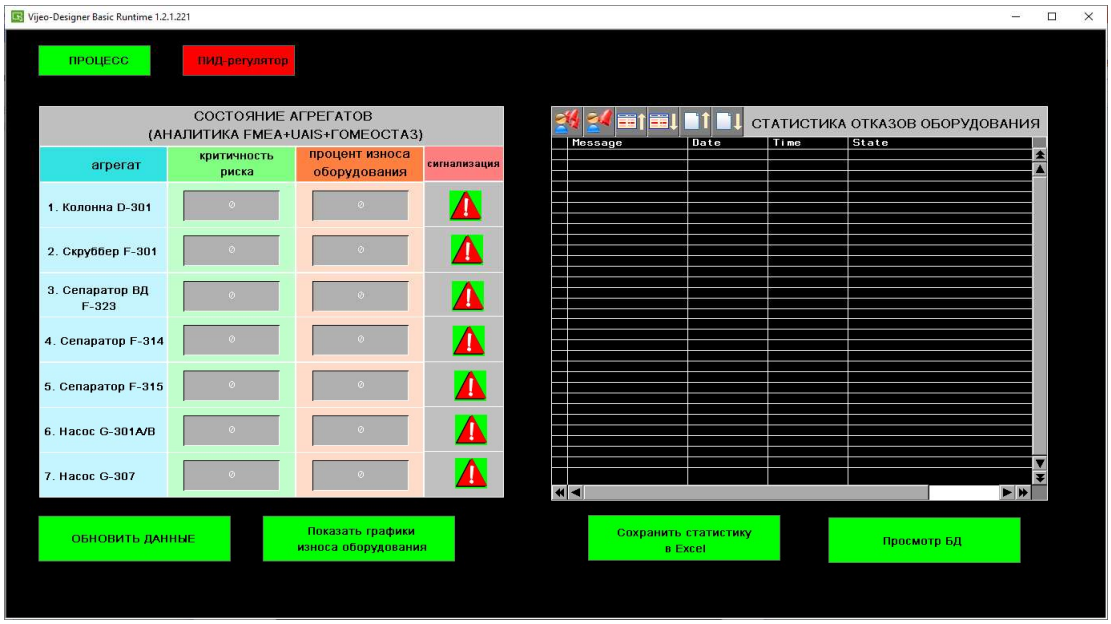


Рис. 5 Когнитивный HMI дисплей интеллектуальной диагностики состояния оборудования с оповещением. Вкладка «Диагностика»

После преобразования вкладки «Диагностика» показатель загруженности дисплея DRAM снизился с 29% до 18%. Помимо снижения *внутренней* когнитивной нагрузки по результатам тестов можно менять и *внешнюю* нагрузку на оператора [3].

Таким образом, применение когнитивных технологий для создания эффективных HMI дисплеев позволяет разработать удобные человеко-машинные интерфейсы, учитывающие *внешнюю* и *внутреннюю* нагрузку на операторов и позволяющие успешно интегрировать искусственный интеллект в реальное производство. Достоинством разработанного подхода при проектировании когнитивных мнемосхем является возможность корректного внедрения алгоритмов унифицированной искусственной иммунной системы для снижения внутренней когнитивной нагрузки оператора при управлении технологическим процессом и диагностики состояния оборудования, а также совместимость с предложенными ранее технологиями по представлению информации для снижения внешней когнитивной нагрузки. Применение представленных когнитивных HMI дисплеев на современных предприятиях нефтегазовой отрасли позволит повысить эффективность производства за счёт оперативного принятия решений и снижения количества аварийных ситуаций по вине человека, а также своевременной профилактической работой с диагностикой оборудования, что существенно снизит простои на производстве.

Исследования проводятся при поддержке Комитета науки Министерства образования и науки Республики Казахстан (Грант №AP09258508).

Литература

1. Y. Sineshchuk, A. Vostrykh, I. Saenko. Designing a Human-Machine Interface Based on the Cognitive Model of Information Systems Users // Proceeding of 2022 International Russian Automation Conference (RusAutoCon). – Sochi, IEEE, 2022. – P. 377–382. // DOI:10.1109/RusAutoCon54946.2022.9896356.

2. Hinss M. F., Brock A. M., Roy R. N. Cognitive effects of prolonged continuous human-machine interaction: The case for mental state-based adaptive interfaces // Front. Neuroergonomics. Sec. Cognitive Neuroergonomics. – 2022, Vol.3(935092). // https:/doi.org/10.3389/fnrgo.2022.935092.
3. Samigulina G. A., Samigulin T. I. Development of a cognitive mnemonic scheme for an optical Smart-technology of remote learning based of Artificial Immune Systems // Computer Optics. – 2021, - Vol. 45(2). – P. 286-295 // DOI: 10.18287/2412-6179-CO-736.
4. https://www.se.com. (Дата обращения: 12.08.2023).
5. Samigulina G. A, Samigulina Z. I. Biologically inspired unified artificial immune system for industrial equipment diagnostic / G. Nicosia et al. (Eds.) // J. Lecture Notes in Computer Science. Machine Learning, Optimization, and Data Science 8th International Workshop, – Certosa di Pontignano, Italy: Springer, 2023. LNCS 13811. – Part II. – P. 77-92. // https://doi.org/10.1007/978-3-031-25891-6.

ПЛАСТМАССА ӨНІМДЕРІН ҚАЙТА ӨҢДЕУДІ БАСҚАРУДЫҢ АВТОМАТТАНДЫРЫЛҒАН ЖҮЙЕСІН ЖАСАУ

Қалибаева Аружан Серікқызы, Карымсакова Нургуль Тлетаевна
E-mail: kalibaeva.aruzhan@gmail.com, nkarymsakova1@gmail.com
Әл-Фараби атындағы Қазақ Ұлттық университеті

Аңдатпа. *Бұл мақалада пластикалық өнімдерін қайталама пайдалану мақсатында қайта өңдеудің автоматтандырылған жүйесі құрылды. Бұл тақырыптың маңыздылығы әлем бойынша миллиондаған пластмасса қалдықтарынан ластанудан қорғау, оларға екіншілік қолданысқа жарату болып табылады.*

Жұмыс барысында пластмасса және оның түрлері, қолданылуы мен оларды қайта өңдеу түрлері баяндалды. Пластмасса өнімдерін қайта өңдеу үшін автоматтандырудың технологиялық, функционалды сұлбасы құрылып, техникалық құралдар кешендерінің сұлбасы құрастырылды. Сонымен қатар, TIA Portal-да ортасында пластмасса өнімдерінің қайта өңдеу процесінің визуализациясы мен бағдарламасы құрылды.

XXI ғасырдың ең өзекті экологиялық мәселелерінің бірі – пластмасса өнімдерін өмірдің кез-келген салаларында пайдалануы. Сол пластмасса өнімдері әдетте пайдаланусыз қалады. Осыған байланысты пластмасса қалдықтарының саны жыл сайын өсіп жатыр. Бүкіл әлем бойынша полиэтилентерефталатты құтылардың қалдықтарын пайдалану әлеулі мәселесі болып тұр.

Пластикалық өнімдерді пайланылған соң өзіндік барлық қасиеттерін сақтайды, сондықтан пластмасса қалдықтарды қайталама қайта өңдеу – экологиялық ең тиімді тәсілі.

Пластмасса өнімдерін қайта өңдеу арқылы қайталама өндіріске шикізат алынады. Ол шикізат негізі пластмасса гранулалары болып табылады.

Пластмасса өнімдерінің көп түрі қайта өңделуге келеді, бірақ оларды алдымен түрі, түсі, белгіленуі бойынша сұрыптау керек. Көп өндіріс орнындарында сұрыптау процесі қолмен жүргізіледі және ол көп уақытты алады. Пластмасса қалдықтарын автоматты түрде сұрыптау қайта өңдеу мәселесін шешуге әлдеқандай көмектесуіне ықпал етеді.

Пластмассаларды қайта өңдеу – бұл пластик қалдықтарын қайталама шикізатқа, энергияға немесе нақты тұтыну тауарлары бар өнімдерге айналдыру процесі.

Пластмасса өнімдерінің қалдықтары, бір жағынан қоршаған ортаны ластаушы көзі болып саналса, екінші жағынан бағалы ресурс болып табылады, Себебі пластмасса өнімдерін қайталама пайдалану үшін қайта өңдеуге ықтималды жарамды болып келеді.

50-ге жуық әр түрлі пластмасса түрлері бар, олар техникалық сипаттамаларымен ерекшеленеді. Полимерлерді және пайдаланылмаған бұйымдарды өңдеудің бірнеше әдістері бар, олардың ішінен келесілерді ажыратуға болады:

1. Пиролизді қолдану. Жоғары температураның әсерінен пластмассаларды қайта өңдеудің бұл әдісінде оттегі қолжетімділігі жоқ арнайы камераларда өңделеді. Бұл процестің өнімі газ, жылу энергиясы және мазут болып табылады.

2. Гидролиз арқылы. Вакуумдық цистерналарда судың, қышқыл ерітінділердің және жоғары температура әсерінен шағын пластик қалдықтары бөлінеді. Бұл энергияны көп тұтынады және ауқымды өңдеу кезінде өтеледі.

3. Екі атомды спирттерді қолдану. Бұл әдіс этиленгликоль мен жоғары температура қолданатын гидролиздің кіші түрі болып табылады. Ол бастапқы шикізатқа аз шығындалады, бірақ нәтижесінде пластмасса тамақ өнімдеріне қолдануға жарамсыз.

4. Метанолизді қолдану. Термиялық қайта өңдеудің ең танымал әдісі. Температурасы жоғары цистерналарда пластмассалар метанолдың көмегімен бөлінеді. Бұл әдіс тек ауқымды өңдеуге де жарамды, сондай-ақ ол жарылғыш болып табылады.

5. Механикалық рециклинг. Пластмасса алдымен сұрыпталады, кейін алдын ала ұнтақтауға, жууға және кептіруге жіберіледі. Алынған фрагменттерді ерітеді және сол жерде қалыптарға құяды немесе басқа салаларда шикізат ретінде пайдаланылатын түйіршіктерге ұсақтайды.

Бұл мақалада пластмасса өнімдерін қайта өңдеу үшін механикалық рециклинг әдісі қолданылады.

1.1 кесте

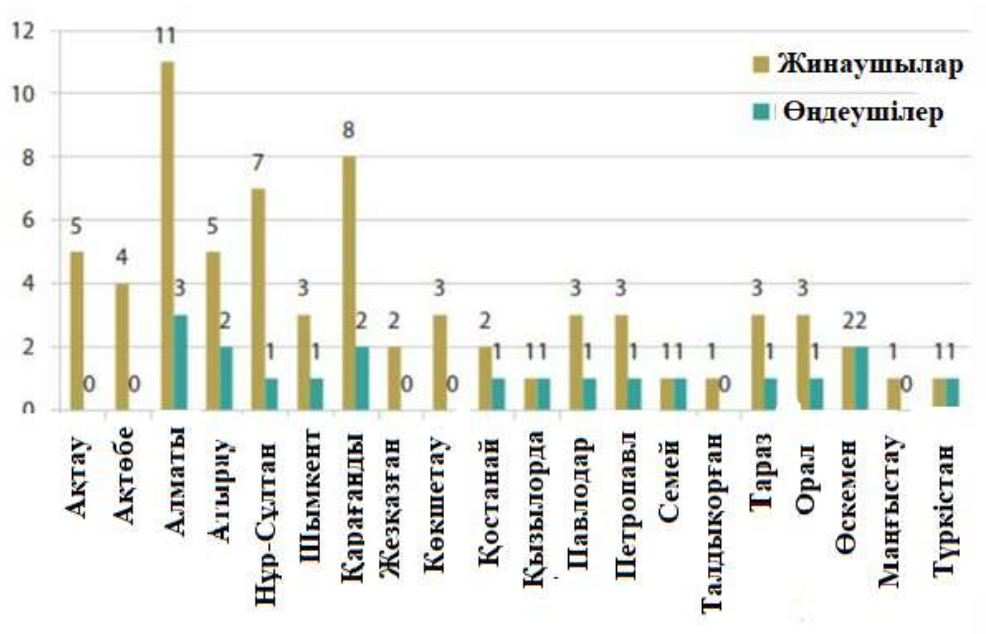
Қолмен және автоматты сұрыптау әдістерінің артықшылықтары мен кемшіліктері

Режим атауы	Артықшылығы	Кемшілігі
Қолмен	-Таңдалған компоненттердің тазалығы. - Шығындардың аздығы	- Процесстің ұзақ орындалуы - Еңбек өнімділігінің аздығы
Автоматты	- Процесстің тез орындалуы - Еңбек өнімділігінің аздығы - таңдалған компоненттердің тазалығы. - Әмбебаптық - Зиянды заттар шығарылуы, булану жоқ	- Жабдықтың жоғары құны; - Процестің жоғары энергия қарқындылығы

Әлемнің көптеген елдеріндегі сияқты Қазақстанда пластикалық қалдықтармен жұмыс істеу мәселесі өте өткір. «Оператор РОП» ЖШС деректері бойынша 2016 жылы пластикалық қаптаманың үлесі 283 387 тоннаны құраса, оны өңдеу көлемі 2017 жылы небәрі 6066 тоннаны (шамамен 2%), 2017 жылы 8994 тоннаны (шамамен 3%) құраған.

2019 жылы Тұрақты дамуға жәрдемдесу орталығы пластик қалдықтарын басқару саласына талдау жүргізді, негізінен ПЭТ және HDPE. Сараптама нәтижесінде бүгінде Қазақстанда осындай қалдықтарды жинап, өңдеумен айналысатын 80-нен астам кәсіпорын бар екені анықталды. Әдетте, бұл шағын және орта кәсіпорындар, жеке кәсіпкерлер, оның ішінде 69 құрастыру және 15 пластик өңдеу компаниялары (1-сурет).

Статистикалық мәліметтерге сүйенсмек қалдықтардың ең көп жиналуы Ақмола облысында – 1527, оның ішінде 742 полигон жойылды – 49%, Алматы облысында – 688, оның ішінде 307 полигон жойылды – 45%, Түркістан облысында – 404. , оның ішінде 122 полигон жойылды – 30%, БҚО – 419, оның ішінде 204 полигон – 49%.

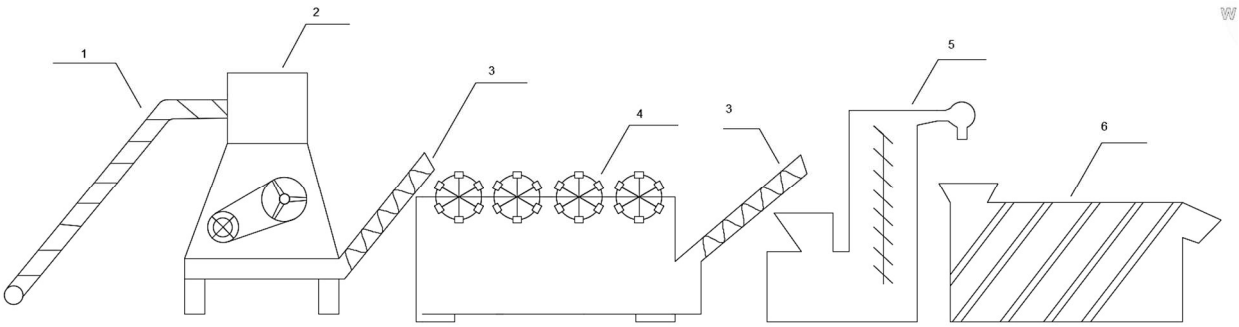


Сурет 1 Қазақстанның әртүрлі қалаларында жұмыс істейтін пластик қалдықтарын жинау және қайта өңдеу компаниялары

Пластмассаның жеке коллекциясын дамыту оны қайта өңдеу үлесін арттыруға белсенді әсер етеді. Бөлек жиналған таза шикізатты ұзағырақ сақтауға болады, бұл әрі қарай тасымалдауды оңтайландырады және қалдықтарды жинайтын компаниялардың шығындарын азайтады. Бөлек жинау сонымен қатар қайталама шикізатты қайта өңдеуді оңтайландырады және алынған өнімнің сапасын жақсартады.

Пластмасса өнімдерін қайта өңдеу процесінің технологиялық сұлбасы және оның жұмыс принципі. Технологиялық сұлбаны әзірлеу технологиялық жобалаудың ең маңызды бөлігі болып табылады. Ол жеке технологиялық процестер мен жабдықтардың өзара байланысы мен сипатын көрсетеді.

Пластмасса өнімдерін қайта өңдеудің сұлбасы келесі 2-суретте келтірілген.



Сурет 2 Пластмасса өнімдерін қайта өңдеу процесінің технологиялық сұлбасы мұндағы 1 – таспалы конвейер, 2 – роторлы ұсатқыш, 4 – флотация ваннасы, 3 – транспортерлік шнек, 5 – кептіргіш центрифугасы, 6 –гранулятор.

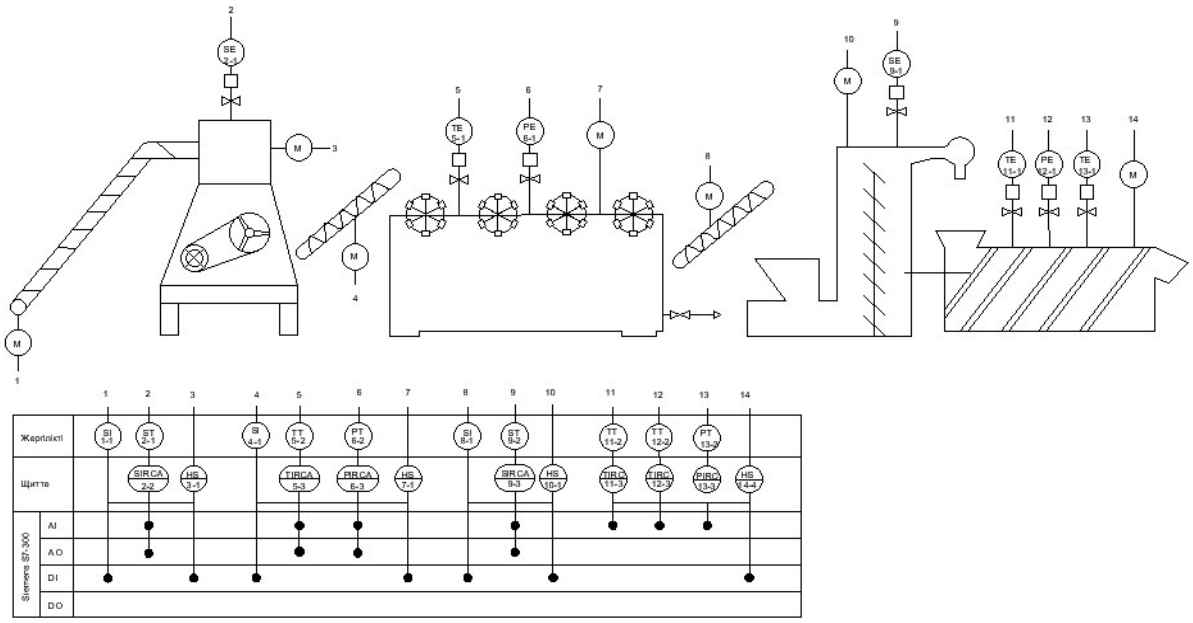
Пластмасса өнімдерін қайта өңдеу процесі рет ретімен орындалатын процес болып табылады. Кез келген пластмасса өнімі қайта өңдеуден бұрын түсі мен түрі бойынша сұрыптаудан өтеді.

Бұл жұмыста пластмасса өнімдері алдын ала сұрыпталған болып келеді, олар (1) таспалы конвейері арқылы (2) ұсатқышқа келіп түседі. Ұсатқышта олар 550-950 айн/мин жылдамдықпен айналатын ротор көмегімен ұсақ бөлшектерге ұсақтайды. Ұсақталған бөлшектер артық ластану мен шаңдардан жуылу үшін 0-60°С-та және 200-600кПа қысыммен жұмыс жасайтын (4) флотация ваннасына (3) транспортерлік шнек көмегімен жіберіледі. Әбден жуылған пластмасса өнімдерінің ұсақталған бөлшектері содан кейін траснпортерлік шнек көмегімен (5) кептіргіш центрифугасы кептірілуге жіберіледі. Кептіргіш центрифугасы 1500-2000 айн/мин жылдамдықпен бөлшектер толық құрғақ күйге енгенше кептіреді. Кептірілген ұсақ бөлшектер содан кейін гранулятор қондырғысына жіберіледі. Гранулятор қондырғысы жылытқыш, суытқыш, қалыптағыш және кескіш секілді бөліктерден тұрады. Грануляторға түскен ұсақ бөлшектер алдымен 150°С дейін қыздырудан өтеді, содан пайда болған балқыма 15-100 Мпа қысыммен экструзия әдісімен қысылып, арнайы жіп тәрізді тесіктерден өтеді. Содан соң пайда болған масса бірден суыту үшін 0-15°С суға суытылады. Суытылған балқыма массасы арнайы кескіштермен гранулаларға кесіледі. Осы процестерден өтіп, пайда болған гранулала келесі рет қолдануға жарамды болып саналады және қайталама өнімдер үшін шикізат болып табылады

Қайта өңдеу процесінің автоматтандырылған функционалдық сұлбасын әзірлеу. Автоматтандырудың функционалдық схемасы (АФС) технологиялық процес-терді автоматтандырудың негізгі конструкторлық құжаты болып табылады.

Бұл сұлбада өлшеу, параметр көрсеткіші, өлшемдерді тіркеу және реттеуді жүзеге іске асыратын датчиктер, атқарушы механизмдер мен қондырғылар бейнелегенген.

Сұлбада ең алдымен конвейер және конвейер қозғалысына жауап беретін моторы (1) белгіленген. Содан кейін ұсатқыш кондырғысында орналасқан жылдамдық датчигі мен оның атқарушы механизмі (2) көрсетілген. Бұл жылдамдық датчигі көрсеткіштерді көрсету, өлшеу, тіркеу, реттеу және сигнализация қызметтерін атқарады. Сонымен қатар, екі режимде: автоматты және қолмен басқаруға болатын қосқыш (3) атқарушы механизмдер орналасқан. Бұл атқарушы механизм тек ұсатқышта ғана емес флотация ваннасында, кептіргіш центрифугасында және гранулятордa орналасқан. Флотация ваннасында пластмассаның ұсақтамаларын жуу температурасын реттеу үшін температура датчигі (5) және жуу кезіндегі су қысымын реттеу мақсатында қысым датчигі (6) орналасқан. Ұсатқыш пен флотация ваннасы арасында және флотация ваннасы мен кептіргіш центрифугасы арасында тасымалдау конвейерлері орналасқан, олардың атқарушы механизмдері моторлары (4,8) сұлбада көрсетілген. Ұсатқышта орналасқан жылдамдық датчигі (9) кептіргіш центрифугада да орналасқан. Кептірілген ұсақталған пластмассалар грануляторға келіп түседі. Грануляторда олар үлкен температурада балқиды, сондықтан олардың балқу температурасын реттеу үшін температура датчигі (11), сол балқымаларды жіп тәрізді формаға келтіру үшін қысым датчигі (12) орналасқан. Пайда болған жіп тәрізді балқыманы әрімен қарай гра-нулаға айналдыру үшін оларды төменгі температурамен суытады. Сол температураны көрсеткіштерін тіркеп, реттеу үшін тағы температура датчигі (13) орналастырылған.



Сурет 3 Пластмасса өнімдерін қайта өңдеу процесінің функционалды сұлбасы

Техникалық құралдар кешенінің сұлбасын әзірлеу

ТП АБЖ келесі деңгейлерден тұрады:

Жоғарғы деңгей – бұл PC-үйлесімді дербес компьютер негізінде оператордың автоматтандырылған жұмыс орны (АЖО) және мәліметтер қорларының сервері. Олар өзара ethernet арқылы байланысады.

Ортаңғы деңгейі бағдарламалаушы логикалық контроллер Simatic Siemens S7-300 арқылы жүзеге асырылады.

ТҚҚ сұлбасының төменгі деңгейінде мына құрылғылар ұсынылған: жылдамдық датчиктері; температура датчиктері; қысым датчиктері; атқарушы механизмдер.

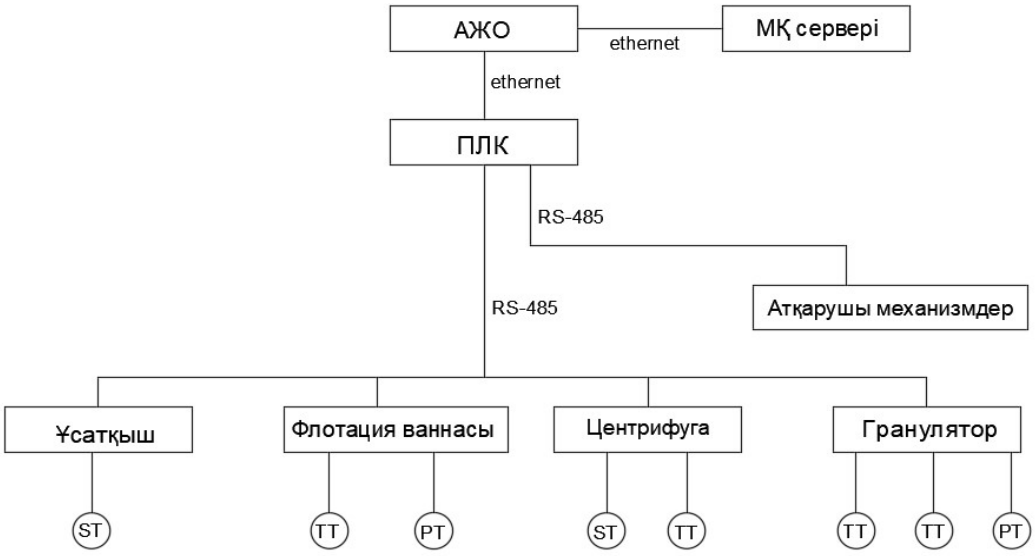
Ортаңғы деңгей мен төменгі деңгей өзара RS-485 протоколы арқылы байланысқан.

Пластмасса өнімдерін қайта өңдеудің АБЖ техникалық құралдар кешенінін сұлбасы 4-суретте көрсетілген.

Айналым жылдамдық датчиктерінің сипаттамаларын салыстыра және талдай келе IFM Electronic DI5028 датчик моделі таңдалды, себебі ұсатқыш роторы 550-950 айн/мин жылдамдықпен жұмыс жасайды, ал датчиктің өлшеу диапазоны – 6000 айналымға дейін.

Пластмасса ұсақтарының дұрыс жуылуы үшін белгілі бір температура режимін қадағалап отыру керек. Салыстыру және талдау барысында «Овен» компаниясының ДТС035 температура датчигі таңдалды, Флотация ваннасының қысымын қадағалап отыру үшін "Теплоприбоп" фирмасының ДДМ-03-1000 ДИ датчигі таңдалынды.

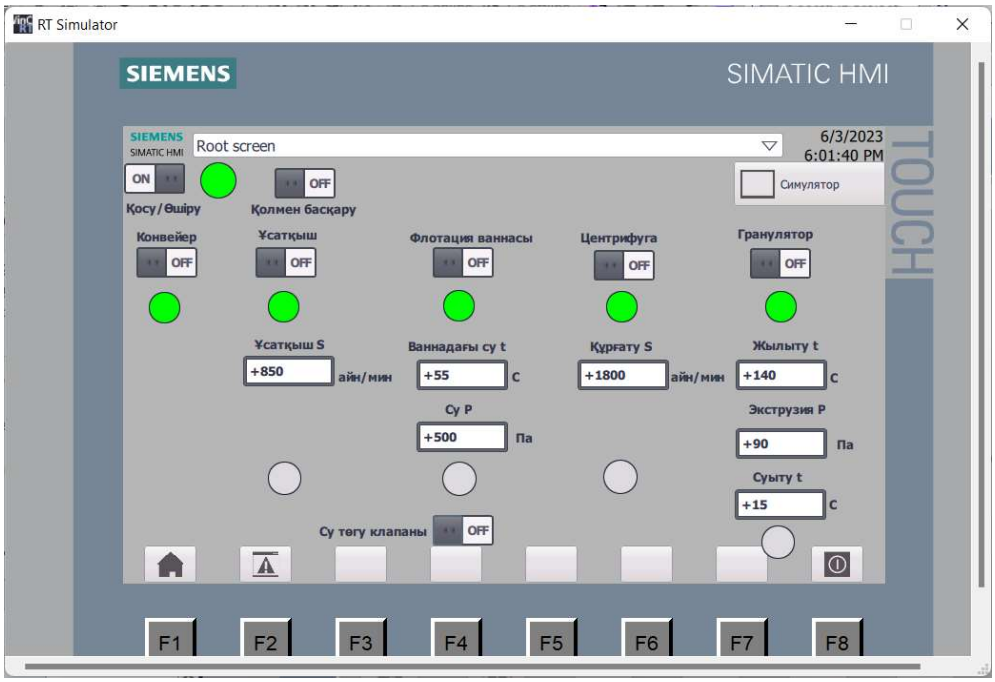
Пластмасса өнімдерін қайта өңдеу үшін материалдар алдымен конвейер арқылы ұсатқышқа келіп түседі. Конвейер жұмысын реттеу үшін үшфазалы асинхронды электрқозғалтқышы қолданылады және АИР 80А2 электрқозғалтқышы таңдалынды.



Сурет 4 ТҚК сұлбасы

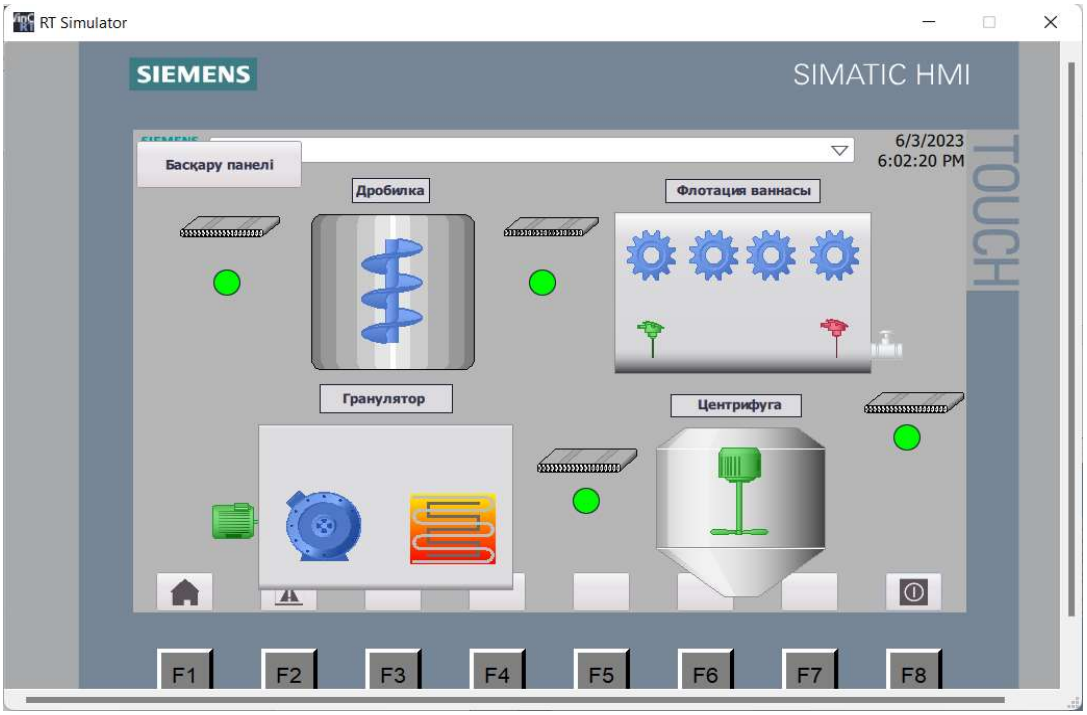
Tia Portal V16 ортасында пластмасса өнімдерін қайта өңдеу процесіне визуализация жүргізу

Пластмасса өнімдерін қайта өңдеу процесін визуализация үшін барлық процессті орындауға 55 әртүрлі типтегі кіріс және шығыс параметрлері ұсынылады. Тегтер тізімі құрылып болған соң, процесстің рет-ретімен дұрыс орындалуын LAD тілінде 22 Network құрылды. Жалпы бұл толық процесстің визуализациясын көрсету үшін екі HMI панелі құрылды: басқару панелі және визуалиция панелі. Бұл екі панелдің көрінісін келесі суреттерден көруге болады.



Сурет 5 Басқару панелі

Бірінші HMI панелі басқару панелі болып табылады. Мұнда қосу/өшіру дискретті батырмасы және процесстің қолмен басқару режиміне ауыстыруға болатын дискретті батырмасы орналасқан. Сондай-ақ, әрбір құрылғы атауы және оларға сәйкес қосылу тұрғанын хабар беретін лампалар, реттелтін параметрлер және тиісті авария сигналдары көрсетілген. Суретте байқанымыздай процесске қатысатын барлық қондырғылар қосулы және параметрлер реттеулі тұр. Сонымен қатар, панельде флотация ваннасының су төгу клапаны және симуляция панеліне өтуге мүмкіндік беретін батырмасы да орналасқан.



Сурет 6 Визуализация панелі

Екінші HMI панельде визуализация келтірілген. Мұнда әрбір қондырғы мен құрылғылар көрсетілген. Жоғарыда айтылғандай жүйенің толық жұмыс істеп жатқанын көруімізге болады. Егер жүйеде авария сигналдары орын алып жатса, авария сигналы қай қондырғыдан шығып жатқаны сәйкес қызыл лампа жанады. Әрбір қондырғыға орнатылған авария лампалары аварияларды анықтауды тездетеді және бұл шешім жұмыс жасауда ыңғайлы болып табылады. Авария орын алған жағдайда жүйе толығымен істен шығады.

Қорытынды

Бұл мақалада пластмасса өнімдерін екіншілік қолданысқа жіберу үшін қайта өңдеуді басқарудың автоматтандыру жүйесін жасау қарастырылды. Сондай-ақ жұмысқа қойылған тапсырмаларға қол жеткізілді. Зерттеу барысында келесідей есептер шешілді:

- Пластмасса өнімдері және оларды қайта өңдеу түрлеріне шолу жасалды;
- Қазақстандағы пластмасса өнімдерін өңдеудің мәселелері қарастырылды;
- Қайта өңдеу үшін қажетті құрылғылар мен қондырғылар таңдалып, бірыңғай автоматтандырылған жүйесі әзірленді. Пластмасса өнімдерін қайта өңдеу үшін технологиялық сұлба құрастырылды;

- Қайталама өнімге шикізат алу үшін қайта өңдеу процесіне автоматтандырудың функционалдық сұлбасы құрылды;
- Таңдалынған құрылғылар мен контроллер негізінде техникалық құралдар кешенінің сұлбасы құрылды;
- Tia portal V16 ортасында пластмасса өнімдерін қайта өңдеу процесінің визуализациясы жасалды;

Қолданылған әдебиеттер тізімі

1. Шевцова А. А. Утилизация изделий из пластика // Современные научные исследования и инновации. 2016. № 1.
2. Все об утилизации пластиковых бутылок. – http://www.makulaturin.ru/vsyo-ob-utilizatsii-plastikovyih-butyilok/
3. Переработка ПЭТ бутылок. – http://to-biz.ru/pererabotka-petbutylok/
4. Керницкий В. И., Жир Н. А. Переработка отходов полиэтилентерефталата (ПЭТ) // Полимерные материалы № 8, 2014. - С. 11-20.
5. Волчкевич Л. И. Автоматизация производственных процессов: Учебное пособие для вузов // Машиностроение, 2007. – 380 с.
6. Автоматизированные системы диспетчерского и технологического управления. – http://en-res.ru/stati/scada.html

ҒИМАРАТТЫҢ МИКРОКЛИМАТЫН БАСҚАРУДЫҢ АВТОМАТТАНДЫРЫЛҒАН ЖҮЙЕСІН ӘЗІРЛЕУ

Құрманова Н. Н., Карымсакова Н. Т.
e-mail: nuraida.kurmanovaa@gmail.com, nkarymsakova1@gmail.com
Әл-Фараби атындағы Қазақ Ұлттық университеті

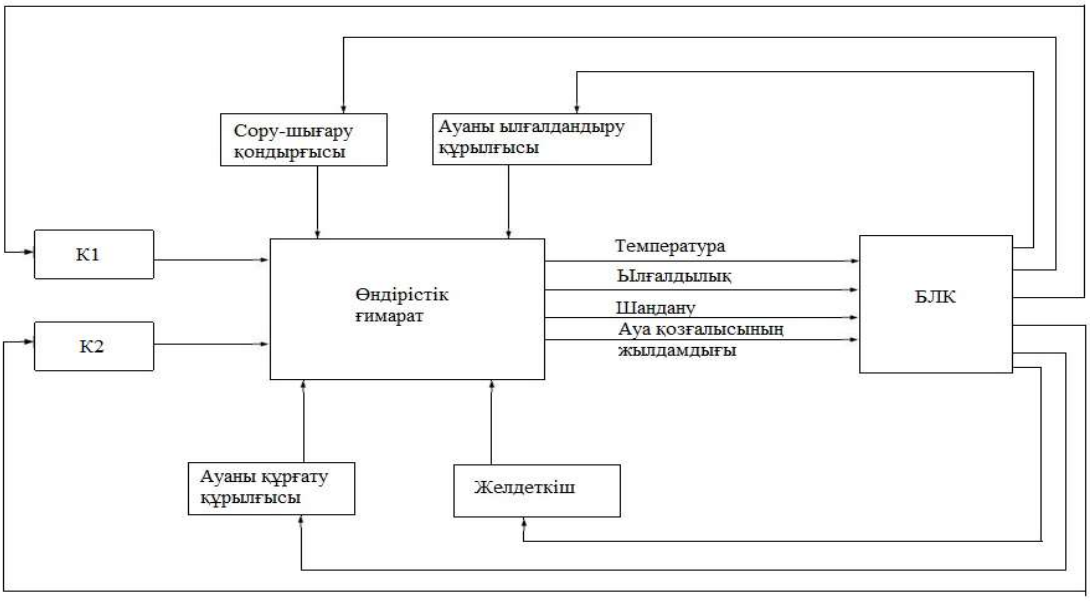
Аңдатпа. *Мақаланың басты мақсаты бұл ғимараттың микроклиматын қамтамасыз етудің автоматты басқарылатын жүйесін құру. Жобаны жүзеге асыру барысында ағынды-құйылмалы желдету жүйесі зерттеліп, қажетті қондырғылар таңдалынады. Желдету жүйесінің принципиалды сұлбасы тұрғызылды. Сонымен қатар ұқсас автоматтандырылған басқару жүйелері егжей-тегжейлі талдаады, соның негізінде технологиялық процестің құрылымдық сұлбасы және автоматтандырудың функционалдық сұлбасы құрылды. Matlab ортасындағы Simulink пакетінде басқару объектісінің моделі құрастырылды. SCADA жүйесінде желдету жүйесінің автоматты басқару бағдарламасы құрылды.*

Адам денсаулығының жағдайы, оның өнімділігі көбінесе бөлменің микроклиматына байланысты. Атмосферада болып жатқан климаттық процестерге тиімді әсер ете алмайтын адамдар өндірістік үй-жайлар ішіндегі ауа факторларын жоғары сапалы бақылауға алады.

Үй-жайлардың микроклиматы - бұл үй-жайлардың ішкі ортасының климаты, ол температурамен, салыстырмалы ылғалдылықпен және ауа қозғалысының жылдамдығымен, сондай-ақ адам ағзасына бірге әсер ететін қоршаған беттердің температурасымен анықталады. Микроклиматқа әсер ететін факторларды екі топқа бөлуге болады: реттелмейтін (белгілі бір аумақтың климатын құрайтын факторлар кешені) және реттелетін (ғимараттар мен құрылыстардың құрылысының сипаттамалары мен сапасы, қарқындылығы, жылытқыштардан жылу сәулелену жиілігі, ауа алмасуы, саны. бөлмедегі адамдардың саны және т.б.). Адамның оңтайлы микроклиматтық жағдайларда ұзақ және жүйелі болуымен дененің қалыпты функционалдық және жылулық күйі терморегуляция механизмдерін шамадан тыс жүктеместен сақталады. Бұл ретте жылулық жайлылық сезіледі (сыртқы ортамен қанағаттану) және өнімділіктің жоғары деңгейі қамтамасыз етіледі. Жұмыс орнында бұл жағдайларға басымдық беріледі. Микроклимат параметрлерінің сапасын қажетті деңгейде бақылау жылыжай кәсіпорындарының жұмысының маңызды бөлігі болып табылады. Агроөнеркәсіп кешенінің белсенді дамуына, осы салада алынатын өнім санының артуына және қазіргі экономикалық жағдайға байланысты бұл мәселеге үлкен көңіл бөлу қажет. Микроклиматты дұрыс сақтау технологиясы өнімділікті арттыруға және өнім сапасын жақсартуға, нәтижесінде жалпы өндірістің тиімділігі мен бәсекеге қабілеттілігін арттыруға мүмкіндік беретін маңызды элемент болып табылады. Ең маңызды сәттердің бірі – энергия ресурстарын ұтымды пайдалану. Бұл өндіріс процесінің өзіндік құнын айтарлықтай төмендетудің қосымша мүмкіндігі. Заманауи автоматтандырылған ауаны кондиционерлеу жүйесі тұрақты жұмыс жағдайын қамтамасыз етіп қана қоймай, сонымен қатар атқарушы жүйенің мүмкіндіктерін барынша пайдалануға және кез келген

конфигурация мен өлшемдегі жылыжайларға оңай біріктірілуі керек. Қазіргі уақытта инженерлердің көпшілігі жылыжайларды жаңғырту кезінде атқарушы жүйелердің санын көбейтуді жөн көреді: желдету жүйелерін жаңарту, перделерді орнату, желдеткіштерді орнату, релелерді орналастыру, жарықтандыру жүйелері, әртүрлі конфигурациялар мен дизайндағы суару жүйелері. Жылыжайлардағы атқарушы жүйелердің санының артуы оларды жұмыс жағдайында ұстауға кететін шығындардың өсуіне әкелетінін атап өткен жөн. Дегенмен, біздің жағдайда, ең бастысы, төмен құны, ауқымдылығы және пайдаланудың қарапайымдылығы. Әзірленген жүйенің бұл ерекшелігі оның кәсіби фермерлер мен осы жүйені өз учаскелерінде қолданатын қарапайым бағбандар арасында танымалдылығын қамтамасыз етеді. Біздің жүйенің төмен құнымен біріктірілген масштабтаудың және пайдаланудың қарапайымдылығының принциптеріне сүйене отырып, пайдаланушылар бірнеше желілік төменгі деңгейлері бар көп блокты жүйелерді жасай алады. Алдыңғы ақпаратты қорытындылай келе, біздің жүйе заттар интернеті тұжырымдамасын жүзеге асырады деп айта аламыз. Зерттеу объектісі – өндіріс орны, соның ішінде тамақ өнеркәсібі. Өндірістік ғимараттың микроклиматы – бұл бөлмелердің ішкі ортасының климаты, яғни ол адам ағзасына бірге әсер ететін температура, салыстырмалы ылғалдылық және ауа ағынының жылдамдығымен, сондай-ақ қоршаған беттердің температурасымен анықталады. Өндірістік ғимараттағы микроклиматты сипаттайтын параметрлер:

- ауа температурасы;
- салыстырмалы ауа ылғалдылығы;
- ауа қозғалысының жылдамдығы;
- шаңдану.



1 сурет Өндірістік ғимараттың микроклиматын басқарудың технологиялық сұлбасы

Микроклиматтың сақталуы ғимаратта жұмыс істеп жүрген жұмысшылардың жұмыс істеу өнімділігін арттырады. Егер де микроклиматтың оптималды мәндер бұзылатын болса өнімділік төмендейді. Яғни температура көбейіп кетсе болмаса ылғалдылық немесе құрғақтылық көп болып кетсе жұмысшылар өзін жайсыз сезінетін

болады. Сондықтан микроклиматтың осы оптималды мәнін ұстап отырудың автоматты жүйесі құрылады.

Мұнда яғни белгілі бір процесс кезек кезегімен жүрмейді, барлық параметр бір-біріне тәуелсіз жұмыс жасай алады. 4 датчик және 6 атқарушы механизм жұмыс жасайды. Мұндағы мақсат осы 4 параметрдің оптималды мәндерін сақтай отырып өндірістік ғимаратта қолайлы микроклиматты қалыптастыру және сақтау.

К1 дегеніміз салқындату режиміндегі өндірістік конционер, ал К2 жылыту режиміндегі өндірістік конционер. Егер біз өндірістің ПӘК-ін арттырғымыз келсе, яғни өнім көп шыққанын қаласақ, ғимаратта микроклиматтың оптималды мәндерін ұстап тұру керек. Мысалы жұмысшылар күнде түнде 12 сағаттан жұмыс істейтін болса олардың өмірлері қауіпсіз болу үшін оптималды аура, режимде жұмыс жасауы керек. Көп жерде мұндай микроклимат сақталмаған.

1 кесте

СанПиН 2.2.4.548-96 өндірістік ғимараттардың микроклиматына қойылатын гигиеналық талаптар

Жыл мезгілі	Жұмыс категориясы	Температура, °С	Салыстырмалы ылғалдылық, %	Ауа жылдамдығының қозғалысы, м/с	Шаңдану, мг/м^3
Салқын	I II III	22...24 21...23 16...18	40...60	0,1 0,1 0,3	2...6
Жылы	I II III	23...25 22...24 18...20	40...60	0,05 0,2 0,4	2...6

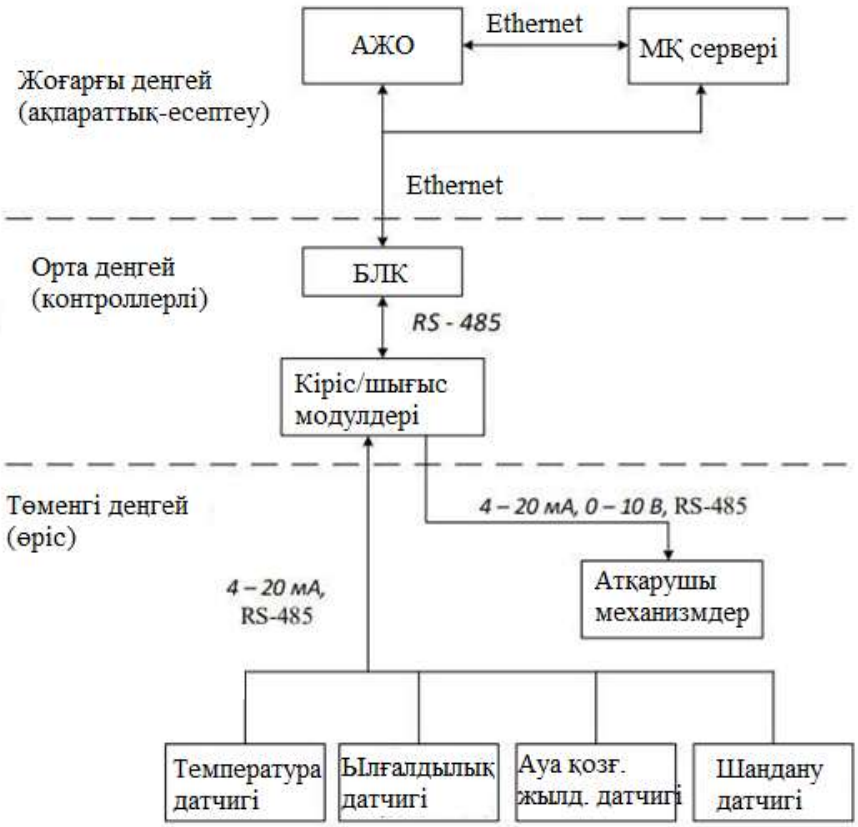
Микроклиматты басқару жүйесі температураны, ауа ылғалдылығын, ауа қозғалыс жылдамдығын және ауадағы шаң концентрациясын көрсететін датчиктарды керек қылады. Температура датчигі ретінде EE451, ауа ылғалдылығын өлшеу датчигі ретінде SHA0-79N114, ауадағы шаң концентрациясын көрсететін датчик ретінде RPS-SD, Ауа қозғалысының жылдамдығы датчигі ретінде EE680 датчиктері таңдалынды. Таңдалыну себептері оның жұмыс температурасының диапазонына, өлшеу қателігіне, орташа жарамдылық мерзіміне қарай отырып таңдалынды. Ең бастысы өлшеу қателігіне назар аударылды, себебі өлшеу қателігі неғұрлым аз болса, өлшеу дәлдігі соғұрлым жоғары болады. Одан кейін контроллердің SIEMENS S7-300 CPU 314 түрі таңдалынды.

Технологиялық процесті бақылау бөлмедегі микроклиматтың қажетті параметр-лерін сақтауға дейін азаяды. Басқарудың құрылымдық схемасы ақпаратты берудің үш деңгейімен ұсынылған (2 сурет).

Төменгі деңгей (өріс) процестің параметрлері туралы ақпаратты жинайтын бастапқы датчиктерден және басқару әрекетін жүзеге асыратын жетектерден тұрады. Іске қосу құрылғылары - жылыту құрылғылары, желдеткіш, ылғалдандырғыштар, ауаны ылғалдандырғыштар.

- Орта деңгей (контроллер деңгейі) келесі әрекеттерді орындайтын бағдар-ламаланатын логикалық контроллерден (PLC) тұрады:
- жоғарғы деңгейдегі командаларды орындау;
- жоғарғы деңгеймен ақпарат алмасу;

– технологиялық процестің параметрлері туралы ақпаратты жинау және алғашқы өңдеу;
– автоматты реттеу.



2 сурет Процестің техникалық құралдар кешенінің құрылымы

Жоғарғы деңгей (ақпараттық-есептеу) Ethernet желісі арқылы PLC-ге қосылған компьютерден тұрады, беру ортасы ретінде мыстан жасалған бұралған жұп қолданылады.

Процесс блок-сұлбасының алгоритмі:

1. Автоматты режимде салқын режимде Iкатегорияда, себебі жеңіл өнеркәсіп, барлық параметр 0 болып тұрады, ылғалдылық 0ге тең болғандықтан ауаны ылғалдандыру құрылғысы қосылып тұрады және температура 0ге тең болғандықтан жылыту режиміндегі өндірістік кондиционер қосылып тұрады. Қалған параметрлер қалыпты режимде жұмыс жасап тұр.

2. Жыл мезгілінің жылы режимінде де дәл солай жұмыс жасалынады.

3. Әр параметрде белгілі бір оптимальді аралық мәндері бар, бірінші температура мәнін берсек (22-24), мысалы ішкі температура 24градустан асып кететін болса, ғимарат іші ысып кетеді, сондықтан салқындату режимдегі өндірістік кондиционер қосылады. 22градустан төмендеп кетсе жылыту режимдегі өндірістік кондиционер қосылады.

4. Жылы мезгілде де дәл солай орындалады (23-25).

5. Екінші ылғалдылық мәнін берсек (40-60%), мысалы ғимарат ылғалдылығы 40тан төмендеп кетсе, ауаны ылғалдандыру құрылғысы қосылады, ал 60тан жоғарылап кетсе,

ауаны құрғату құрылғысы қосылады, себебі ғимарат ішіне ылғал жетіспей немесе ылғал көп болып кетсе жұмыс істеу өнімділігі төмендеп кетеді.

5. Жылы мезгілде де дәл солай орындалады (40-60).

6. Үшінші ауа қозғалыс жылдамдығы датчигін реттеліп көріледі.

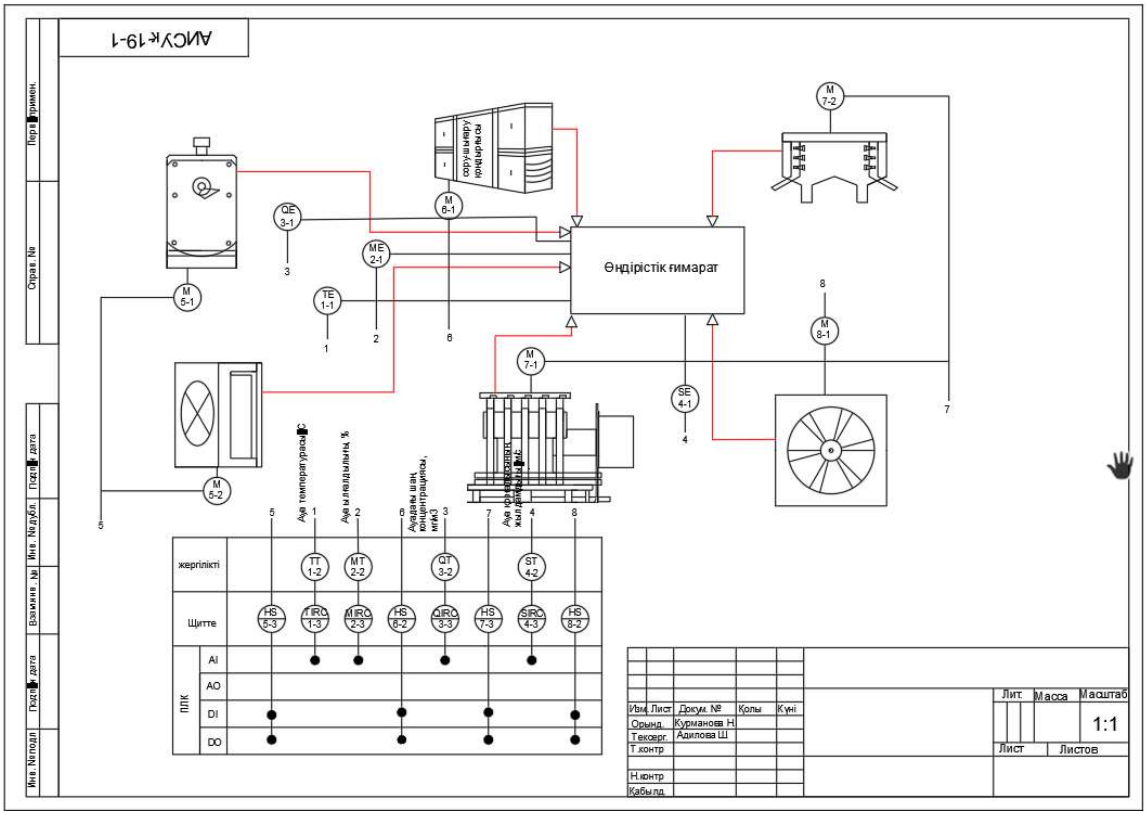
7. Салқын мезгіл кезінде ауа қозғалысының жылдамдығы қосылып тұрады, себебі 0 көрсетіп тұрады, оптималды мәні 0,3 тен асып кетсе сәйкесінше вентиляция қосылады.

8.Жылы мезгілде де тура солай орындалады.

9. Салқын мезгіл кезінде шаңдану көбейіп кетсе, яғни, оптималды 7 мәнінен асып кеткен жағдайда сору-шығару қондырғысы қосылады.

10. Жылы мезгілде де солай орындалады.

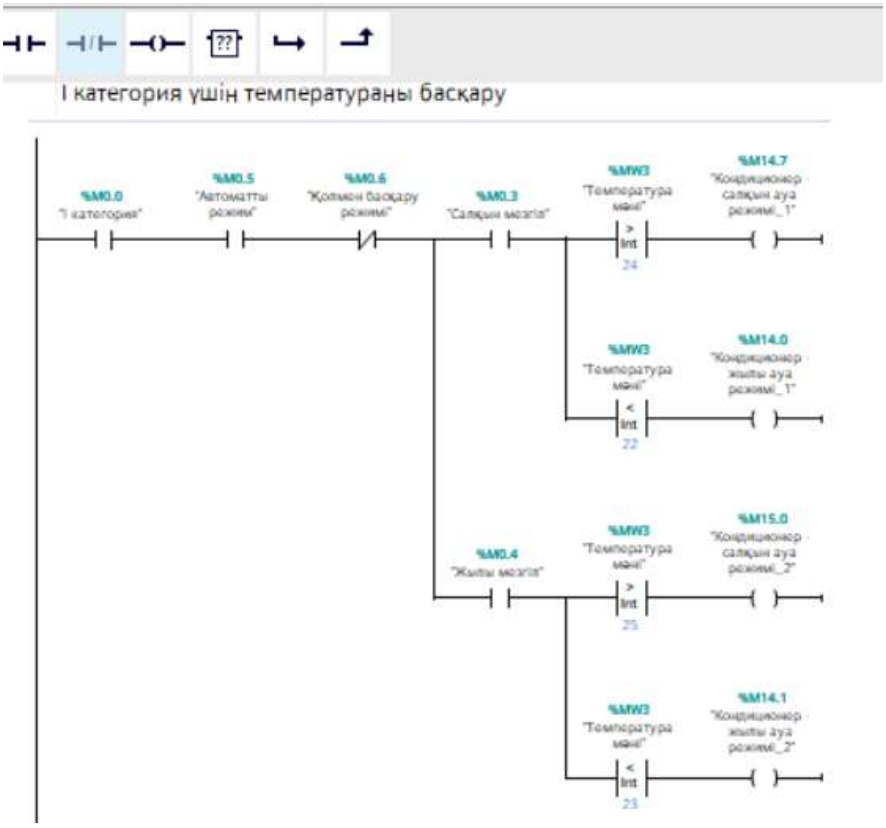
11. Қолмен басқару кезінде кез-келген батырманы басу арқылы сіз микроклимат басқара аласыз.



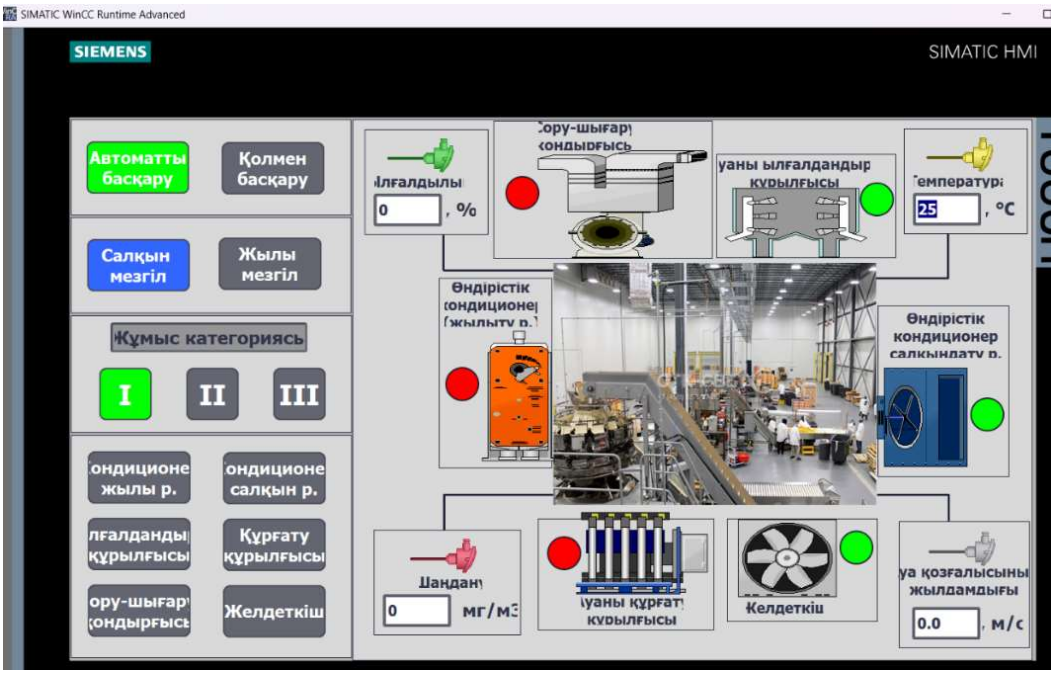
3 сурет Өндірістік ғимараттың микроклиматын басқарудың функционалдық сұлбасы

Бағдарламалық қамтамасыз етуді әзірлеу барысында Wincc ортасы таңдалынды. Жұмысты бастар алдында символдар кестесінде ұсынылатын кіріс және шығыс параметрлері көрсетіледі.

Жұмыстың орындалу барысының бір бөлігі: 1-категориядағы жұмыстар үшін автоматты режимдегі салқын және жылы мезгілдер үшін бөлмедегі ауа температурасын реттеу алгоритмі. Температураның шектік мәндері кестеден алынды. Ауа температурасы өндірістік кондиционерлермен реттеледі.



4 сурет I категория үшін температураны басқару



5 сурет Салқын мезгіл кезіндегі I жұмыс категориясында температура оптималды мәннен асып кетуі кезіндегі визуализациясы

Қорытынды

Мақаланы жазу барысында ғимараттың микроклиматын басқарудың автоматтандырылған жүйесін әзірлеу қарастырылды. Сәйкесінше мақаланы орындау үшін қойылған тапсырмалар анықталды.

Мақаланы орындау кезінде келесі тапсырмалар шешілді:

- интернет сайттардың ішінен ғимаратта микроклимат қалыптастыру және оны басқару бойынша жұмыстар жүргізілді, мысалға айтатын болсақ басқарудың бір атқарушы механизмін алып, оның сипаттамасы және түрлері айтылды және талдау жүргізілді;
- ғимараттың микроклиматын басқарудың технологиялық сұлбасы сызылды және оның процесі қалай жүретіні көрсетілді;
- әр параметрдің оптималды мәндер аралықтары стандарт бойынша көрсетілді;
- ғимараттың микроклиматын басқаруды автоматтандырудың функционалдық сұлбасы жасалынды;
- функционалды сұлбаға қарап негізгі параметрлер: температура, ауа ылғалдылығы, ауа қозғалысының жылдамдығы, шаңдану мәндері толықтай зерттелініп, соған сәйкес датчиктер және басқарудың техникалық құралдары мен SIEMENS контроллері таңдалынып алынды;
- Ғимараттың микроклиматын басқару жүйесінің жұмыс істеу алгоритмінің блок-сұлбасы әзірленді;

- Tia Portal, WinCC ортасында ғимарат ішіндегі микроклиматты басқарудың автоматтандырылған жүйесінің визуализациясы жасалды.

Әдебиеттер тізімі

1. Кокорин О. Я., Варфоломеев Ю. М. Системы и оборудование для создания микроклимата помещений; ИНФРА-М-М, 2017. – 272с.
2. Микроклимат производственных помещений. Терморегуляция. Влияние микроклимата на работоспособность человека: сайт. 2017. – URL: https://cyberpedia.su/12x45f7.html (дата обращения: 19.03.2021). – Текст: электронный.
3. Иванов, А. А. Автоматизация технологических процессов и производств. Учебное пособие / А. А. Иванов. – М.: Форум, Инфра-М, 2015. – 224 с.
4. Клюев, А. С. Проектирование систем автоматизации технологических процессов / А. С. Клюев, Б. В. Глазов, А. Х. Дубровский. – М.: Энергия, 2015. – 512 с.
5. Селевцов, Л. И. Автоматизация технологических процессов: Учебник / Л. И. Селевцов. - М.: Academia, 2019. - 160 с.
6. Хисаров Б. Д., Погребняк В. В., Искакова Г. Т. Бакалавр дәрежесін алуға арналған дипломдық жоба. 5В070200 – Автоматтандыру және басқару мамандығы студенттеріне дипломдық жобаны орындауға арналған әдістемелік нұсқау. Алматы: АЭжБУ, 2016. - 30 б.
7. Степанов А. Применение средств автоматизации // Автоматизация в промышленности элктрон. науч. журн. 2015. - С.17-20.

ОБЗОР АРХИТЕКТУРЫ ПОИСКОВЫХ СИСТЕМ РЕАЛЬНОГО ВРЕМЕНИ

Мергенғали Е. Қ.
e-mail: estony.9999@gmail.com
Казахский национальный университет имени аль-Фараби, Алматы, Казахстан

Аннотация. *В этой статье представлен всесторонний обзор архитектур поисковых систем реального времени, проливающий свет на важнейшие компоненты, проблемы и технологии, лежащие в основе этих систем. Поисковые системы, работающие в режиме реального времени, стали незаменимыми в нашем мире, ориентированном на данные, и обслуживают самые разные отрасли — от социальных сетей до финансов. Ключевые компоненты, включая уровни приема, механизмы индексирования, механизмы запросов, распределенное хранилище и алгоритмы ранжирования поиска, разъяснены, что дает читателям четкое понимание сложной работы этих систем. В статье также освещаются серьезные проблемы, с которыми сталкиваются поисковые системы реального времени, такие как управление высокой скоростью передачи данных, достижение низкой задержки, обеспечение согласованности данных в распределенных системах и балансирование масштабируемости с локальностью данных. Эти проблемы подчеркивают сложность разработки и поддержки архитектур поисковых систем реального времени.*

Ключевые слова: *поисковая система реального времени, архитектура поисковой системы реального времени, механизмы индексирования, машинное обучение.*

Введение

В современном информационном мире спрос на поисковые системы, работающие в режиме реального времени, резко возрос. Пользователи ожидают мгновенного доступа к самой актуальной информации, что делает поисковые системы, работающие в режиме реального времени, важнейшим инструментом для получения информации. Эти поисковые системы поддерживают все: от новостных веб-сайтов и платформ социальных сетей до электронной коммерции и финансовых приложений. В этой статье мы углубимся в архитектуру, лежащую в основе поисковых систем реального времени, изучим их компоненты, проблемы и технологии, которые ими управляют.

Поисковые системы реального времени — это специализированные системы, предназначенные для индексации и извлечения данных практически в реальном времени. В отличие от традиционных поисковых систем, которые периодически обновляют свои индексы, поисковые системы в режиме реального времени постоянно принимают, индексируют и обслуживают свежие данные по мере их появления. Эти системы имеют решающее значение для приложений, требующих мгновенного доступа к текущей информации, таких как платформы социальных сетей, новостные порталы и инструменты мониторинга.

Ключевые компоненты поисковых систем реального времени

Уровень приема. В основе любой поисковой системы реального времени лежит уровень приема. Этот компонент отвечает за сбор и обработку данных из различных источников, таких как веб-каналы, базы данных, API социальных сетей и т. д. Уровень приема часто использует такие технологии, как Apache Kafka, RabbitMQ или специально созданные конвейеры для эффективной обработки потоков данных большого объема. В настоящее время ежедневно создаются огромные массы данных. Например, пользователи социальных сетей и Интернета вещей генерируют большие объемы разнообразных данных, которые необходимо передавать, записывать и обрабатывать на высокой скорости. Эти данные представлены в различных типах структур, а именно структурированные, полуструктурированные и неструктурированные данные. На уровне архитектуры больших данных эти различные источники данных расположены на уровне источников данных, который является отправной точкой для дальнейшей обработки больших данных. В [1] предлагаются методы, связанные с модельно-ориентированной разработкой «MDE», чтобы предложить универсальное метамоделирование как для слоев источников данных, так и для слоев приема больших данных.

Механизм индексирования. Механизм индексирования отвечает за преобразование необработанных данных в документы, доступные для поиска. Он анализирует, маркирует и организует данные в инвертированный индекс — структуру данных, которая сопоставляет термины с их расположением в документах. Elasticsearch, Apache Solr и Amazon Elasticsearch Service — популярные варианты индексации в реальном времени. В [2] описывается проектирование и разработка системы индексирования, кластеризации и поиска научных документов и показывается, что реляционные базы данных, такие как MySQL, становятся все более неспособными удовлетворить потребности пользователей в быстром поиске данных и то, что Elasticsearch компенсирует задержку поиска, предлагая пользователям совместимость с быстрым поиском, обеспечивая при этом высокое качество. В статье был разработан веб-сервер Java Spring с Bootstrap, jQuery и Foamtree, визуализирующий данные с помощью Kibana, позволяющий оркестровать используемые технологии, обеспечивая эффективные функции поиска и анализа.

Механизм запросов. Механизм запросов обрабатывает запросы пользователей и извлекает соответствующие результаты из индексированных данных. Он должен быть способен обрабатывать большой объем одновременных запросов и возвращать результаты за миллисекунды. Для этого используются такие технологии, как Apache Lucene и распределенные поисковые системы.

Разнообразие данных является одной из основных характеристик больших данных, что обуславливает рост числа мультимодальных данных. Например, в приложениях электронной коммерции продукт часто описывается одновременно с помощью текста, изображения, видео и т. д. Однако традиционные текстовые поисковые системы в основном ориентированы на текстовые данные и не могут предложить высокопроизводительную индексацию и поиск по мультимодальным данным. В [3] предлагается высокопроизводительная гибридная структура индекса под названием MMH-index для улучшения Apache Lucene, Java-библиотеки с открытым исходным кодом, обеспечивающей мощные функции индексации и поиска с мультимодальной индексацией и поиском. Индекс MMH оптимизирует классический инвертированный индекс в Lucene с помощью инновационного дизайна, называемого растровым изображением модальности. Добавляя растровые изображения модальности в словарь инвертированного индекса, MMH-индекс может уменьшить избыточные данные в

индексе и среднюю длину списков публикаций, тем самым уменьшая потребление места и временные затраты на мультимодальные запросы.

Распределенное хранилище. Поисковые системы реального времени часто полагаются на распределенные системы хранения для хранения и репликации индексированных данных для обеспечения отказоустойчивости и масштабируемости. Для этой цели обычно используются распределенные базы данных, такие как Apache Cassandra, HBase и Amazon DynamoDB.

Оптимизация хранения больших данных необходима для повышения общей эффективности систем больших данных за счет максимального использования ресурсов хранения. Hadoop предоставляет решение для хранения и анализа больших объемов данных, но может столкнуться с проблемами управления хранилищем из-за его распределенного характера и управления большими объемами данных. В [4] предлагается подход, основанный на мультиагентной системе, способной собирать информацию об активности доступа к данным в кластере HDFS (Hadoop Distributed File System). Система классифицирует данные в зависимости от их температуры (горячие или холодные) и принимает решения о репликации данных на основе их классификации.

Алгоритмы ранжирования поиска. Чтобы предоставить пользователям релевантные результаты поиска, поисковые системы реального времени используют сложные алгоритмы ранжирования, которые учитывают такие факторы, как релевантность, новизна, популярность и предпочтения пользователей. Модели машинного обучения часто интегрируются для улучшения ранжирования результатов.

В [5] описывается алгоритм ранжирования RankNet, в которой оптимизирована функция потерь для повышения эффективности ранжирования при поиске. RankNet сравнивался с алгоритмом машины опорных векторов ранжирования (SVM) и традиционным алгоритмом RankNet. Результаты показали, что по мере увеличения количества возвращаемых запросов, эффективность поискового ранжирования всех трех алгоритмов поискового ранжирования имела тенденцию к снижению. При том же количестве возвращенных запросов улучшенный алгоритм RankNet показал лучшую производительность.

Проблемы в архитектуре поисковых систем реального времени

Скорость передачи данных. Обработка высокой скорости входящих потоков данных является серьезной проблемой. Уровни приема должны быть масштабируемыми и отказоустойчивыми, чтобы предотвратить потерю данных или задержки во время пиковых нагрузок.

Задержка. Поисковые системы, работающие в режиме реального времени, должны предоставлять ответы на запросы пользователей с малой задержкой. Минимизация времени обработки запросов при сохранении точности является постоянной проблемой. Современные подходы используют централизованный индекс выборки в агрегаторе для выбора узлов обслуживания индекса (ISN) для поддержания качества и оперативности. В [6] предлагается Cottage - скоординированная структура между агрегатором и ISN - для оптимизации задержек и качества при веб-поиске. В Cottage есть две отдельные модели нейронных сетей на каждом ISN для прогнозирования вклада качества и задержки соответственно. Затем эти результаты прогнозирования отправляются обратно агрегатору для оптимизации задержки и качества. Ключевой задачей является интеграция прогнозов в агрегаторе для определения оптимального динамического бюджета времени для выявления медленных и низкокачественных ISN для улучшения задержки и эффективности поиска

Согласованность данных. Обеспечение согласованности данных в распределенных системах может оказаться сложной задачей. Инженеры должны тщательно проектировать механизмы репликации и синхронизации данных.

Крупномасштабные вычисления с интенсивным использованием данных требуют интеграции сложной технологии с распределенными файловыми системами, чтобы предоставить клиентам надежный и масштабируемый высокопроизводительный доступ к хранимым данным. Клиенты физически совместно используют устройства хранения данных и на этих клиентах распределенные файловые системы берут на себя ответственность за обеспечение скоординированного доступа и согласованного представления общих данных. В [7] представлен протокол распределенной блокировки, который позволяет нескольким узлам одновременно записывать свои данные в отдельные части данных файла, обеспечивая при этом согласованное представление данных, кэшированных клиентом.

Масштабируемость. По мере роста объемов данных поисковые системы реального времени должны масштабироваться горизонтально, чтобы поддерживать производительность. Балансировать между необходимостью в распределенной обработке и локальной может быть непросто.

Масштабируемость — одна из важных целей, решаемых системами обработки распределенных потоков. Системы распределенной потоковой обработки выполняют потоковые приложения, которые сегментируются и распределяются между несколькими узлами в кластерах, чтобы обслуживать интенсивные и растущие сценарии использования потоковой обработки. Из-за преимуществ, предоставляемых облачными инфраструктурами, системы распределенной потоковой обработки часто адаптируются для развертывания в облачных средах. Оркестраторы контейнеров, как правило, обеспечивают модульность и простоту управления контейнерными приложениями. Динамическая масштабируемость и динамическое планирование потоковых приложений становятся ключевыми моментами в обеспечении эффективности облачных систем распределенной обработки потоков. В [8] обобщены исследования, проведенные в области динамического масштабирования и динамического планирования систем потоковой обработки, связанных с облаками.

Актуальность и персонализация. Нахождение баланса между предоставлением актуальной информации и предотвращением информационной перегрузки — деликатная задача. Функции персонализации могут помочь адаптировать результаты к индивидуальным предпочтениям пользователя.

Технологии, управляющие поисковыми системами в реальном времени

Apache Kafka — это распределенная платформа потоковой передачи событий, которая играет решающую роль в приеме данных и потоковой обработке для поисковых систем в реальном времени.

В [9] проведено исследование того, как стабильно и эффективно обмениваться, и собирать данные в географически удаленных независимых сетях с использованием Apache Kafka, и как обеспечить простой анализ таких данных, чтобы пользователи могли эффективно их использовать.

[10] синтезирует и консолидирует существующие знания, экономя драгоценное время и усилия исследователей при поиске соответствующей информации в нескольких источниках. Ощутимые преимущества этого обзора включают предоставление консолидированной базы знаний по исследовательским темам Apache Kafka, освещая

практические идеи и новые подходы, выявляя связанные с ними исследовательские проблемы и служа надежным справочником для сообщества Apache Kafka.

Elasticsearch — это хорошо масштабируемая система поиска и аналитики, работающая в режиме реального времени, которая обеспечивает уровни индексации и запросов многих поисковых систем, работающих в режиме реального времени.

Apache Lucene — это основная технология многих поисковых систем, обеспечивающая мощные возможности индексирования и выполнения запросов.

Машинное обучение. Методы машинного обучения, такие как обработка естественного языка (NLP) и системы рекомендаций, все чаще интегрируются в поисковые системы реального времени для повышения релевантности и персонализации результатов.

Включив NLP в поисковые системы реального времени, эти системы могут лучше понимать и интерпретировать запросы пользователей и содержимое документов. Это приводит к более точным и контекстно-зависимым результатам поиска, что в конечном итоге улучшает взаимодействие с пользователем.

Рекомендательные системы используют алгоритмы машинного обучения для анализа поведения и предпочтений пользователей. Интегрируя системы рекомендаций в поисковые системы в режиме реального времени, эти системы могут предлагать пользователям релевантный контент на основе их прошлых взаимодействий и интересов. Этот аспект персонализации не только повышает релевантность результатов, но и увеличивает вовлеченность пользователей. В целом, интеграция методов машинного обучения, таких как НЛП и систем рекомендаций, в поисковые системы реального времени отражает стремление отрасли предоставлять пользователям узкоспециализированные и контекстуально релевантные результаты поиска, тем самым улучшая общее качество поиска.

Заключение

Поисковые системы, работающие в режиме реального времени, находятся на переднем крае предоставления пользователям мгновенного доступа к самой последней информации из различных источников. Их архитектуры сложны и развиваются, что вызвано постоянной необходимостью сбалансировать скорость, точность и масштабируемость. Поскольку объем и скорость данных продолжают расти, поисковые системы, работающие в режиме реального времени, останутся жизненно важным инструментом как для бизнеса, так и для потребителей, определяя то, как мы взаимодействуем с информацией и получаем к ней доступ в эпоху цифровых технологий.

Литература

1. Erraissi Allae & Belangour, A. Data sources and ingestion big data layers: Meta-modeling of key concepts and features // International Journal of Engineering and Technology(UAE), 2018. - 7. 3607-3612. 10.14419/ijet.v7i4.19653.
2. Walter-Tscharf, Franz Frederik Walter Viktor. Indexing, Clustering, and Search Engine for Documents Utilizing Elasticsearch and Kibana // Lecture Notes on Data Engineering and Communications Technologies, 2022. - PP. 897 – 910, DOI: 10.1007/978-981-19-2069-1_62
3. Liu R., Liang J., Jin P., Wang Y. MMH-index: Enhancing Apache Lucene with High-Performance Multi-Modal Indexing and Searching // Proceedings of the 30th ACM International Conference on Multimedia, 2018. - PP. 7279 – 7289, DOI: 10.1145/3503161.3548768

4. Sais M., Rafalia N., Mahdaoui R., Abouchabaka J. Distributed storage optimization using multi-agent systems in Hadoop // E3S Web of Conferences, 412, 2021. - 01091 DOI: 10.1051/e3sconf/202341201091
5. Zhi S. S., Wang H. H. A search ranking algorithm for web information retrieval, 2023. - 29 (2) - PP. 113 – 124, DOI: 10.1504/IJCNDS.2023.129225
6. Zhou L., Ramakrishnan K. K. Balancing Latency and Quality in Web Search, 2021. - DOI: 10.1109/NAS51552.2021.9605375
7. No J. Data consistency protocol for distributed file systems / 2009. - PP. 253 - 258 DOI: 10.1109/IDAACS.2009.5342984
8. Senthuran A., Hettiarachchi S. A Review of Dynamic Scalability and Dynamic Scheduling in Cloud-Native Distributed Stream Processing Systems // Lecture Notes in Electrical Engineering, 601, 2021. - PP. 1539 - 1553 DOI: 10.1007/978-981-15-1420-3_161
9. Park S., Huh J. A Study on Big Data Collecting and Utilizing Smart Factory Based Grid Networking Big Data Using Apache Kafka // IEEE Access, 2023. - PP. 1-1, DOI: 10.1109/ACCESS.2023.3305586
10. Raptis T. P., Passarella A. A Survey on Networked Data Streaming with Apache Kafka // IEEE Access, 2023. - PP. 85333 - 85350, DOI: 10.1109/ACCESS.2023.3303810

ИДЕНТИФИКАЦИЯ ИМЕНОВАННЫХ СУЩНОСТЕЙ НА ПРИМЕРЕ УЗБЕКСКОГО ТЕКСТА: МЕТОДЫ И ПРИМЕНЕНИЕ

Барахнин В.Б.[1,2], Менглиев Д.Б.[1,3]*
*e-mail: shogunuz@gmail.com
[1]*Новосибирский государственный университет, Россия*
[2]*Федеральный исследовательский центр информационных и вычислительных технологий, Россия*
[3]*Ургенчский филиал Ташкентского университета информационных технологий им. Мухаммада ал-Хоразмий, Узбекистан*

Аннотация. *Статья посвящена разработке и адаптации технологии распознавания именованных сущностей для узбекского языка. В контексте активного развития технологий обработки естественного языка и необходимости создания высококачественных инструментов для многоязычной адаптации, особое внимание уделяется языкам с ограниченными ресурсами, таким как узбекский. В исследовании представлен авторский подход, сочетающий словарные и морфологические методы. Результаты работы способны стать весомым вкладом в расширение инструментария лингвистической обработки узбекского языка.*

Введение

В последние десятилетия наука и технология обработки естественного языка (NLP) активно развиваются, обеспечивая создание совершенно новых инструментов и методик для изучения и обработки человеческой речи. Основной акцент в области NLP традиционно делается на таких языках, как английский, китайский и другие широко распространённые языки, что обусловлено как их глобальным статусом, так и доступностью огромных данных для исследований. Однако важность адаптации технологий обработки естественного языка для менее распространенных и малоизученных языков не может быть недооценена.

Узбекский язык — один из таких менее распространенных языков, который, несмотря на свою богатую историю и культуру, остается относительно малоизученным в контексте современных NLP-технологий [1]. Этот язык имеет свои уникальные лингвистические характеристики, которые требуют особого подхода при создании методов его обработки. Недостаток лингвистических ресурсов и исследовательской инфраструктуры создает определенные трудности для применения стандартных методов NLP.

В свете этого растет актуальность исследований, направленных на адаптацию и разработку новых подходов к обработке узбекского языка [2]. В частности, задача распознавания именованных сущностей (Named Entity Recognition, NER) стоит особенно остро. Эта задача является одной из основных в области NLP и подразумевает определение и классификацию именованных сущностей (например, имен собственных, географических названий, дат) в тексте [3]. Эффективное решение этой задачи необходимо для ряда приложений, таких как автоматический поиск информации, машинный перевод, извлечение информации и многие другие.

Осознавая важность и сложность задачи, проанализировав особенности узбекского языка [4], в данной статье авторы предлагают новый подход к распознаванию именованных сущностей для узбекского языка, сочетающий в себе словарные и морфологические методы. Предлагаемая работа стремится заполнить пробел в литературе по этой теме и предложить эффективный метод, адаптированный к специфике узбекского языка.

Авторы надеются, что результаты исследования будут интересны не только специалистам в области обработки естественного языка, но и всем, кто интересуется узбекским языком и культурой. Помимо практической ценности, работа подчеркивает важность дальнейшего развития исследований в области NLP для языков с ограниченными ресурсами и ставит задачу активизации научного сообщества в этом направлении

Аналогичные работы

Статья [5] подробно рассматривает тонкости распознавания именованных сущностей в контексте обработки естественного языка. Основное внимание уделяется быстрому обнаружению слов, их классификации, пониманию текстового содержания и выделению именованных сущностей. Исследователи детально описывают методы автоматического выявления именованных сущностей в узбекском контенте, демонстрируя конкретные примеры и соответствующие классификации сущностей. Кроме того, авторы рассматривают функционал программного продукта «Анализатор узбекских именованных сущностей», стараясь раскрыть принципы его работы и включенный в него материал.

Тем не менее, статья имеет ряд недостатков. Прежде всего, несмотря на представление программного решения для NER, авторы лишь поверхностно затрагивают его внутреннюю структуру. Остается неясным, какие алгоритмы и инструменты были интегрированы, что создает определенную неясность вокруг основной архитектуры программы. Если программное решение базируется на методах машинного обучения, авторы не уточняют, какой конкретный набор данных был выбран для обучения и валидации. Этот пробел мешает полному пониманию качества выбранного набора данных, его представительности и возможных предвзятостей, что, в свою очередь, влияет на оценку надежности и доверия к системе. Кроме того, рукопись не предоставляет убедительных доказательств эффективности предложенных методов. Отсутствие четко определенных критериев производительности, таких как точность, общее количество рассмотренных узбекских фраз и другие релевантные показатели, порождает сомнения в практической применимости рекомендуемого подхода.

Также стоит отметить, что значительная часть дискуссии об алгоритмах, кажется, сильно опирается на предыдущие работы [6−8]. Содержание статьи во многом повторяет существующие методы, что вызывает вопросы относительно ее уникального вклада в академическое обсуждение в данной области.

Работа [9] глубоко изучает механизмы автоматического извлечения знаний, нацеленные на усиление процесса принятия решений в условиях обилия данных. Исследование включает методики, охватывающие морфологическую оценку текста, представление текстовых документов в виде гиперграфов, а также использование метода случайного блуждания для выявления семантически связанных пар слов. Результаты систематизированы в матрицу, демонстрирующую коэффициенты сродства слов.

Кроме того, эта статья детально раскрывает сложности обучения нейронной сети, задачей которой является извлечение лингвистических структур и определение

потенциальных дескрипторных значений для именованных текстовых сущностей. Этот нейронный механизм способен извлекать информацию на основе выбранного дескриптора, такого как географические указатели, в итоге выдавая географическую номенклатуру в качестве результата.

Центральное место этой статьи − создание алгоритма для извлечения ключевых слов из согласованного корпуса текста на определенную тему, создавая основу для формирования онтологии в конкретной области. Это способствует точному поиску в объемных текстовых корпусах.

Методика, описанная в этой работе, в теории может быть применима к узбекскому языку благодаря его родству тюркоязычному семейству языков. Однако, важно учитывать лингвистические особенности узбекского языка, такие как его сингармонизм, который может оказать значительное влияние на конечные результаты. Добавляет сложности несоразмерное наличие лингвистических ресурсов: языки, такие как казахский, обладают обширными наборами данных для машинного обучения, чего не может похвастаться узбекский. Этот очевидный дефицит ресурсов для узбекского языка подчеркивает актуальность разработки методов, не зависящих от больших наборов данных, для эффективного применения алгоритмов машинного обучения. В результате настоящее исследование представляет альтернативную стратегию, специально разработанную для языков с ограниченными ресурсами, таких как узбекский, с акцентом на NER без необходимости больших объемов данных.

Предлагаемый подход

В представленной статье авторами рассматривается алгоритм распознавания именованных сущностей для узбекского языка, основанный на словарном подходе. Основная последовательность действий алгоритма представлена следующим образом:

1. Исходный текст подвергается предварительной обработке, в ходе которой осуществляется его деление на отдельные предложения, далее происходит токенизация, преобразующая каждое предложение в массив слов.
2. Каждый полученный токен сопоставляется со словарем размеченных слов. В случае успешного сопоставления слова добавляются в результирующий словарь с указанием соответствующей морфологической категории.
3. Слова, которые не были идентифицированы на предыдущем этапе, подвергаются морфологическому анализу, основанному на стемматизации. После этого процесса они повторно сравниваются со словарем размеченных слов. При успешном сопоставлении данные слова также интегрируются в результирующий словарь.
4. В завершение работы алгоритма формируется список идентифицированных слов с указанием их морфологической категории и принадлежности группе именованных сущностей. Данная информация может быть использована для глубокого лингвистического анализа текста или других научных и практических задач.

Этот алгоритм представляет собой авторский подход к задаче распознавания именованных сущностей и обеспечивает эффективную интеграцию словарного метода с методами морфологического анализа, что особенно актуально для языков с ограниченными ресурсами, таких как узбекский.

Оценка результатов

Для оценки эффективности предложенного словарного подхода к распознаванию именованных сущностей в узбекском языке было проведено ряд экспериментов на корпусе текстов, собранном из новостных порталов Узбекистана.

Главной задачей экспериментов было определение доли правильно распознанных именованных сущностей по отношению к общему количеству реальных сущностей в текстах. Важным показателем служила также доля ложных срабатываний, то есть слов и фраз, неверно интерпретированных как именованные сущности.

Основные показатели оценки:

Точность (Precision): Отношение числа верно определенных именованных сущностей к общему числу сущностей, выделенных алгоритмом.

Полнота (Recall): Отношение числа верно определенных именованных сущностей к общему числу именованных сущностей в тексте.

В рамках испытательных работ алгоритма было протестировано 1000 предложений, где алгоритм выявил 2259 сущностей из 2351 имеющихся. При этом из выявленных алгоритмом 2259 сущностей правильно распознанными оказались 2148.

Таким образом, оценка точности равна 95%, тогда как оценка полноты равна 91%.

Результаты тестирования свидетельствовали о значительной эффективности словарного подхода для большинства текстов. Однако были выявлены определенные категории именованных сущностей, с которыми алгоритм справлялся менее успешно. Это, в основном, касалось многокомпонентных имен и фраз, имеющих неоднозначную интерпретацию в разных контекстах.

Словарь, который активно применяется в процессе работы алгоритма, содержит свыше 80 тысячи корней-слов, из которых более 15 тысяч размечены для применения их в задачах распознавания именованных сущностей. Пример размеченного слова: *bola – существительное в единственном числе, персона.*

Схема алгоритма представлена на рис. 1.

Рис. 1 Схема работы алгоритма

Заключение

В представленной статье был описан словарный подход к распознаванию именованных сущностей для узбекского языка. Этот подход, комбинируя классические методы лингвистического анализа с современными технологическими решениями, показал свою эффективность на реальных данных.

Тем не менее, как и любой научный метод, предложенный алгоритм не лишен недостатков и областей для дальнейших улучшений. Некоторые категории именованных сущностей, особенно многокомпонентные и неоднозначные, требуют более детального изучения и доработки.

Дальнейшие исследования в этой области могут включать в себя расширение и уточнение словарных баз, интеграцию дополнительных морфологических и семантических методов, а также адаптацию алгоритма к другим языкам тюркской группы.

Литература

1. Kuriyozov E. Construction and evaluation of sentiment datasets for low-resource languages: the case of Uzbek / E. Kuriyozov, S. Matlatipov, M. A. Alonso, C. Gómez-Rodríguez // Z. Vetulani, P. Paroubek, M. Kubis (eds) Human Language Technology. Challenges for Computer Science and Linguistics. LTC 2019. – Lecture Notes in Computer Science. – Vol. 13212. –DOI:10.1007/978-3-031-05328-3_15
2. Mengliev D. Development of intellectual Web system for morph analyzing of Uzbek words / D. Mengliev, V. Barakhnin, N. Abdurakhmonova // Applied Sciences. – 2021. – Vol. 11, Iss. 9117. – DOI:10.3390/app11199117.
3. Mansurova, M. Named entity extraction model based on the random walk method / M. Mansurova, V. Barakhnin, M. Kyrgyzbayeva, N. Kadyrbek // 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST). – 2021. – Pp. 1–6. – DOI:10.1109/SIST50301.2021.9465992.
4. Abdurakhmonova, N. Z. Developing NLP tool for linguistic analysis of Turkic languages / N. Z. Abdurakhmonova, A. S. Ismailov, D. Mengliev // 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Yekaterinburg, Russian Federation. – P. 1790–1793. – DOI:10.1109/SIBIRCON56155.2022.10017049/
5. Elov, B. Identifying NER (named entity recognition) objects in Uzbek language texts / B. Elov, M.Samatboyeva // Science and innovation international scientific journal, 2023.– Vol. 2, iss. 4. – P. 44–57. – DOI: 10.5281/zenodo.7834009.
6. Manwatkar Ch. A beginner's guide to named entity recognition (NER), online guide, 2020 December. – https://medium.com/swlh/a-beginners-introduction-to-named-entity-recognition-ner-2002b1a010c1 (дата обращения 29.09.2023).
7. Kim S. (J.). Introduction to the medspaCy, the medical Named Entity Recognition(NER) package, online guide, 2023 March. – https://medium.com/geekculture/introduction-to-the-medspacy-the-medical-named-entity-recognition-ner-package-e7c6f0f06496 (дата обращения 29.09.2023).
8. Tang Y. The best way to do named entity recognition (NER), online guide, 2022 September. – https://blog.devgenius.io/the-best-way-to-do-named-entity-recognition-ner-d1d7905bc8ba (дата обращения 29.09.2023).
9. Mansurova M. Named entity extraction from semistructured data using machine learning algorithms / M. Mansurova, V. Barakhnin, Ye. Khibatkhanuly, I. Pastushkov // Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence). – V. 11684. Computational Collective Intelligence. 11th International Conference, ICCCI 2019. Proceedings, Part II. – Cham: Springer, 2019. – P. 58–69. – DOI:10.1007/978-3-030-28374-2_6.

РЕЛЯЦИОННЫЕ МОДЕЛИ ОБРАБОТКИ ТЮРКСКИХ ЯЗЫКОВ

Тукеев У.А.
e-mail: ualsher.tukeyev@gmail.com
Казахский Национальный Университет им. Аль-Фараби, Казахстан

Аннотация. *В данной работе описывается вычислительная модель морфологии, основанная на полных наборах окончаний (CSE – Complete Set of Endings) для тюркских языков. Данная модель позволяет обработку текста естественного языка представить в виде реляционных моделей (таблиц), что обеспечивает простоту и универсальность алгоритмов обработки для ряда задач обработки естественных языков (ОЕЯ). Предлагаемый подход позволяет пользователю использовать универсальные (управляемые данными) алгоритмы и программы для ряда задач ОЕЯ, таких как определение основ слов (стемминг), морфологический анализ текста, сегментация текста, машинный перевод.*

Введение

В данной работе описывается вычислительная модель морфологии, основанная на полных наборах окончаний (CSE – Complete Set of Endings) для тюркских языков. Предлагаемый подход позволяет пользователю (лингвисту) использовать универсальные (управляемые данными) алгоритмы и программы для ряда задач ОЕЯ, таких как определение основ слов (стемминг), морфологический анализ текста и сегментация текста. Одна из ключевых особенностей этого подхода заключается в том, что для нового языка только лингвистический ресурс этого языка в виде полной системы окончаний должен быть подготовлен в форме вычислительной реляционной модели данных. Затем используется универсальная программа для решения соответствующей задачи, основанная на разработанных данных. Этот подход особенно важен для большого количества языков с ограниченными ресурсами, для которых исходных данных (корпусов) для методов машинного обучения все еще недостаточно.

Модели морфологии естественных языков

Существуют три общепринятые модели морфологии естественных языков [1], а именно: «Item and Arrangement – Элемент и расположение» (IA-модель); «Item and Process – Элемент и процесс» (IP-модель); «Word and Paradigm – Слово и парадигма» (WP-модель).

IA-модель фокусируется на агглютинативном характере словоформ. Его основной инструмент моделирования выполняет линейную сегментацию словоформ на морфемы. Морфема — это минимальная значимая неделимая часть слова. Рассматривая морфемы как минимальные единицы грамматического описания, IA-модель хорошо подходит для описания морфологии агглютинативных языков.

IP-модель фокусируется на концепции динамической природы алломорфов, вводя один или несколько уровней представления словоформ. Каждая морфема словоформы обязательно имеет единственное глубокое представление, а также правила перехода к более поверхностным уровням представления с учетом контекста, при котором возможны алломорфные вариации представления морфемы. Таким образом, в зависимости

от контекста поверхностные представления морфем словоформы будут различаться. Рассматривая морфемы и символы глубокого представления («глубокие фонемы») как минимальные единицы грамматического описания, IP-модель позволяет упростить описание морфологии флективных языков.

WP-модель фокусируется на концепции флекции по парадигме. В этой морфологической модели слово рассматривается как единое целое, а не как комбинация основы и окончания. Флексия в WP-модели рассматривается по сходству, а минимальной единицей грамматического описания является словоформа.

На практике при реализации задач обработки естественного языка активно используются модели и методы конечных преобразователей (FST) и методы машинного обучения.

В методах FST современные методы представляют собой двухуровневую (TWOL) морфологическую вычислительную модель [2], которая в основном основана на IP-модели для морфологии. Для реализации этой технологии были разработаны программные средства, которые используются для многих языков. Для использования этих инструментов были разработаны специальные языки пользовательского интерфейса для исходных данных (правила технологии двухуровневой морфологии). Однако освоение и использование пользовательского языка для задания исходных данных для основанных на правилах методов, основанных на двухуровневой морфологии, - довольно трудоемкий процесс. Это является серьезным препятствием для широкого использования лингвистами TWOL технологий, основанных на правилах IP-модели морфологии для стемминга, сегментации и морфологического анализа, особенно для языков с ограниченными ресурсами.

В отличие от предыдущих парадигм, в вычислительной CSE-модели морфологии, предложенной в данной работе, минимальными единицами грамматического описания морфологии являются окончания и основы слов. Окончания слов могут быть представлены последовательностью морфем или, в простейшем случае, одной единственной морфемой. Окончания в CSE-модели рассматриваются как единое целое.

CSE - модель морфологии

Построение CSE-модели морфологии и ее использование для анализа языка основаны на выводе полного набора окончаний для данного языка. Кроме того, в полной версии модели необходимо собрать набор основ слов данного языка, присоединяя к которым возможные окончания можно получить полный набор словоформ языка. Таким образом, для описания морфологии языка необходимо указать либо полный набор словоформ, либо полный набор окончаний и основ. Последний вариант, конечно, предпочтительнее, так как представление морфологии с полным набором окончаний и основ более экономично по объему описания, чем перечисление всех возможных словоформ [3].

Давайте теперь рассмотрим процесс вывода полного набора окончаний на примере казахского языка, относящийся агглютинативным языкам, также, как и все тюркские языки.

Все аффиксы казахского языка можно разделить на два класса: аффиксы к именным основам (существительные, прилагательные, числительные) и аффиксы к основам глаголов (глаголы, причастия, наречия, наклонения, залоги).

Схема вывода окончаний для каждого класса аффиксов рассматривается отдельно. Однако следующая четырехэтапная процедура одинакова для всех случаев:

– определение комбинации возможных размещений основных типов аффиксов;
– выбор размещений основных типов аффиксов (осуществляется путем проверки их семантической приемлемости в языке);
– перечисление возможных вариантов окончаний для каждого варианта семантически приемлемого размещения основных типов аффиксов;
– объединение окончаний в полный набор окончаний для данного языка.

Для каждого нового языка процесс вывода полной системы окончаний необходимо рассматривать полностью с учетом особенностей этого языка.

3.1 Определение возможных размещений основных типов аффиксов.

Рассмотрим сначала схему получения возможных размещений базовых типов аффиксов на примере именных основ [3]. Набор аффиксов к именным основам слов в казахском языке имеет четыре типа: - аффиксы множественного числа (обозначаются K); - притяжательные аффиксы (обозначаются T); - падежные аффиксы (обозначаются C); - личные аффиксы (обозначаются J). Основание будет обозначаться буквой S. Рассмотрим все возможные варианты размещения типов аффиксов: из одного типа, из двух типов, из трех типов и из четырех типов. Количество размещений определяется по формуле (1):

$$A_{nk} = n! / (n-k)! \quad (1)$$

Тогда количество размещений будет определяться следующим образом:

$$A_{41} = 4! / (4-1)! = 4;\ A_{42} = 4! / (4-2)! = 12;\ A_{43} = 4! / (4-3)! = 24;\ A_{44} = 4! / (4-4)! = 24.$$

Следовательно, есть 64 возможных размещения базовых типов аффиксов.

3.2 Определение семантически приемлемых размещений основных типов аффиксов.

Теперь рассмотрим выбор семантически приемлемых размещений типов аффиксов [3]. Типы базовых аффиксов (K, T, C, J) семантически действительны по определению.

Размещение из двух типов базовых аффиксов может быть следующим:

KT, KC, KJ, TC, TJ, TK, CJ, CT, CK, JK, JT, JC.

Анализ семантики размещения из двух типов аффиксов показывает, что допустимы только следующие размещения (выделены жирным шрифтом): KT, TC, CJ, KC, TJ, KJ. Действительно, например, TK - после притяжательных аффиксов, аффиксы множественного числа не используются, CK - после падежных аффиксов, множественные аффиксы не используются, JC - после личных аффиксов, падежные аффиксы не используются, CT - после падежных аффиксов, притяжательные аффиксы не используются, JT- после личных аффиксов притяжательные аффиксы не используются.

Размещение из трех типов аффиксов для семантической приемлимости проверяется по правилу:

если размещение из трех типов аффиксов содержит недопустимые размещения двух типов аффиксов, то это размещение недопустимо.

Тогда, будет 4 приемлемых размещения из трех типов аффиксов (KTC, KTJ, TCJ, KCJ).

Размещение из четырех типов аффиксов для семантической приемлимости проверяется по правилу:

если размещение из четырех типов аффиксов содержит недопустимые размещения из трех типов аффиксов, то это размещение недопустимо.

Допустимое размещение из четырех типов аффиксов - 1 (KTCJ).

Для каждого из полученных вариантов семантически приемлемого размещения основных типов аффиксов производится перечисление возможных вариантов окончаний путем полного перебора. Например, комбинации возможных окончаний для КТ: К * Т = (6 аффиксов К) * (5 аффиксов (разных) Т), что дает 30 окончаний КТ.

Объединение всех языковых окончаний в единый список языковых окончаний дает полный набор окончаний языка.

1. Вычислительные реляционные модели для сегментации и морфологического анализа текстов на основе CSE-модели морфологии

На основе предложенной CSE-модели для морфологии были построены вычислительные реляционные модели для сегментации текста и морфологического анализа, машинного перевода.

4.1 Вычислительная реляционная модель сегментации текста

Вычислительная реляционная модель для сегментации текста представляет собой реляционную (табличную) модель, состоящую из двух столбцов [4]. Первый столбец «окончаний» содержит полный набор языковых окончаний, а второй столбец «сегментированных окончаний» содержит окончания языка, разбитые на аффиксы (Таблица 1).

Таблица 1

Вычислительная реляционная модель для сегментации казахских окончаний (фрагмент).

Окончания слов	Сегментированные окончания	Примеры
gandarmensızder	gan@@dar@@men@@sızder	bar-gandarmensızder (you are wıth people who goıng)
largamyn	lar@@ga@@myn	apa-largamyn (I am to my sısters)
arlardasyndar	ar@@lar@@da@@syndar	aıt-arlardasyndar (You are wıth whom wıll speak)
atyndargamyn	atyn@@dar@@ga@@myn	bar-atyndargamyn (I am to whom who wıll go)
rlergemın	r@@ler@@ge@@mın	tole-rlergemın (I am to whom who wıll pay)

4.2 Вычислительная реляционная модель морфологического анализа текста

Вычислительная модель для морфологического анализа представляет собой реляционную (табличную) модель, состоящую из двух столбцов. Здесь также первый столбец «окончания» содержит полный набор окончаний языка, а второй столбец представляет собой последовательность морфологических характеристик, описывающих морфологический анализ соответствующих окончаний (таблица 2). Для описания морфологических характеристик используются теги системы машинного перевода на основе правил платформы Apertıum [5].

Таблица 2

Вычислительная модель данных для морфологического анализа казахского языка (сегмент таблицы).

Endıngs	Morphologıcal analysıs	Comments
largamyn	<NB>*lar<pl> *ga<dat>*myn<p1>	NB – nomınal base type; pl - plural; dat - datıve case; p1- 1-st person
gandarmensızder	<VB>*gan<pp>*dar <pl>* men<ınst>*sızder<p2><frm>	VB-verbal base; pp-past partıcıple; pl – plural; ınst- ınstrumental case; p2 - 2nd person; frm- formalıty
arsyn	<VB>*ar<fut> *syn<p2>	VB-verbal base; fut – future tense; p2 - 2nd person

На основе предложенного подхода каждый новый язык будет иметь свои собственные вычислительные модели для сегментации и морфологического анализа. В работе [6] описаны универсальные алгоритмы для стемминга, сегментации, морфологического анализа, основанные на CSE-модели морфологии тюркских языков. Данные универсальные алгоритмы и программы основаны на использовании вышеописанных реляционных моделях обработки тюркских языков.

Библиография

1. Плунгян В.А. Общая морфология: Введение в проблематику: Учебное пособие. Изд. 2-е, исправленное // М.: Едиториал УРСС, 2003. - 384 с.
2. Koskennıemı K. Two-level morphology: A general computatıonal model of word-form recognıtıon and production // Tech. rep. Publıcatıon No. 11. Department of General Lınguıstıcs, Unıversıty of Helsınkı. Tukeyev U., Karıbayeva A. Inferrıng the Complete Set of Kazakh Endıngs as a Language Resource // Hernes M., Wojtkıewıcz K., Szczerbıckı E. (eds) Advances ın Computatıonal Collectıve Intellıgence. ICCCI 2020. Communıcatıons ın Computer and Informatıon Scıence, vol 1287. Spırnger, Cham, 2020. - https://doı.org/10.1007/978-3-030-63119-2_60.
3. 4. Tukeyev U., Karıbayeva A., Zhumanov Zh. Morphologıcal Segmentatıon Method for Turkıc Language Neural Machıne Translatıon // Cogent Engıneerıng, Volume 7, 2020 - https://doı.org/10.1080/23311916.2020.1856500
4. Forcada M. L. et al. Apertıum: A free/open-source platform for rule-based machıne translation // Machıne Translatıon 25(2):127-144 pp., DOI: 10.1007/s10590-011-9090-0.
5. 6.Tukeyev U., Karıbayeva A., Turganbayeva A., Amırova D. Unıversal Programs for Stemmıng, Segmentatıon, Morphologıcal Analysıs of Turkıc Words // Nguyen N.T., Ilıadıs L., Maglogıannıs I., Trawıñskı B. (eds) Computatıonal Collectıve Intellıgence. ICCCI 2021. Lecture Notes ın Computer Scıence, vol 12876. Spırnger, Cham. - https://doı.org/10.1007/978-3-030-88081-1_48.

AUTOMATIC SPEECH RECOGNITION FOR THE KAZAKH LANGUAGE

Kairatuly B. , Mansurova M.
email: kairatuly_bauyrzhan1@live.kaznu.kz , madina.mansurova@kaznu.edu.kz
Al-Farabi Kazakh National University, Kazakhstan

Abstract. *We've created a specialized automatic speech recognition (ASR) system for the Kazakh language, which is considered low-resource and presents unique challenges for traditional ASR techniques due to the limited availability of training data and its multilingual nature. To address this, we utilized transfer learning and carefully fine-tuned a set of models derived from pre-trained wav2vec 2.0 and Whisper checkpoints. These models were originally trained on a vast collection of languages and hundreds of thousands of hours of audio data using unsupervised and weak supervised methods. For the fine-tuning process, we employed 332 hours of annotated Kazakh speech data collected from various speakers. The resulting word error rates (WERs) achieved for the wav2vec 2.0 and Whisper models were 19.5 and 18.1, respectively. These low WERs demonstrate the effectiveness of transfer learning in the context of ASR for low-resource languages.*

Keywords: *Automatic speech recognition, wav2vec 2.0, Whisper, Kazakh*

Introduction

Until a relatively short time ago, the development of robust Automatic Speech Recognition (ASR) systems for small and low-resource languages posed an exceedingly daunting challenge. These languages frequently suffered from an insufficiency of data necessary for the training and evaluation of ASR systems, thereby complicating the achievement of substantial recognition accuracy. The primary objective of this study is to thoroughly investigate and convincingly illustrate the methodology employed in developing an ASR system for the Kazakh language, while taking into thorough consideration the precise constraints encountered when dealing with a low-resource language, such as the limited availability of training material.

Over the past few years, various methods to tackle Automatic Speech Recognition (ASR) for the Kazakh language have been introduced, as referenced in sources [1], [2], [3], [4], and [5]. Nonetheless, the latest advancements in machine learning (ML) and natural language processing (NLP) have yielded the capacity to construct considerably more precise ASR systems specifically tailored for small and low-resource languages. In particular, these recent ML developments have resulted in notable enhancements in techniques, including but not limited to, unsupervised and semi-supervised learning methods, which effectively facilitate the transfer of knowledge from larger and well-endowed languages to smaller languages characterized by fewer resources.

Drawing inspiration from the latest developments in ASR research, the present study will be grounded in two of the most promising approaches currently available: wav2vec 2.0 [6], a creation of Meta, and Whisper [7], which has been meticulously crafted by OpenAI. For both of these approaches, substantial pre-trained model checkpoints are readily accessible, and these checkpoints can be meaningfully employed for the purpose of adaptation to a low-resource language, such as Kazakh, with exceptional efficacy.

Recent advancements in ASR

In recent years, notable strides have been made in the domain of Automatic Speech Recognition (ASR) owing to the application of deep learning methodologies. These techniques have demonstrated remarkable efficacy in ASR, leading to swift enhancements in both precision and functionality of ASR systems. Presently, self-supervised transformer-based models tailored for audio processing have substantially elevated the performance standards of ASR systems. These models exhibit a notable capacity to effectively handle unlabeled data, a distinguishing advantage compared to traditional ASR systems, such as the HMM-based approaches exemplified by Kaldi, which demand extensive quantities of labeled data and a pronunciation lexicon for training.

With the advent of pre-trained multilingual base models, significant headway can be achieved in the development of ASR systems for low-resource languages, even in cases where training data is scarce. In this context, we will concentrate our efforts on two highly effective algorithms, namely wav2vec 2.0 and Whisper, both of which offer notable utility in the context of the Kazakh language, primarily because they provide pre-trained checkpoints derived from multi-lingual training.

Wav2vec 2.0 operates on the principle of learning by predicting the contextual information surrounding a given audio segment. This method employs a technique referred to as "time-depth separable convolution," which entails the application of a series of convolutional neural networks (CNNs) to the audio waveform data. CNNs are aptly trained to extract features from audio data at multiple scales. For the acquisition of contextual representations in speech data, both CNNs and recurrent neural networks (RNNs) are harnessed, and these RNNs can be efficiently trained on substantial volumes of unannotated data, making this model particularly advantageous for deployment in low-resource language scenarios characterized by a dearth of annotated data.

Conversely, the Whisper approach utilizes a transformer-based sequence-to-sequence (seq2seq) architecture to glean contextual speech data representations. This approach exhibits robustness to noise and other distortions, a crucial feature when dealing with audio data of suboptimal quality. Noteworthy distinctions between the two models lie in their architectural variances, with wav2vec 2.0 being designed primarily for training on extensive amounts of unannotated audio files, whereas Whisper can seamlessly accommodate both annotated and unannotated data, often referred to as 'weakly supervised' [8]. Both methodologies hold considerable promise for application in low-resource languages such as Kazakh, demonstrating the potential for commendable performance even with relatively scant amounts of annotated data.

It is pertinent to note that a recent study [9] has introduced a wav2vec 2.0 model tailored for the Kazakh language, yielding outcomes akin to the findings in the present study.

Data and methods

In the process of assembling our training dataset, we meticulously curated a crowdsourced open-source speech corpus specifically tailored to the Kazakh language. This corpus, known as the Kazakh Speech Corpus (KSC) [10], encompasses an extensive collection of resources, comprising approximately 332 hours of meticulously transcribed audio data. These audio recordings encompass a vast array of over 153,000 individual utterances, thoughtfully contributed by a diverse cross-section of individuals hailing from various geographical regions, spanning diverse age groups, and representing both genders. To guarantee the utmost quality and authenticity, the KSC dataset underwent rigorous scrutiny and examination by native Kazakh speakers. The primary objective behind assembling this comprehensive dataset is to

facilitate its utilization in the training of automatic speech recognition systems, thereby fostering advancements in this domain.

Regarding the training process, our selection of recipes for fine-tuning wav2vec 2.0 and Whisper models was guided by the comprehensive framework developed by Hugging Face, as detailed in their work [11]. This framework is recognized for providing cutting-edge scripts dedicated to the refinement of these models. The dataset was thoughtfully partitioned into two distinct subsets: an 80% training set and a 20% test set. To prevent overfitting during the fine-tuning process, early stopping mechanisms were judiciously employed.

It is imperative to acknowledge that fine-tuning these highly parameterized models imposes significant computational demands. To further elevate the precision of the wav2vec 2.0 model, a language model was meticulously constructed. This language model was nurtured on a corpus of text data, meticulously curated from diverse sources including parliamentary debates, speeches, radio news articles, commentaries, Wikipedia, and others. This corpus amassed a substantial count of approximately 130 million word tokens. The subsequent n-gram weights derived from this language model were subsequently integrated into the decoding pipeline. Their role therein is to make informed decisions in instances where the probabilities stemming from acoustic recognition are inconclusive. The utilization of such a language model has demonstrably enhanced the error rate by a noteworthy margin, ranging between 2% to 5%.

In the realm of Automatic Speech Recognition (ASR) system evaluation, the quality assessment predominantly hinges on the Word Error Rate (WER). WER quantifies the errors present in the transcribed output, encompassing substitutions (S), deletions (D), and insertions (I). The WER is calculated as the sum of these three error types, divided by the total number of words (N) present in the reference transcription. A lower WER value signifies a superior quality ASR system.

Mathematically, the WER can be expressed as:

$$WER = \frac{S+D+I}{N} * 100$$

This metric serves as a robust tool for comparative analysis of different ASR models and provides valuable insights into potential avenues for system enhancement and refinement.

Results

In this section, we will present the outcomes from our analysis of four meticulously fine-tuned models. Among these models, two of them are constructed upon the wav2vec 2.0 XLS-R architecture, boasting parameter counts of 300 million and a staggering 1 billion. Our journey with the Whisper model initially involved experiments with the relatively diminutive Whisper checkpoints, specifically the small and medium variants. However, our initial trials did not yield the desired outcomes. Consequently, we shifted our focus to the Whisper-large checkpoint, equipped with a formidable 1.5 billion parameters.

The performance metrics for all four of our models, as well as the original Whisper model, are meticulously documented in Table 2. These metrics, based on rigorous testing against a comprehensive dataset, consistently register Word Error Rates (WERs) below the 20% threshold, firmly positioning them within the realm of state-of-the-art systems for major languages. Notably, the wav2vec 2.0 model with 1 billion parameters demonstrates a marked superiority over its 300 million-parameter counterpart.

Furthermore, we incorporated a language model into the equation, resulting in a further reduction in WERs by several percentage points. Impressively, the lowest WER recorded across

all models stands at an impressive 9.5%. The performance statistics for the uncased version of the Whisper model exhibit a remarkable degree of similarity. However, when scrutinizing the cased variant, we observe a substantial increase in WER. This discrepancy is likely attributed to challenges stemming from punctuation and capitalization intricacies.

Table 1.

WERs for Kazakh ASR models trained with and without a language model.

Model	WER w/o language model	WER w. language model
wav2vec 2.0 (300m)	23.5	21.5
wav2vec 2.0 (1b)	20.7	19.5
Whisper-large uncased	18.1	n.a.
Whisper-large-v2 cased	19.6	n.a.

To illustrate the proficiency of these models, we have thoughtfully included two more extensive examples within Table 2, accompanied by their authentic, unaltered transcriptions. When scrutinizing the text, it becomes abundantly clear that both models exhibit an exceptional level of accuracy, with discrepancies predominantly confined to uncommon vocabulary choices, less emphasized function words, or the closing segments of words.

Table 2.

Recognition samples: Ground truth and outputs from wav2vec 2.0 and Whisper models.

Ground truth	Wav2vec 2.0	Whisper-large-v2
мұның ар жағында менің күдігімді растай түсетін дәлелдер жасырынып жатыр әрине өкінішке орай ол жайлап орнынан тұрды	мұнаң әр жағында менің күдігімды түсетін дәлелдер жасырынып жатыр әрине өкінішке ұрай ол жайлап орнанан тұрды	Мынан әр жағындай менің күдігімді растай түсетін делилтер жастырынып жаттыр әрине, үкіншікке орай ол жайлап орднынан тұрды.

In stark contrast to the wav2vec 2.0 model, the Whisper model excels in handling multilingual speech with remarkable efficacy. This advantage can be attributed to Whisper's comprehensive cross-lingual pre-training approach, which draws upon both audio and textual datasets, granting it superior capabilities in handling multilingual input when compared to wav2vec 2.0. While the latter model has undergone training across a plethora of languages, its multilingual adaptability is predominantly rooted in acoustic characteristics and does not extend to textual representations. Consequently, the quality of its output hinges substantially on the specific composition of the training data utilized during the fine-tuning phase.

Conclusion

In conclusion, this study has successfully developed a specialized automatic speech recognition (ASR) system tailored for the Kazakh language, a low-resource language presenting unique challenges for conventional ASR approaches. By harnessing the power of transfer learning and fine-tuning models derived from pre-trained wav2vec 2.0 and Whisper checkpoints, originally trained on a diverse array of languages and massive audio datasets, we

were able to address these challenges effectively. Our fine-tuned models, utilizing 332 hours of annotated Kazakh speech data, achieved impressive word error rates (WERs) of 19.5 for wav2vec 2.0 and 18.1 for Whisper. These remarkably low WERs underscore the efficacy of transfer learning in the realm of ASR for low-resource languages. This research contributes to advancing ASR capabilities in linguistically diverse and resource-constrained contexts.

References

1. O.Khomitsevich, V.Mendelev, N.Tomashenko, S.Rybin, I.Medennikov, S.Kudubayeva. A bilingual Kazakh-Russian system for automatic speech recognition and synthesis // Proceedings of the International Conference on Speech and Computer, Springer: Berlin/Heidelberg, 2015. - pp. 25-33.
2. A.Dawel, Y.Nurmemet, Y.Liu. On language model construction for LVCSR in Kazakh // Comput. Eng. Appl. 2016. – P.178-181.
3. O.Mamyrbayev, M.Turdalyuly, N.Mekebayev, K.Mukhsina, A.Keylan, B.BabaAli, G. Nabieva, A.Duisenbayeva, B.Akhmetov. Continuous speech recognition of Kazakh language // Proceedings of the ITM Web of Conferences. EDP Sciences, Timisoara, Romania, 2019. - Volume 24, p. 01012.
4. O.Mamyrbayev, M.Turdalyuly, N.Mekebayev, K.Alimhan, A.Kydyrbekova, T.Turdalykyzy, Automatic recognition of Kazakh speech using deep neural networks // Proceedings of the Asian Conference on Intelligent Information and Database Systems, Yogyakarta, Indonesia, 2019. - pp. 465-474.
5. B.Amirgaliyev, D.Kuanyshbay, O.Baimuratov, M.Kutubayeva. Development of Automatic speech recognition for Kazakh language using transfer learning // Int. J. Adv. Trends Comput. Sci. Eng. 2020. – Vol.9, P. 5880-5886. http://dx.doi.org/10.30534/ijatcse/2020/249942020
6. A. Baevski, Y. Zhou, A. Mohamed, and M. Auli. Wav2vec 2.0: A framework for self-supervised learning of speech representations // Advances in Neural Information Processing Systems, 2020.- vol. 33, pp. 12 449–12 460.
7. A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, Robust speech recognition via large-scale weak supervision // arXiv preprint arXiv:2212.04356, 2022.
8. J. C. Vásquez-Correa and A. Álvarez, Novel Speech Recognition Systems Applied to Forensics within Child Exploitation: Wav2vec2.0 vs. Whisper // ENGINEERING, Preprint, Dec. 2022.
9. K. Bauyrzhan, M. Madina and O. Assel, Fine-Tuning the Wav2vec2 Model for Kazakh Speech: A Study on a Limited Corpus // 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, 2023. - pp. 124-128, doi: 10.1109/SIST58284.2023.10223504.
10. Y. Khassanov, S. Mussakhojayeva, A. Mirzakhmetov, A. Adiyev, M. Nurpeiissov and H. A. Varol, "A crowdsourced open-source Kazakh speech corpus and initial speech recognition baseline", Proc. of the Conference of the European Chapter of the Association for Computational Linguistics, pp. 697-706, 2021. - https://aclanthology.org/2021.eacl-main.58.
11. T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz et al., Transformers: State-of-the-art natural language processing // Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, 2020.- P. 38–45.

ЕДА - ПОДГОТОВКА ДАННЫХ ДЛЯ МАШИННОГО ОБУЧЕНИЯ И ПРОГНОЗИРОВАНИЯ ДИНАМИКИ ВРЕДИТЕЛЕЙ ПШЕНИЦЫ

[1]Аканова А.С., [1]Шарипова С.Е., [2]Оспанова Н.Н., [2]Казанбаева А.С., [1]Анарбекова Г.А.
e-mail: akerkegansaj@mail.ru
[1]Казахский агротехнический исследовательский университет имени С. Сейфуллина, Астана, Казахстан
[2]Toraighytov University, Павлодар, Казахстан

Аннотация. *В данной статье рассматривается применение методов EDA (Exploratory Data Analysis) для анализа и предварительной обработки данных перед обучением нейронной сети для прогнозирования динамики численности вредителей (Phyllotreta vittula) в Казахстане. В работе использовались данные о численности вредителей, а также метеорологические данные за период с 2013 по 2022 годы. Были проведены исследования, позволяющие выявить зависимости между численностью вредителей и метеорологическими условиями, а также определить наиболее значимые факторы, влияющие на динамику численности вредителей. А также была проведена стандартизация данных и работа с выбросами.*

В настоящее время хлебная блошка является одним из наиболее опасных вредителей зерновых культур в Казахстане.

Полосатая хлебная блошка предпочитает яровые пшеницу и ячмень, озимую рожь, просо, рис. Озимую пшеницу блошка почти не трогает. Насекомое обитает на северо-западе России, в центральноевропейских регионах страны, в Сибири и на Кавказе. В 2021 году хлебная блошка активно поражает посевы зерновых в приволжских регионах.

В отличие от множества других вредителей опасность для посевов представляют взрослые жуки; личинки же вреда не приносят. Жуки хлебной блошки поедают мякоть листьев зерновых, в которой запасены питательные вещества, соскребая её с наружней стороны листа. На листьях остаются прозрачные или пожелтевшие полосы. Растения желтеют или становятся бледно-зелёными, засыхают. Наиболее опасны хлебные блошки в фазе до кущения.

При повреждении 50% поверхности листьев урожай падает на 20%. Куст рождает меньше колосьев, в них развивается меньше зёрен. Если поражено 70% поверхности листьев, растение гибнет. Численность хлебной блошки, при которой экономический ущерб урожаю начинает превышать затраты на борьбу с вредителем – 20-30 жуков на 1 м² в фазе всходов в засушливые годы и 50-60 во влажные. Оценивать численность можно с помощью сачка или клеевой ловушки. Можно провести контроль визуально, оценивая повреждения листьев.

В связи с этим, возникает необходимость в разработке эффективных методов борьбы с этим вредителем. В данной статье рассматривается применение методов EDA (Exploratory Data Analysis) для обучения нейронной сети для прогнозирования динамики численности вредителей (Phyllotreta vittula) в Казахстане. Для этого были использованы данные о численности вредителей и метеорологические данные за период с 2004 по 2022 годы. Целью исследования является подготовка данных для обучения нейронной сети

для прогнозирования динамики численности вредителей на основе метеорологических данных. Результаты данной работы могут быть использованы для разработки аналитической системы мониторинга численности вредителей и принятия решений в области сельского хозяйства.

Растительные болезни приводят к убыткам в сельском хозяйстве и, следовательно, в экономике. Это требует создания моделей прогнозирования для выявления и оценки болезней растений. Грибковые инфекции, являющиеся наиболее распространенными болезнями, могут быть успешно контролированы при условии их раннего обнаружения. В работе [1] представлена разработка экспертной системы для прогнозирования различных грибковых болезней растений (порошковая мильт, антракноз, ржавчина и гниль корней/поражение листьев). Для классификации болезней используется многослойная нейронная сеть, которая не только эффективно обнаруживает болезни растений, но также может существенно увеличить урожайность. Предложенная техника включает в себя три важных этапа: предварительную обработку набора данных, исследовательский анализ данных и модуль обнаружения. Экспериментальные результаты показывают, что модель превосходит несколько существующих методов по точности. Средняя точность прогнозирования каждой болезни составляет более 98%. Эта работа также подтверждает возможность использования данной техники для более быстрого обнаружения болезней растений по доступной цене.

Применение EDA-анализа для предварительного анализа входящих данных является первым шагом при разработке аналитической системы прогнозирования динамики численности вредителей в Казахстане на основе нейронных сетей. Всестороннее изучение и анализ входящих данных, а также подготовка их для внесения в нейронную сеть является очень важным шагом.

Так, в работе [2] представлена новая методология SubTab, разработанная для повышения эффективности исследования данных путем создания кратких и информативных подтаблиц из больших наборов данных. SubTab обеспечивает создание подтаблиц, обозначаемых как T_sub, состоящих из уменьшенного числа k строк и l столбцов, извлеченных из большой таблицы данных размером n x m, где k и l тщательно выбираются для выявления значимых паттернов данных как внутри, так и между столбцами, тем самым упрощая процесс исследования данных. Кроме того, данное исследование демонстрирует универсальность SubTab в ускорении интерпретации результатов запросов, что позволяет пользователям быстро осознавать результаты и принимать информированные решения относительно последующих запросов, повышая тем самым удобство и эффективность рабочего процесса анализа данных.

Здравоохранительная индустрия генерирует огромное количество данных, так называемых "больших данных", которые содержат скрытые знания или закономерности для принятия решений. Огромный объем данных используется для принятия решений, которые более точны, чем интуиция. Анализ EDA обнаруживает ошибки, находит подходящие данные, проверяет предположения и определяет корреляцию между объясняющими переменными. В рассматриваемом исследовании [3] EDA рассматривается как анализ данных, исключающий выводы и статистическое моделирование. В данной работе рассматривается использование анализа данных для прогнозирования сердечных заболеваний с использованием алгоритма K-средних на наборе данных с 209 записями и 8 атрибутами. Результаты показывают точность прогнозирования. Также в работе [4] было рассмотрено применение алгоритмов машинного обучения для улучшения точности прогнозирования сердечной недостаточности путем использования процедур предварительной обработки данных, стандартизации и настройки гипер-

параметров, что подтверждается результатами эксперимента на стандартном наборе данных о сердечных заболеваниях.

Ученые в исследовании [5] фокусируются на анализе финансовых бухгалтерских операций с использованием методов исследовательского анализа данных для определения прибыльности и на применении прогностической аналитики для прогнозирования общей выручки, используя различные модели машинного обучения с учителем, включая дерево решений. Результаты исследования указывают на эффективность дерева решений как модели для анализа производительности в этой области.

В работе [6] исследуется характеризация влияния случайно распределенных сланцевых барьеров на процесс добычи нефти и газа в различных типах месторождений. Проводится построение 2D резервуарной модели с различными сценариями случайных сланцевых барьеров и извлечение уникальных параметров для их параметризации. Затем разрабатывается нелинейная модель с использованием алгоритма машинного обучения для прогнозирования коэффициента извлечения нефти, что демонстрирует высокую точность в сравнении с результатами численного термального моделирования резервуара.

Исследование [7] о преступлениях против женщин в Индии за период с 2001 по 2020 годы, проведенное с использованием методов EDA и линейной регрессии, показывает, что насилие в отношении женщин остается на высоком уровне, и оно сильно зависит от различных факторов. Предложенная модель анализирует данные и прогнозирует преступления против женщин в различных категориях IPC с высокой точностью.

В работе [8] рассматривается обнаружение болезней листьев растений с использованием методов обработки изображений и машинного обучения. Для этой цели в исселдовании используются сверточные нейронные сети (CNN) и глубокие сверточные нейронные сети (DCNN).

Исследование [9] рассматривает новый метод машинного обучения, объединяющий кластеризацию без учителя (K-средних) и ансамблевые методы машинного обучения (случайный лес) для прогнозирования упругих логов в сложной геологической среде нижнего Индусского бассейна Пакистана, с точностью на уровне 98% на реальных данных. Такую же высокую точность показывает работа [10], где были использованы инструменты анализа данных и машинного обучения в Python, примененных к набору данных Iris, позволящие создать оптимизированную модель для будущих прогнозов.

Работа [11] представляет исследование, основывающееся на анализе данных о качестве воды в реке Бхавани за пять лет. Используя различные методы визуализации данных, была проведена исследовательская аналитика для выявления источников вариации в качестве воды и подготовки данных для построения прогнозной модели.

В исследовании [12] рассматривается влияние искусственного интеллекта как нового общего метода изобретения на экономику и организацию научно-исследовательской деятельности. Авторы обсуждают историю развития искусственного интеллекта, а также представляют анализ изменений в инновационной деятельности, связанных с глубоким обучением, и делают выводы о возможных изменениях в организации процесса инноваций и необходимости разработки политических мероприятий для поддержки этого нового метода изобретения.

Исследование [13] фокусируется на прогнозировании потребления энергии и производства возобновляемой энергии на основе данных за одиннадцать лет с использованием рекуррентных нейронных сетей (RNN). Оно исследует взаимосвязь между потреблением энергии и производством возобновляемой энергии и может быть использовано для прогнозирования будущих сценариев для различных обществ. А в

работе [14] рассматривается разработка модели глубокой сверточной нейронной сети (DCNN) для идентификации и классификации болезней виноградных кустов на основе RGB-изображений листьев. Предложенная модель показала высокую точность в 99,34% при идентификации и классификации общих болезней виноградных кустов по изображениям листьев, что делает ее полезной для точного земледелия и применения в реальном времени на мобильных устройствах. Похожая работа [15] сосредотачивается на прогнозировании продаж мебели в розничном магазине с использованием различных моделей прогнозирования, включая классические методы временных рядов и более современные методы, такие как LSTM и CNN. Результаты исследования показывают, что метод Stacked LSTM продемонстрировал лучшую производительность среди других методов, а также выделяются хорошие результаты моделей Prophet и CNN.

В работе [16] рассматривается оценка урожайности сельскохозяйственных культур с использованием данных спутникового наблюдения и методов машинного обучения, особенно нейронных сетей. В результате данного исследования было обнаружено, что современные технологии спутниковой связи и методы машинного обучения могут предоставить доступные и точные решения для прогнозирования урожая. Исследование [17] представляет два новых метода прогнозирования временных рядов на основе эволюционных искусственных нейронных сетей (EANN) с использованием алгоритма оценки распределения (EDA) для оптимизации структуры нейронных сетей и выбора временных задержек. Эти методы, названные SEANN и TEANN, продемонстрировали лучшие результаты в прогнозировании по сравнению с другими методами, при этом они способствуют созданию более простых моделей нейронных сетей и требуют меньше вычислительных ресурсов.

Рассмотренные исследования успешно использовали первичный этап анализа данных, в ходе которого исследователи изучают данные, выявляют их основные характеристики и структуру. Этот процесс помогает понять данные, выявить аномалии, обнаружить корреляции и распределения, и определить, какие дополнительные шаги могут потребоваться для более глубокого анализа.

Таким образом, для обучения нейронной сети прогнозирования динамики численности вредителей в Казахстане были использованы такие входные данные, как год, местность, температура воздуха в рассматриваемый период, площадь местности, и целевая переменная – заселенность хлебной полосатой блошки на квадратный метр (zasel_o) с 2013 года по 2022 год.

Для оценки данных были выявлены статистические показатели данных, такие как среднее значение, квартили (25%, 50%, 75%), минимальные и максимальные значения по каждому признаку, которые указаны в рисунке 1.

	t_april	t_may	t_june	t_july	t_august	t_september	audan_v	zasel_v	audan_l	zasel_l	audan_o	zasel_o
count	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000	88.000000	88.000000	100.000000	100.000000	100.000000	100.000000
mean	9.416345	18.645439	25.154054	26.594212	24.289202	15.956876	0.108523	0.066989	0.215800	0.140090	0.093300	0.048370
std	3.445187	3.455032	2.975544	2.818422	2.591275	2.707258	0.034923	0.057687	0.243651	0.251891	0.045528	0.057414
min	-0.066667	10.746000	18.550000	19.931500	18.399200	10.416700	0.100000	0.000000	0.050000	0.000000	0.040000	0.000000
25%	6.892708	16.534250	23.170800	24.594725	22.331625	13.839550	0.100000	0.020000	0.100000	0.015000	0.050000	0.000000
50%	9.533335	18.643150	25.260450	26.651200	23.965700	15.948950	0.100000	0.100000	0.100000	0.080000	0.100000	0.030000
75%	12.053150	21.676400	27.428100	28.460650	26.625975	17.862525	0.100000	0.100000	0.200000	0.100000	0.100000	0.100000
max	17.300000	25.866900	30.741700	31.866900	29.044400	22.037500	0.300000	0.300000	1.600000	1.600000	0.300000	0.300000

Рис. 1 Статистические показатели данных

После анализа статистических показателей было выявлено, что данные признаков имеют очень разные форматы, поэтому было принято решение стандартизировать данные. Для этой цели была использована библиотека Python StandardScaler (from sklearn.preprocessing import StandardScaler):

```
scaler = StandardScaler()
df= scaler.fit_transform(df)
```

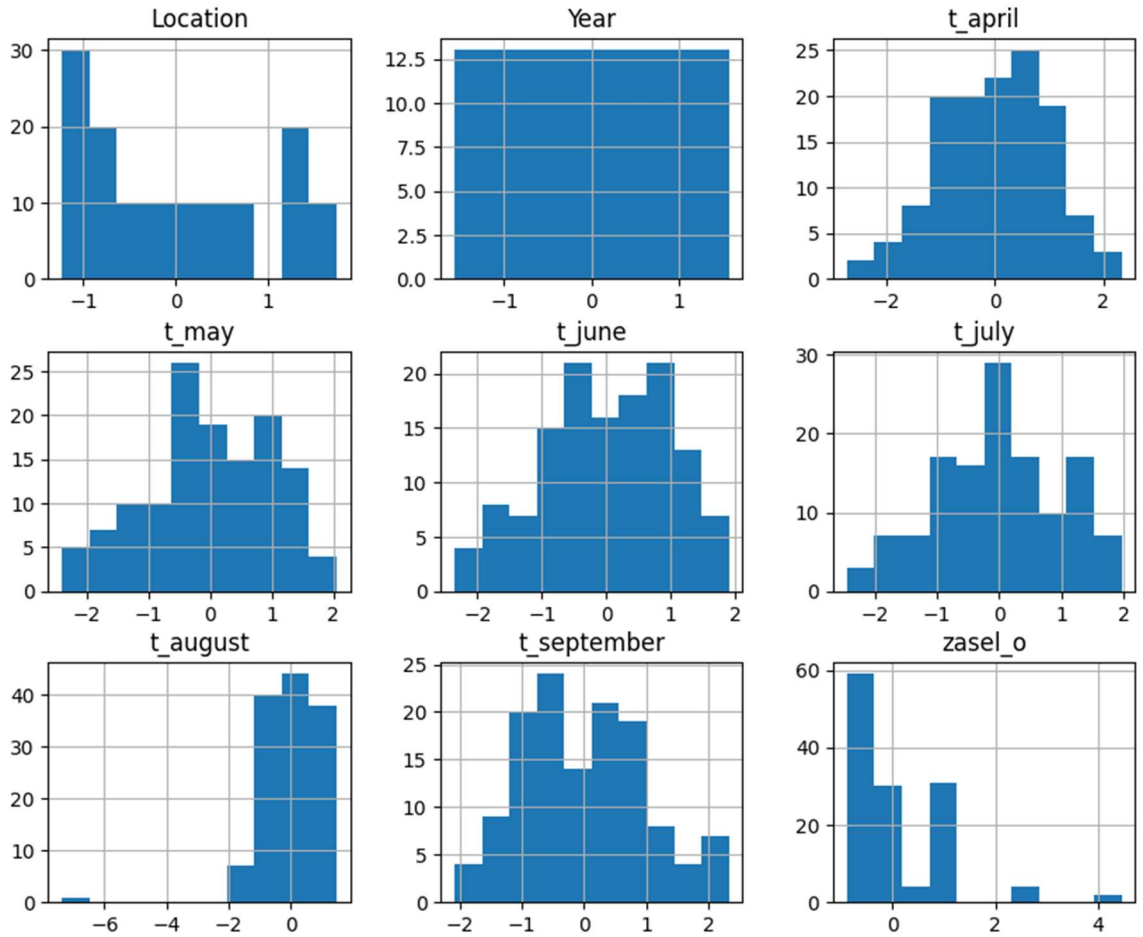


Рис. 2 Гистограммы признаков

Стандартизация позволяет привести все признаки к одному масштабу, что делает их сравнение и интерпретацию более надежными.

Для анализа нормальности распределения данных были построены гистограммы для всех признаков, которые отображены на рисунке 2.

Из представленных гистограмм можно сделать вывод, что есть выбросы и отклонения. Дополнительно были построены графики «ящик с усами», для точного определения выбросов. Как видно из рисунка 3, они есть в таких признаках, как t_august и zasel_o. Данные выбросы были удалены.

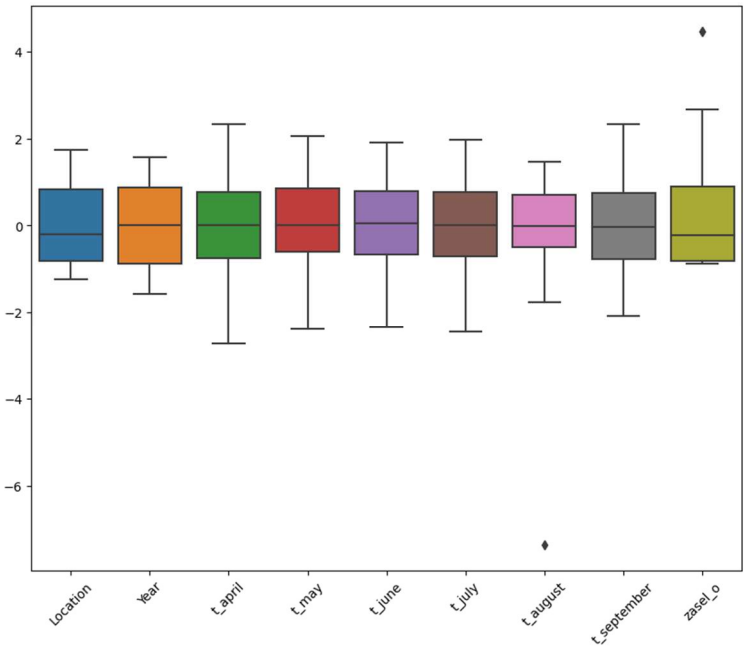


Рис. 3 Оценка выбросов в датасете

Таким образом, проанализированные и обработанные данные готовы к подаче на обучение в нейронную сеть для прогнозирования динамики численности вредителе в Казахстане.

Благодарности

Статья подготовлена в рамках проекта AP19675312«Аналитическая система прогнозирования динамики численности вредителей зерновых культур в Казахстане на основе нейросетевой модели» на базе КАТИУ имени С.Сейфуллина

Библиография

1. Kumar M., Kumar A., Palaparthy V. S. Soil sensors-based prediction system for plant diseases using exploratory data analysis and machine learning //IEEE Sensors Journal. – 2020. – Т. 21. – №. 16. – С. 17455-17468.
2. Amsterdamer Y. et al. Selecting Sub-tables for Data Exploration //2023 IEEE 39th International Conference on Data Engineering (ICDE). – IEEE, 2023. – С. 2496-2509.
3. Indrakumari R., Poongodi T., Jena S. R. Heart disease prediction using exploratory data analysis //Procedia Computer Science. – 2020. – Т. 173. – С. 130-139.
4. Jain V., Agrawal M. Heart Failure Prediction Using XGB Classifier, Logistic Regression and Support Vector Classifier //2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT). – IEEE, 2023. – С. 1-5.
5. Chakri P. et al. An exploratory data analysis approach for analyzing financial accounting data using machine learning //Decision Analytics Journal. – 2023. – Т. 7. – С. 100212.
6. Kumar A., Hassanzadeh H. A qualitative study of the impact of random shale barriers on SAGD performance using data analytics and machine learning //Journal of Petroleum Science and Engineering. – 2021. – Т. 205. – С. 108950.
7. Gupta U., Sharma R. Analysis of criminal spatial events in india using exploratory data analysis and regression //Computers and Electrical Engineering. – 2023. – Т. 109. – С. 108761.

8. Nandhini S. et al. Analysis on classification and prediction of leaf disease using deep neural network and image segmentation technique //Annals of the Romanian Society for Cell Biology. – 2021. – Т. 25. – №. 6. – С. 9035-9041.
9. Ali M. et al. Reservoir characterization through comprehensive modeling of elastic logs prediction in heterogeneous rocks using unsupervised clustering and class-based ensemble machine learning //Applied Soft Computing. – 2023. – С. 110843.
10. Singh J. et al. Exploratory Data Analysis for Interpreting Model Prediction using Python //2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON). – IEEE, 2022. – С. 1-6.
11. Nair J. P., Vijaya M. S. Exploratory Data Analysis of Bhavani River Water Quality Index Data //Proceedings of International Conference on Communication and Computational Technologies: ICCCT 2022. – Singapore : Springer Nature Singapore, 2022. – С. 971-987.
12. Cockburn I. M., Henderson R., Stern S. The impact of artificial intelligence on innovation: An exploratory analysis //The economics of artificial intelligence: An agenda. – University of Chicago Press, 2018. – С. 115-146.
13. Yazdan M. M. S., Saki S., Kumar R. Untangling Energy Consumption Dynamics with Renewable Energy Using Recurrent Neural Network //Analytics. – 2023. – Т. 2. – №. 1. – С. 132-145.
14. Math R. K. M., Dharwadkar N. V. Early detection and identification of grape diseases using convolutional neural networks //Journal of Plant Diseases and Protection. – 2022. – Т. 129. – №. 3. – С. 521-532.
15. Ensafi Y. et al. Time-series forecasting of seasonal items sales using machine learning–A comparative analysis //International Journal of Information Management Data Insights. – 2022. – Т. 2. – №. 1. – С. 100058.
16. Khalil Z. H., Abdullaev S. M. Neural network for grain yield predicting based multispectral satellite imagery: comparative study //Procedia Computer Science. – 2021. – Т. 186. – С. 269-278.
17. Donate J. P., Cortez P. Evolutionary optimization of sparsely connected and time-lagged neural networks for time series forecasting //Applied Soft Computing. – 2014. – Т. 23. – С. 432-443.

Сведения об авторах

Каз	Рус	Англ
1	2	3
Аканова Акерке Сапаровна PhD Компьютерлік ғылымдар кафедрасы, аға оқытушы	Аканова Акерке Сапаровна PhD старший преподаватель кафедры «Компьютерные науки»,	Akanova Akerke Saparovna PhD S.Seifullin Kazakh Agrotechnical University
С. Сейфуллин атындағы Қазақ агротехникалық университеті, Жеңіс көш. 62, Астана қ., Қазақстан Республикасы 010000 Компьютерлік ғылымдар кафедрасы, аға оқытушы akerkegansaj@mail.ru +7 705 448 06 80	Казахский агротехнический университет имени С. Сейфуллина, ул. Жениса 62, г. Астана, Республика Казахстан 010000 akerkegansaj@mail.ru +7 705 448 06 80	Department of Computer Science, Senior lecturer, Senior lecturer 010000, Zhenis St. 62, Astana, Republic of Kazakhstan akerkegansaj@mail.ru +7 705 448 06 80

1	2	3
Оспанова Назира Нұрғазықызы Профессор, педагогика ғылымдарының кандидаты, доцент Торайғыров университеті Павлодар 140000 Қазақстан Республикасы nazira_n@mail.ru +77011664573	Оспанова Назира Нургазыевна Профессор, кандидат педагогических наук, доцент Торайгыров университет Павлодар Республика Казахстан nazira_n@mail.ru +77011664573	Ospanova Nazira Nurgazyevna Professor, Candidate of Pedagogical Sciences, Associate Professor Toraighyrov University Pavlodar 140000 Republic of Kazakhstan nazira_n@mail.ru +77011664573
Казанбаева Альбина Советовна Постдокторант, PhD докторы Манаш Қозыбаев атындағы Солтүстік Қазақстан университеті Петропавл қ. 150000 Қазақстан Республикасы akazanbaeva83@mail.ru +77022727018	Казанбаева Альбина Советовна Постдокторант, доктор PhD Северо-Казахстанский университет имени Маната Козыбаева г. Петропавловск 150000 Республика Казахстан akazanbaeva83@mail.ru +77022727018	Kazanbayeva Albina Sovetovna Postdoctoral fello, PhD North Kazakhstan University named after Manash Kozybayev Petropavlovsk 150000 Republic of Kazakhstan akazanbaeva83@mail.ru +77022727018
Шарипова Салтанат Еркиновна С. Сейфуллин атындағы Қазақ агротехникалық университеті Жеңіс көш. 62, Астана қ., Қазақстан Республикасы, 010000 salta_pvl@bk.ru +77759060037	Шарипова Салтанат Еркиновна Казахский агротехнический университет имени С. Сейфуллина, ул. Жениса 62, г. Астана, 010000 Республика Казахстан salta_pvl@bk.ru +77759060037	Saltanat Sharipova S. Seifullin Kazakh Agrotechnical Research University Astana S.Seifullin Kazakh Agrotechnical University Zhenis St. 62 010000, Astana, Republic of Kazakhstan salta_pvl@bk.ru +77759060037
Анарбекова Галия Абаевна С. Сейфуллин атындағы Қазақ агротехникалық университеті, докторант Жеңіс көш. 62, Астана қ., Қазақстан Республикасы, 010000 galiya2810@mail.ru +7 700 596 53 54	Анарбекова Галия Абаевна Казахский агротехнический университет имени С. Сейфуллина, кафедра «Компьютерные науки», докторант ул. Жениса 62, г. Астана, 010000 Республика Казахстан galiya2810@mail.ru +7 700 596 53 54	Anarbekova Galiya Abayevna S.Seifullin Kazakh Agrotechnical University PhD student Zhenis St. 62 010000, Astana, Republic of Kazakhstan galiya2810@mail.ru +7 700 596 53 54

РАЗВИТИЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СОВРЕМЕННЫХ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМАХ

Жеңісбек Ә.А.
e-mail:adilbek.zhenisbek1@gmail.com
Институт автоматики иинформационных технологий КазНИТУ

Аннотация. *Развитие технологий искусственного интеллекта оказывает значительное воздействие на информационно-телекоммуникационные системы. В данной статье рассматривается актуальная тема развития искусственного интеллекта в информационно-телекоммуникационных системах и его влияние на информационную безопасность. Авторы анализируют вызовы, возникающие при интеграции искусственного интеллекта, представляют методы обеспечения информационной безопасности в условиях использования искусственного интеллекта, а также рассматривают этические аспекты.*

Статья также представляет результаты экспериментов, демонстрирующие эффективность использования искусственного интеллекта в информационно-телекоммуникационных системах.

Актуальность. В современном мире информационно-телекоммуникационные системы играют ключевую роль в различных сферах, включая экономику, здравоохранение, и научные исследования. С ростом объемов данных и требований к вычислительным ресурсам, технологии искусственного интеллекта становятся незаменимыми инструментами для оптимизации и повышения эффективности этих систем. Однако, этот переход к использованию искусственного интеллекта вызывает существенные вызовы и вопросы в области информационной безопасности.

Целью данной статьи является анализ развития технологий искусственного интеллекта в информационно-телекоммуникационных системах, выявление проблем и рисков, связанных с этим развитием, и представление методов обеспечения информационной безопасности при использовании искусственного интеллекта.

Современные проблемы применения искусственного интеллекта в информационных системах. Применение искусственного интеллекта в информационных системах стало активно развиваться в последние десятилетия. С этим развитием появились новые проблемы, требующие внимания и решения:

- Интеграция с существующей инфраструктурой: Переход к использованию искусственного интеллекта в существующих системах требует глубокой интеграции с существующей инфраструктурой. Это может привести к проблемам совместимости и несоответствия стандартам.
- Высокая доступность и надежность: Системы, использующие искусственного интеллекта, требуют высокой доступности и надежности. Сбои в работе искусственного интеллекта могут иметь серьезные последствия, поэтому разработка методов обнаружения и восстановления при ошибках становится критически важной.

Искусственный интеллект и информационная безопасность. Использование искусственного интеллекта также имеет значительное влияние на информационную

безопасность. С одной стороны, искусственный интеллект может использоваться для обнаружения и предотвращения кибер угроз, но с другой стороны, он может быть использован в кибератаках. Основные аспекты этой проблемы включают:

- Защита данных: Использование искусственного интеллекта для анализа и обработки больших объемов данных требует усиленных мер по защите данных от утечек и несанкционированного доступа. Методы шифрования и аутентификации становятся более сложными и важными.
- Кибератаки с использованием искусственного интеллекта: Злоумышленники могут использовать искусственный интеллект для усовершенствования атак и обхода систем защиты. Разработка методов обнаружения и предотвращения кибератак с использованием искусственного интеллекта становится неотъемлемой частью обеспечения безопасности.

Результаты экспериментов. Для демонстрации эффективности использования в информационно-телекоммуникационных системах проведены эксперименты. Ниже представлены результаты:

Таблица 1.

Результаты экспериментов по оптимизации сетевого трафика

Метод	Уменьшение трафика (%)	Время выполнения (мс)
Без использования	-	500
ИИ-оптимизация	30	250
Традиционные методы	20	300

Формула 1. Функция потерь для оптимизации сетевого трафика

$$\mathrm{L} = \sum_{i=1}^{N} (w_i * x_i) + \lambda * R \tag{1}$$

Где L - функция потерь, w_i - вес параметра, x_i- значение параметра, N - количество параметров, λ - коэффициент регуляризации, R - регуляризационная функция.

Исследование алгоритмов машинного обучения в информационно-телекоммуникационных системах. В данном разделе мы проанализируем некоторые ключевые алгоритмы машинного обучения, которые активно используются в информационно-телекоммуникационных системах и являются важными компонентами развития информационно-телекоммуникационных систем:

1. Алгоритм решающего дерева. Решающие деревья - это мощный инструмент для классификации и регрессии в информационных системах. Формула 2 представляет структуру решающего дерева:

$$Y = \mathcal{F}(X_1, X_2, \ldots, X_n) \tag{2}$$

Где Y - целевая переменная, $X_1, X_2, \ldots, X_n$- признаки, используемые для прогнозирования. Решающие деревья могут быть эффективными при обработке больших объемов данных в информационно-коммуникационных системах.

2. Метод опорных векторов (SVM). Метод опорных векторов является мощным инструментом для задач классификации и регрессии. Он строит гиперплоскость,

максимально разделяющую классы данных. Формула 3 представляет модель SVM:

$$\mathrm{F(X)} = \sum_{i=1}^{N} \theta_i * K(X, X_i) + b \tag{3}$$

Где F(X) - функция решения, θ_i - коэффициенты, $K(X, X_i)$ - ядро, b - смещение. SVM может быть эффективным методом для классификации данных в информационно-телекоммуникационных системах.

3. Глубокое обучение (Deep Learning). Глубокое обучение представляет собой нейронные сети с большим числом слоев. Оно демонстрирует выдающиеся результаты в области обработки изображений, естественного языка и многих других задач. Формула 4 представляет модель глубокой нейронной сети:

$$Y = \mathrm{F}(W^T * X + b) \tag{4}$$

Где Y - выход, W^T - матрица весов, X - вход, b - вектор смещения. Глубокое обучение может быть применено для анализа данных и прогнозирования в информационно-телекоммуникационных системах.

Таблица 2

Сравнение производительности алгоритмов машинного обучения

Алгоритм	Точность (%)	Время обучения (сек)
Решающее дерево	90	10
SVM	95	20
Глубокое обучение	98	120

В таблице 2 представлено сравнение производительности алгоритмов машинного обучения в информационно-телекоммуникационных системах. Это демонстрирует, что разные алгоритмы имеют свои преимущества и недостатки, и выбор алгоритма должен зависеть от конкретной задачи и требований системы.

В заключении развитие технологий искусственного интеллекта предоставляет множество возможностей для улучшения информационно-телекоммуникационных систем. Однако, эти возможности сопровождаются новыми проблемами и вызовами, которые требуют внимания и решения.

Эффективное использование искусственного интеллекта требует баланса между высокой доступностью, надежностью, и безопасностью. Результаты экспериментов показывают, что применение искусственного интеллекта может значительно повысить эффективность информационно-телекоммуникационных систем.

В данной статье была рассмотрена актуальная тема развития машинного обучения в информационно-телекоммуникационных системах. Были рассмотрены ключевые алгоритмы машинного обучения и проведено сравнение их производительности. Машинное обучение имеет огромный потенциал для оптимизации информационно-телекоммуникационных системах и повышения их эффективности.

Литература

1. Смит, Дж. Р. Искусственный интеллект и информационная безопасность // Издательство, 2020.
2. Джонсон, А. Технологии искусственного интеллекта в телекоммуникациях // Журнал телекоммуникаций, 2020.
3. Браун, С. Интеграция искусственного интеллекта в сети связи // Конференция по телекоммуникациям, 2021.
4. Гарсия, М. Этические аспекты применения искусственного интеллекта // Журнал информационной безопасности, 2017.
5. Ли, Ч. Анализ рисков при использовании ИИ в кибератаках // Конференция по информационной безопасности, 2022.
6. Ким, Ю. Тенденции развития информационно-телекоммуникационных систем // Журнал телекоммуникаций, 2020.
7. Хан, З. Применение машинного обучения для обнаружения угроз в сетях связи // Конференция по информационной безопасности, 2002.
8. Роджерс, П. ИИ и телекоммуникации: современные вызовы и перспективы // Журнал телекоммуникаций, 2012.
9. Чжан, Г. Искусственный интеллект в системах управления телекоммуникациями // Конференция по информационной безопасности, 2020.
10. Миллер, Э. Обеспечение информационной безопасности в эпоху искусственного интеллекта // Книга, 2010

«ҚҰҚЫҚҚА ҚАЙШЫ ИНТЕРНЕТ-КОНТЕНТ» КӨПТІЛДІ ОНТОЛОГИЯ ҚОСЫМШАСЫН ҚҰРУ

[1]Мамырбаев О.Ж., [2]Хайрова Н.Ф., [3]Ыбытаева Г.С.
[1]Ақпараттық және есептеуіш технологиялар институты, Қазақстан, Алматы
[2]«Харьков политехникалық институты» ұлттық техникалық университеті, Украина, Харьков [3]Сәтбаев университеті, Қазақстан, Алматы

Аңдатпа. *Мақалада «Құқыққа қайшы интернет-контент» көптілді онтологиясын әзірлеу, оны визуализациялау және пайдалану шектеулері ұсынылған. Қылмыстық лексикамен жасалған көптілді терминологиялық тезаурус және қылмыстық маңызы бар ақпараттың құрылған мәтіндік корпустары кіретін құқыққа қайшы мазмұнның онтологиясын толтырудың негізгі көздері қарастырылады.*

1 Кіріспе

Қазіргі уақытта онтология әр түрлі пәндік салалардағы (ПС) білім қорын ұсыну мәселелерін шешу үшін жиі қолданылады. Онтология семантикалық желілермен қатар белгілі бір пәндік саладағы білімді көрсету үшін ыңғайлы абстракция болып табылады [1]. Бұл жұмыс онтологияны толықтыру міндетімен тығыз байланысты. Онтологияны толықтыру дегеніміз – әртүрлі дереккөздерді автоматты түрде талдау және мәліметтер қоры ПС онтологиясына негізделген ақпараттық жүйенің мазмұнын табылған ақпаратпен толтыру. Мұндай жүйелерде алынған ақпарат берілген онтологияның түсініктері мен қатынастарының даналары ретінде ұсынылады.

Онтологияны автоматты түрде құру әдістері белгіленген пәндік саланың арнайы лингвистикалық корпустарын қолдануға негізделген. Мұндай корпустарда арнайы семантикалық белгілеу болуы немесе оқиғалардың түрлерін сипаттайтын белгілермен семантикалық аннотациялануы керек, мысалы, әлеуметтік-саяси (SPE) және <Person>, <Organization>, <Location>, <Time>, <Vehicle>, <Weapon> және басқалары.

Мақала келесідей ұйымдастырылған. 2-бөлім әдебиеттерді шолуға арналған. 3-бөлімде «Құқыққа қайшы интернет-контент» онтологиясын құрудың негізгі көздері көрсетіледі, 4-бөлімде көптілді онтологиясын құру сипатталады және соңғы бөлімде зерттеу нәтижелері талқыланады.

2 Байланысты жұмыстар

Онтологияны әзірлеу мен қолдауға қойылатын талаптарды қанағаттандыруға бағытталған жобалар саны үнемі өсіп келеді.

MicroCosmos онтологиясы (кейінірек OntoSem деп аталды) – ең танымал онтологиялық ресурстардың бірі. Бұл онтология «онтологиялық семантика» деп аталатын тәсілдің бөлігі ретінде жасалған [2]. Онтология мәтінді автоматты түрде өңдеу қосымшаларында қолдануға және мәтіндік сөйлемдердің мазмұнын мағыналық, тілге тәуелсіз ұсынуға арналған. Кіріс мәтін үшін алдын-ала өңдеу, морфологиялық талдау,

синтаксистік талдау, семантикалық талдау жасалады, оның нәтижелері мәтіндік-семантикалық көрініс түрінде ұсынылады.

Авторлар онтологияның көлемін шектеуге көп күш жұмсағандықтан, Микрокосмос онтологиясының (OntoSem) мөлшері шамамен 6 мың ұғымды құрайды, олардың әрқайсысы орта есеппен 16 қасиетпен сипатталады. Жүйенің сөздік қоры бірнеше ондаған мың сөздер құрайды.

Қазіргі уақытта лексикалық семантиканы сипаттау саласындағы белгілі жобалардың бірі – белгілі лингвист Чарльз Филлмордың [3] басшылығымен фреймдік семантика тұжырымдамасы аясында құрылған FrameNet лингвистикалық ресурсы. Жобаның мақсаты – кадрлық семантикаға негізделген онлайн-лексикалық ресурс құру және оған мәтіндер корпусы түрінде қор беру. 2009 жылы ресурста 11000-нан астам лексикалық бірліктермен байланысқан 960 иерархиялық ұйымдастырылған фреймдер болды.

Осы контекстегі терминдер мен қатынастарды түсіндіру үшін қылмысқа байланысты онтологиялар да бар және олар кейінірек кейбір қолданыстағы жүйелердегі білімді көрсету үшін қолданылады. Мысал ретінде жағдайды бағалау және аналитикалық мультимодальды платформа (MOSAIC) жобасы [4]; LEAs өзара әрекеттесуінің еуропалық онтологиясын және көптілді қылмыс онтологиясын бір уақытта қолданатын CAPER [5]; және ePOOLICE жобасы [6], COPKIT жобасы [7], ASGARD жобасы [8], TENSOR жобасы [9].

Онтологияны құру және қолдау әр түрлі жобаларда танымал назарға айналды. Осындай жобалардың бірі – Интерполдың жіктеу схемаларына сәйкес есірткі қылмыстарымен күресуге арналған CAPER онтологиясын әзірлеу. Бұл көптілді онтология төрт негізгі ұғымнан тұрады – «Қылмыстар», «Әдістер», «Маңызды шарттар» және «Елдер» – оның құрылымында кластар түрінде ұсынылған. Онтологияның қазіргі нұсқасында 346 түйін бар және қазіргі уақытта итальян, испан, ағылшын және иврит тілдеріне аудармалар, синонимдер және жаргон терминологиясын жинау жұмыстары жүргізілуде.

Сонымен қатар, құқық қорғау органдарына қылмыстық әрекеттерді талдау мен алдын-алуда қолдау көрсету үшін бірнеше басқа жобалар құрылды. Мысалы, COPKIT жобасы тергеу мен стратегиялық талдауға көмектесу үшін деректерге негізделген полиция технологияларын әзірледі. Оның құралдар жинағы жедел және стратегиялық деңгейлерде ерте ескерту/алдын ала әрекет ету парадигмасын сақтай отырып, білімді қалыптастыруға және пайдалануға мүмкіндік береді.

ASGARD жобасы сонымен қатар криминалистикалық тергеу үшін киберқылмыс деректерін қоса алғанда, үлкен деректерді алуға, біріктіруге, бөлісуге және талдауға арналған класындағы ең жақсы құралдар жинағын әзірледі. Тағы бір жоба, TENSOR, терроризмге қатысты контентті біріктіру және интернеттегі қауіптерді анықтау үшін бірыңғай семантикалық инфрақұрылым жасады. Бұл құрылым онтологияны және құқық қорғау органдарына террористік әрекеттерді ерте анықтау, жалдау үшін жоспарлау және алдын алу функцияларын ұсынатын семантикалық негіздеудің бейімделу механизмін қамтиды. Дегенмен, бұл құралдардың әдетте екі негізгі кемшілігі бар. Біріншіден, олар көптілді емес және қолданыстағы аннотацияланған қылмыстық есептер негізінен ағылшын тілінде қол жетімді. Екіншіден, кейбіреулерінде визуализация мүмкіндіктері шектеулі, атап айтқанда аталған субъектілер арасындағы танылған байланыстардың графикалық көрінісі.

Талдау көрсеткендей, қазіргі уақытта қылмыстық тақырыпқа байланысты орыс және қазақ тілдерінің жалпыға қолжетімді онтологиялары жоқ.

Пәндік саланың көптілді онтологияларының толық болмауы онтологияны жасаушылардың қазіргі жұмыс бағытын көрсетеді. Сондай-ақ, мұндай лексикалық ресурстардың да, оларды әзірлеу модельдерінің де ашықтығы онтологияның сапасын жақсартуға, сондай-ақ тиісті салалардағы мәтінді автоматты түрде өңдеу мәселелерін шешуге айтарлықтай үлес қосуға мүмкіндік беретінін атап өткен жөн.

3 «Құқыққа қайшы интернет-контент» онтологиясын құрудың негізгі көздері

«Құқыққа қайшы интернет-контент» онтологиясын қалыптастырудың негізгі көздері қылмыстық жаңалықтарды қамтитын мәтіндерден тұратын әзірленген параллель қазақ-орыс корпусы болып табылады [10] және XML форматында әзірленген қылмыстық лексиканың көптілді терминологиялық тезаурусы [11].

Тезаурустың негізгі лексикасы ағылшын, украин, қазақ және орыс тілдеріндегі қылмыстық тақырыптағы мәтіндерден қолмен алынды. Жеті негізгі тақырыптық категория бөлінді: «*Movement*» (Жол қозғалысы), «*Traffic Accident*» (Жол-көлік оқиғасы), «*Injure*» (Зиян келтіру), «*Offense*» (Құқық бұзушылық), «*Arrest*» (Қамауға алу), *«Trial»* (Соттың істі қарауы) және *«Police Department»* (Полиция қызметі), бұл тезаурусты тар тақырыптық етуге мүмкіндік берді. Категориялардың бұл таңдауы тезаурусты толтыру үшін қолданылатын ақпараттық ресурстардың тақырыбына байланысты. Зерттелген қылмыстық жаңалықтар үш қылмыстық бағытқа қатысты: «*Police*» (Полиция), «*Transfer*» (Қозғалыс), «*Crime*» (Қылмыс) және олардың жоғарыда аталған ішкі түрлері.


```
<vocabulary>
    <nouns>
        <term id="1">
            <lemma lang="ru">стрельба</lemma>
            <domain>OFFENSE</domain>
            <synset lang="ru">обстрел, выстрел</synset>
            <definition lang="ru">учебные занятия по ведению огня из различных видов оружия;  ведение огня,
применение огнестрельного оружия для выполнения поставленной задачи (Пулевая с., стендовая с., с. из
пистолета, с. из лука)</definition>
            <example lang="ru">Два человека получили ранения при стрельбе в Таразе</example>
            <hypernims lang="ru">['приведение в действие', 'движение']</hypernims>
            <hyponims lang="ru">['контрвыстрел', 'разряд', 'отстрел', 'выстрел', 'выстрел из пистолета',
'выстрел в голову', 'выстрел из снаряда', 'перестрелка']</hyponims>
            <lemma lang="en">shooting</lemma>
            <synset lang="en">firing, fire, gunfire</synset>
            <definition lang="en">the act of firing a projectile</definition>
            <example lang="en">his shooting was slow but accurate</example>
            <hypernims lang="en">['actuation', 'propulsion']</hypernims>
            <hyponims lang="en">['countershot', 'discharge', 'firing', 'firing off', 'gunfire', 'gunshot',
'headshot', 'potshot', 'shellfire', 'shoot']</hyponims>
            <lemma lang="ka">атыс</lemma>
            <synset lang="ka">ату, оқ жаудыру, атылыс</synset>
            <definition lang="ka">оқ атылғанда шығатын дыбыс, тарсыл; көздесу, нысанаға алып, оқ тигізу
</definition>
            <example lang="ka">Алматының Ақбұлақ мөлтекауданында атыс болып, бес адам қаза тапты</example>
            <hypernims lang="ka">['іске қосу','қозғаушы күш']</hypernims>
            <hyponims lang="ka">['қарсы атыс', 'ату', 'басына ату', 'снарядтан ату', 'атыспақ', 'атысу',
'атысып қалу']</hyponims>
            <lemma lang="ua">стрілянина</lemma>
            <synset lang="ua">стрільба, пальба</synset>
        </term>
```


1-сурет Қылмыстық лексиканың көптілді тезаурусының үзіндісі

1-суретте қазіргі уақытта 600-ге жуық негізгі сөздерді (330 зат есім, 107 сын есім және 170-ке жуық етістік) және 2500-ден астам негізгі сөз синонимдерін қамтитын тезаурустың үзіндісі көрсетілген. Болашақта тезаурусты биграммаларды қолдану арқылы кеңейту жоспарлануда [12].

Әлбетте, тар мамандандырылған онтологияның автоматты және автоматтандырылған генерациясы пәнге бағытталған мәтіндік корпустарға негізделуі керек. Бұл зерттеуде құрылған көптілді онтологияға арналған мәтіндерден нақты лексикалық ресурстарды алу үшін қылмыстық тақырыптарға арналған екі корпус қолданылды.

Бірінші көптілді корпусқа орыс, украин және ағылшын тілдерінің мәтіндері кіреді. Оны толтыру үшін ақпарат 2018 жылдың маусымынан 2020 жылдың қазанына дейін Python BeautifulSoup кітапханасының талдаушысын пайдалана отырып, жаңалықтар интернет-сайттарынан алынды. Корпустағы әрбір мәтін қылмыстық тақырыппен байланысты және Ukr_texts, Eng_tests, Ru_tests үш каталогының біріне орналастырылған.

Украин тіліндегі мәтіндер автоматты түрде «Украинская правда» ресми сайтынан, сондай-ақ «Главком» сайтынан жүктелді. Украинаның ішкі корпусында 3147 мәтін бар.

Орыс тіліндегі мәтіндер «Редпост» жаңалықтар сайтынан, Харьковтың қоғамдық-саяси аймақтық басылымынан, атап айтқанда «Преступность и происшествия» бөлімінен алынды. Корпустың бұл бөлігінде 5506 мәтін бар. Ағылшын тіліндегі мәтіндер Корпус Кристи, Техас штатындағы «Caller Times» газетінің «Қылмыс» бөлімінен алынған. Қазіргі уақытта бұл ішкі корпуста 300 мәтін бар.

Онтологияны қалыптастыру үшін база ретінде пайдаланылатын екінші көптілді корпус – төрт жылдан астам уақыт бойы дамып келе жатқан параллель қазақ-орыс корпусы [10]. Осыған байланысты мәтіндердің сапалы параллельді көптілді корпустарын құру қазіргі лингвистиканың ең өзекті және прогрессивті бағыттарының бірі екенін атап өткен жөн.

Осы параллель корпусты кеңейту және толықтыру 2021 жылдың наурызы мен қарашасы аралығында Қазақстанның ақпараттық интернет-кеңістігінің zakon.kz, caravan.kz, lenta.kz, nur.kz төрт жаңалықтар сайттарының парсер көмегімен жүзеге асырылды. Бұл сайттарда қылмыстық ақпаратқа, соның ішінде тонау, кісі өлтіру, жол-көлік оқиғалары және т.б. сияқты қылмыстарға қатысты көптеген жаңалықтар мақалалары бар. Қазіргі уақытта параллель қазақ-орыс корпусының көлемі орыс тіліндегі 3000 мәтінді және қазақ тіліндегі 3000 мәтінді, оның ішінде мағынасы бойынша тураланған қазақ-орыс сөйлемдерін қамтитын 2000 мәтінді құрайды.

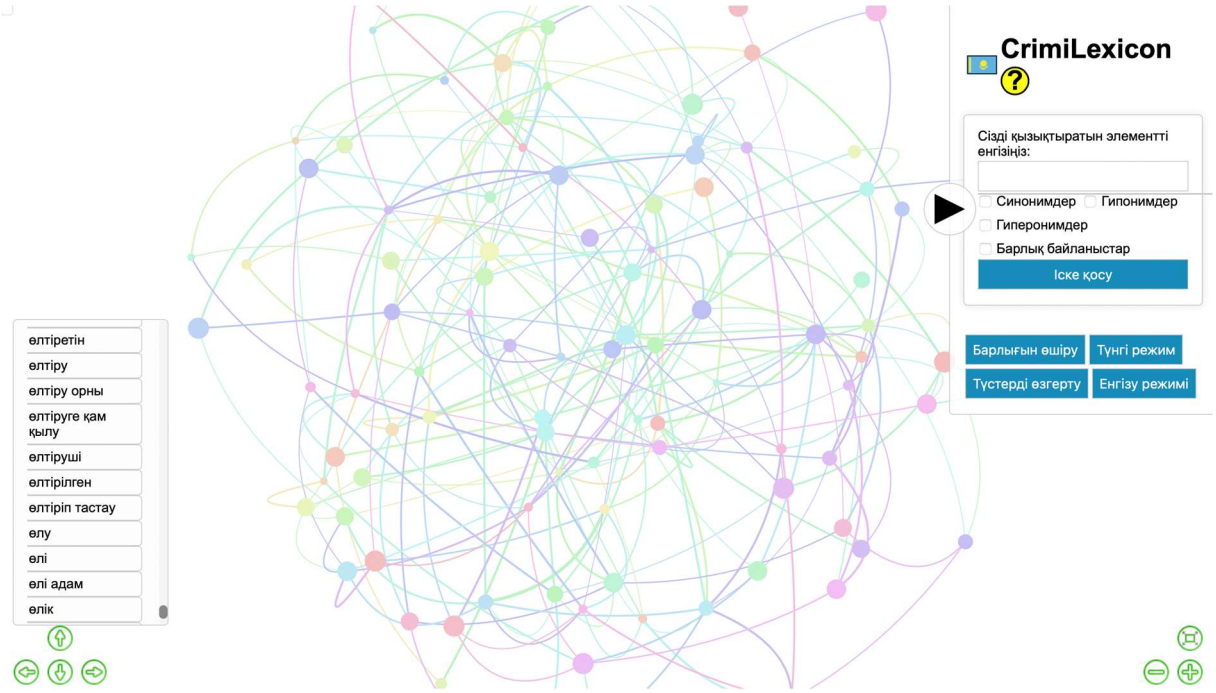
4 «Құқыққа қайшы интернет-контент» көптілді онтологиясын құру

Құқыққа қайшы интернет-контентін іздеу мен талдаудың интеграцияланған технологиясында қылмыстық боялған оқиғаның триггерін анықтау және түрді анықтау бұл триггер және/немесе оқиға көп тілді «Құқыққа қайшы интернет-контент» онтологиясына негізделген [11]. Онтология құқыққа қайшы әрекеттер мен қылмысқа қатысты терминдердің синонимдік сөздігіне негізделген.

Онтологиялық визуализация динамикалық және интерактивті веб-қосымша ретінде жасалған және 2-суретте көрсетілген.

Сөздік файлынан алынған онтологияның қазіргі нұсқасы 12 885 нысанды қамтиды, олардың әрқайсысында бірегей 10 таңбалы идентификатор бар [13]. Онтология төрт негізгі класты қамтиды: *Category*, *Domain*, *Term* және *Word*. *Category* класы терминнің сөз табын анықтауға арналған, сәйкес мағыналары бар – зат есім, етістік, сын есім. *Domain* класы құқыққа қайшы әрекеттер лексикасының жіктелуінің семантикалық кластарының түрі мен ішкі түрін анықтайды. *Term* класы онтология анықтаған ұғымдарды қалыптастыру үшін қолданылады. Бұл класс қазақ, орыс, ағылшын және украин тілдеріндегі тұжырымдаманың әртүрлі лексикалық сөйлемдерін қамтиды. Мысалы, мұндай n80 тұжырымдамасы орыс тіліндегі *«нож»* сөзін, ағылшын тіліндегі

«knife» сөзін және қазақ тіліндегі *«пышақ»* сөзін білдіреді. Word класы онтологиядағы жеке сөздерді білдіреді және тілді анықтайтын екі әріптен тұратын атрибуттары, анықтамасы (*definition*), мысалы (*example*) бар.



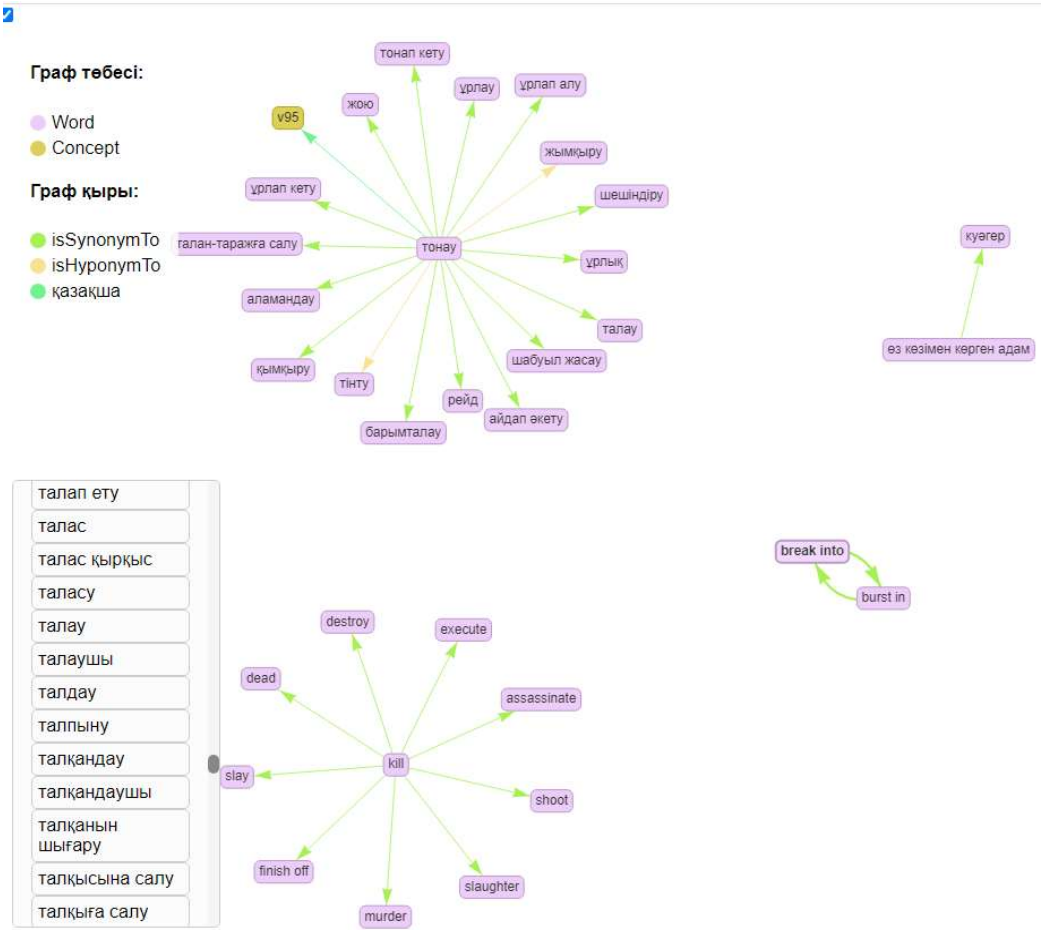
2-сурет «Құқыққа қайшы интернет-контент» көптілді онтологияны визуализациялау

3-суретте онтология субъектілер және олардың байланыстары туралы құрылымдық ақпаратты графикалық түрде кодтайтын білім графигі (KG) түрінде берілген.

3-сурет Көптілді онтологияның білім графигі түріндегі визуализациясы

Пайдаланушы интерфейсін жақсарту және пайдаланушының онтологиямен өзара әрекеттесуін жеңілдету үшін қосымша Query of Crime (QC) әзірленген сұратым тілімен толықтырылған, бұл пайдаланушыға күрделі сұратымдарды оңай және дәл тұжырымдауға мүмкіндік береді. QC синтаксисі синонимдерді (*syn*()), гипонимдерді

(*hypo*()), гиперонимдерді (*hyper*()), терминдерді (*term*()) және домендерді (*domain*()) іздеу құрылымдарын қамтиды. *All()* – жақшадағы сөз/ұғым/терминмен барлық байланыстарды визуализациялайды. {Cypher-ге сұратым} – егер QC мүмкіндіктері қажеттіліктерді қанағаттандырмаса және күрделі сұранысты орындау қажет болса, онда сұранысты бұйра жақшаға (n (бірінші нүкте), r (қатынас) және m (екінші нүкте)) сұрау операндтары ретінде енгізу керек. Барлық қатынастарды визуализациялауға, сондай-ақ мәліметтер қорының бірнеше қосымша элементтерін көрсетуге мүмкіндік беретін QC тіліндегі сұратым мысалы 4-суретте көрсетілген.



4-сурет «Құқыққа қайшы интернет-контент» онтологиясына күрделі сұратымды іске асыру

Көп тілді онтологияның қолданылу саласы – бұл құқық қорғау органдарының интеграцияланған ақпараттық-криминалистикалық жүйелері, сондай-ақ интернет-мазмұнның ашық бөлігіндегі құқыққа қайшы ақпаратты ақпараттық іздеу және талдау үшін қолданылатын басқа мемлекеттік органдардың интеграцияланған ақпараттық-аналитикалық жүйелері, атап айтқанда әртүрлі блогтар, әлеуметтік желілер, көп тілді веб-БАҚ-тың онлайн-топтық талқылау қауымдастықтары.

«Құқыққа қайшы интернет-контент» көптілді онтологиясы http://multilingual.ontology.iict.kz:38000/ мекен-жайы бойынша еркін және жалпыға қол жетімді. Онтологияның құрылымы оны мәтіндік құжаттардың мағынасын модельдеу кезінде компьютерлік лингвистика мен жасанды интеллект тәсілдерін дамытуға, сондай-ақ

мәтіндерді семантикалық талдау әдістерін жасауға, құрылымдалмаған мәтіндерге негізделген онтологияларды автоматты түрде құруға және ақпарат іздеуге пайдалы құрал етеді.

5 Қорытынды

Пәндік салалардың қол жетімді онтологиясының толық болмауы онтологияны жасаушылардың қазіргі жұмыс бағытын көрсетеді. Сондай-ақ, мұндай лексикалық ресурстардың да, оларды әзірлеу модельдерінің де ашықтығы онтологияның сапасын жақсартуға, сондай-ақ тиісті салалардағы мәтінді автоматты өңдеу мәселелерін шешуге айтарлықтай үлес қосуға мүмкіндік беретінін атап өткен жөн.

Әзірленген онтология құқыққа қайшы әрекеттер туралы білімді мәтіндік мазмұнды талдау және берілген ақпараттық домен мәтінінің әртүрлі элементтері арасындағы байланыстарды анықтау үшін тиімді пайдалануға болатын білімнің семантикалық графиктері түрінде ұсынуға мүмкіндік береді. «Құқыққа қайшы интернет-контент» онтологиясын құру Интернет ортасында таратылатын көп тілді құқыққа қайшы мәтіндік ақпаратты автоматты түрде іздеу мен талдаудың дәлдігін арттыруға мүмкіндік береді.

6 Алғыс

Жұмыс Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің қаржылық қолдауымен орындалды (№AP09259309).

Әдебиет

1. Bessmertny I. Knowledge visualization based on semantic networks / Program Comp. Soft, 2010. – Vol. 36, no. 4. – P. 197-205.
2. Nirenburg S., Raskin V. Ontological Semantics / MIT Press, 2004.
3. Fillmore C.J., Miriam R.L, Petruck J.R., Abby W. Framenet in Action: The Case of Attaching / International Journal of Lexicography, 2003. – 16.3. – P. 297-332.
4. Adderley R., Seidler P., Badii A., Tiemann M., Neri F., Raffaelli M. Semantic Mining and Analysis of Heterogeneous Data for Novel Intelligence Insights / Fourth Int. Conf. Adv. Inf. Min. Manag, 2014. – Vol. 1. – P. 36-40.
5. Casanovas P., Arraiza J., Melero F., González-Conejero J., Molcho G., Cuadros M. Fighting Organized Crime Through OpenSource Intelligence: Regulatory Strategies of the CAPER Project / Front. Artif. Intell. Appl., 2014. – Vol. 271. – P. 189-198.
6. Brewster B., Andrews S., Polovina S., Hirsch L., Akhgar B. Environmental scanning and knowledge representation for the detection of organised crime threats / Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 8577 LNAI, 2014. – P. 275-280.
7. COPKIT project. – https://copkit.eu/ (01.02.2022)
8. ASGARD project. – https://www.asgard-project.eu (01.02.2022)
9. TENSOR project. – https://tensor-project.eu (10.02.2022)
10. Khairova N., Kolesnyk A., Mamyrbayev O., Mukhsina K. The aligned Kazakh-Russian parallel corpus focused on the criminal theme / CEUR Workshop Proceedings, 2019. – P. 116-125.
11. Khairova N., Kolesnyk A., Mamyrbayev O., Ybytayeva G., Lytvynenko Yu. Automatic Multilingual Ontology Generation Based on Texts Focused on Criminal Topic / Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems, 2021. – Vol.1. – P. 108-117.
12. Ыбытаева Г.С., Мамырбаев О.Ж., Хайрова Н.Ф., Жумажанов Б.Ж.. Қазақ тіліндегі мәтіндерде коллокацияларды анықтаудың статистикалық әдістерін талдау / Мат. VI

Межд. науч. конф. «Информатика и прикладная математика». – Алматы, Казахстан, 2021. – С. 256-262.

13. Ybytayeva G., Mamyrbayev O., Khairova N., Rizun N., Berdali S., Mukhsina K. Creating a Thesaurus "Crime-Related Web Content" Based on a Multilingual Corpus / 7th International Conference on Computational Linguistics and Intelligent Systems, Kharkiv, Ukraine, April 20-21, 2023. CEUR-WS. – P. 77-87.

A NEW AUTOMATICALLY CONTROLLED WIND POWER STATION OF THE PARALLEL MANIPULATOR

Kuanysh Abzhaparov, Zhiyengali Adilkhan, Korganbayeva Zhuldyz
e-mail: Zhiyengali.a@gmail.com
Automation and Control Department, Satbayev University, 22a Satbayev Str., Almaty 050013 Republic of Kazakhstan

Abstract. *The article considers a wind power station, which significantly differs from the other existing wind power station in that it has rocking an automatically controlled sail of a toroidal shape and a manipulator converter of the wind kinetic energy to the mechanical energy of forward motion. Having analyzed aerodynamic variables, timing chart, and having built a simplified dynamic model, the authors determined kinematic properties, which are necessary for the future design engineering, automatic control, and research of the wind power station. The study gives a description of the acting laboratory-scale model of the sail wind power station and performance test results.*

1 Introduction

A review of existing technology [1-3] for converting wind energy into electrical energy shows that the main problem is unpredictability of wind, its speed, and gusts, often changing over the short periods of time.

Currently, the most common devices are wind turbines with the turbine rotor blades moving under the action of wind lifting forces as a working body. It is known that wind turbines have a small range of wind speeds for efficient operation, an adverse effect of the dynamics of moving bodies located at an altitude of over10 m on a cumbersome tower, and other known shortcomings.

Saphon Energy, a Tunisian startup, has invented a sailing bladeless wind power installation ‘Saphonian’ [4], in which under the action of wind pressure the sail oscillates and drives the pistons. However, in this sailing wind power installation the voluminous structure, including the mobile components, is located on a tall mast, which causes devastating fluctuations of the system.

2 Design and Operation Mode of the Wind power installation

In the new WPI sail, toroidal shape has an aerodynamic profile in the cross sections and can be executed in the form of a hollow body of dense elastic material (parachute fabric) filled with gas or can be made of lightweight synthetic or composite material. For example, in Fig.1 a, the sail made of foam is presented. Depending on the wind speed, the volume of the hollow sail changes by injection or bleed gas or in other ways to change the windage. The presence of such a sail allows the WPI to start functioning at a wind speed of 1.5-2 m/s. Under the action of the wind, the sail makes spatial periodic movements depending on the direction and speed of the wind.

In the WPI, the sail is located at an altitude of more than 10 m and is rigidly connected to the upper platform of the parallel six-movable manipulator (Fig.1,b) created earlier platform robot SHOLKOR [5]. This manipulator allows you to transform the spatial movements of the sail into a series of translational movements. Figure 1, b shows the physical model of the sholkor robot parallel manipulator. Here the lower platform 1 is connected to the upper platform 2 by means of six actuators 3. The feature of the model is that for the connection of actuators with platforms applied a flexible connections protected by a patent [6], allowing to exploit the complex connections without grease in the open air and the aquatic environment. Figure 2 shows the scheme of wind power installation, used in the future as a calculation. Here, sail 13 is connected by three spacer rods 12 to the upper platform of the manipulator. This design forms the working body of the WPI. The working body together with the manipulator is further called the manipulator converter (MC). The movable platform 2 of the manipulator is connected to the lower fixed platform 1 by means of six actuators (rod-cylinder) 3-8. Each actuator represents the connection of piston 9 and cylinder 10. A flexible connection 11 connects the actuators with the manipulator platforms and among themselves.

a) *b)*

Fig 1 Device model of WPI

Here, in the WPI the sail captures the kinetic energy of the air mass, and the manipulator converts this energy into mechanical energy of the forward motion of six rods relative to the cylinders. The novelty of the present sail WPI is the presence of an airfoil section sail with an automatically controlled aerodynamic surface area. This assumes automatic control of the WPI operation as a whole. In this regard, the aim is to justify the functionality of sailing wind power installation. To do this, the justification for choosing the shape of the dimensions and the sail is performed sequentially. Then we consider a simplified dynamic model of the working body (hereinafter WB), which allows to correctly solve the problem of dynamic analysis of the working body and MC WPI for the subsequent solution of problems of calculation of the design of the WPI as a whole and the design of the automatic control system.

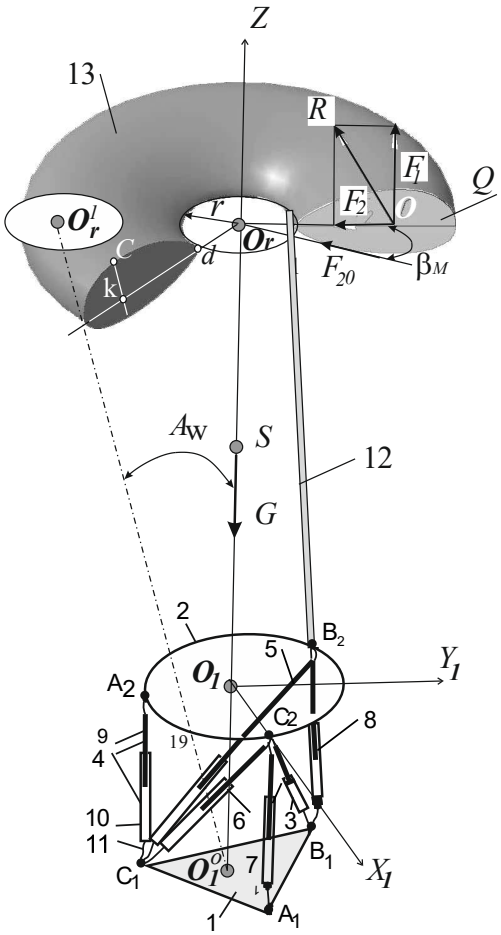


Fig 2 Diagram of a sail WPI

3 The research of the WB's Kinematics

The following objective is to establish the position of the A_2,B_2,C_2 nodes of the upper platform, as well as the speed of actuator rods' movement, if the position of the WB is set by the coordinates ($y_S z_S$) of the centre S and by the rotation (φ_2) relative to the axis passing through S. To solve this problem, applied in robotics, homogeneous transformation matrices are used. As noted earlier the reference system $C_1X_0Y_0Z_0$ is fixed; the mobile system $O_1X_1Y_1Z_1$ is connected with the WB represented in the form of a trihedral prism SA_2, B_2, C_2.

We establish the T matrix of the composition of transformations, which is obtained as a result of multiplying the transformation matrices during elementary rotations and shifts. Below is a sequence of simple movements. In brackets are matrices' designations of elementary transformations:

1. Shift*(A_1)*along the axes of the fixed system until the coincidence C_1andO_1.
2. Rotation*(A_2)*around the axis$O_1Z_1$to an angle until the axisO_1Y_1 does not for m angle α with the axisC_1Y_0.
3. Shift*(A_3)*to the value y_S,
4. Shift *(A_4)* to the value z_s
5. Shift*(A_5)* to the value L_S along the axisO_1Z_1.
6. Rotation*(A_6)*around the axis$O_1X_1$to the angleφ_1.
7. Shift*(A_7)*to the value -L_S along theaxisO_1Z_1.
8. Rotation*(A_8)*around axisO_1Z_1 to an angle –α.

The T matrix of composition of transformations is established by the equality

$$T=(A_1)\ (A_2)\ (A_3)\ (A_4)\ (A_5)\ (A_6)\ (A_7)\ (A_8), \tag{1}$$

$$T=\begin{pmatrix}1&0&0&x_{O1}\\0&1&0&y_{01}\\0&0&1&z_{O1}\\0&0&0&1\end{pmatrix}\cdot\begin{pmatrix}C\alpha&-S\alpha&0&0\\S\alpha&C\alpha&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\cdot\begin{pmatrix}1&0&0&0\\0&1&0&y_S\\0&0&1&0\\0&0&0&1\end{pmatrix}\cdot\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&z_S\\0&0&0&1\end{pmatrix}\cdot$$
$$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&L_S\\0&0&0&1\end{pmatrix}\cdot\begin{pmatrix}1&0&0&0\\0&C\varphi_1&-S\varphi_1&0\\0&S\varphi_1&C\varphi_1&0\\0&0&0&1\end{pmatrix}\cdot\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&-L_S\\0&0&0&1\end{pmatrix}\cdot\begin{pmatrix}-C\alpha&S\alpha&0&0\\-S\alpha&-C\alpha&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix} \tag{2}$$

As a result of matrix multiplication (1), the T transformation matrix is obtained with elements in the form.

$$\begin{aligned}&t_{11}=S^2(\alpha)\cdot C(\varphi_1)-C^2(\alpha);\ t_{12}=C(\alpha)\cdot S(\alpha)+C(\alpha)\cdot S(\alpha)\cdot C(\varphi_1);\\&t_{13}=S(\alpha)\cdot S(\varphi_1);\ t_{14}=x_{01}-y_S S(\alpha)-L_S S(\alpha)\cdot S(\varphi_1);\ t_{21}=-C(\alpha)\cdot S(\alpha)-\\&-C(\alpha)\cdot S(\alpha)\cdot C\ (\varphi_1);\ t_{22}=S^2(\alpha)-C^2(\alpha)\cdot C(\varphi_1);\ t_{23}=-C(\alpha)\cdot S(\varphi_1);\\&t_{24}=y_{01}+y_S C(\alpha)+L_S C(\alpha)\cdot S(\varphi_1);\ t_{31}=-S(\alpha)\cdot S(\varphi_1);\\&t_{32}=-C(\alpha)\cdot S(\varphi_1);\ t_{33}=C(\varphi_1);\ t_{34}=L_S+z_{01}+z_s-L_S C(\varphi_1);\\&t_{41}=t_{42}=t_{43}=0;\ t_{44}=1.\end{aligned} \tag{3}$$

Coordinates of any point of the WB at plane-parallel motion can be established in the fixed reference system.

$$\boldsymbol{\rho}^{\mathbf{0}}=T\boldsymbol{\rho}^{1} \tag{4}$$

where $\boldsymbol{\rho}$ is homogeneous column vector made of coordinates in the system with the index indicated at the top.

For example coordinates of the nodal point C_2 the upper platform of the manipulator in the fixed system $C_1X_0Y_0Z_0$ corresponding to the movement of a platform when changing coordinates of the WB $[y_S z_S \varphi_2]^T$ and an angle α is established by matrix equality

$$\begin{bmatrix}x_{C2}^0 & y_{C2}^0 & z_{C2}^0 & 1\end{bmatrix}^T=T\begin{bmatrix}x_{C2}^1 & y_{C2}^1 & z_{C2}^1 & 1\end{bmatrix}^T\cdot \tag{5}$$

The speed of node point A_2, B_2, C_2 is established by time differentiation of an equality (4)

$$\dot{\boldsymbol{\rho}}^{\mathbf{0}}=\dot{T}\boldsymbol{\rho}^{1}$$

At the same time

$$\dot{T}=\sum_{i=1}^{3}U_i\dot{q}_i=U_1\dot{y}_S+U_2\dot{z}_S+U_3\omega_2^{\cdot} \tag{7}$$

where

$$U_1 = A_1 A_2 \frac{dA_3}{dy_S} A_4 A_5 A_6 A_7 A_8,$$
$$U_2 = A_1 A_2 A_3 \frac{dA_4}{dz_S} A_5 A_6 A_7 A_8, \quad (8)$$
$$U_3 = A_1 A_2 A_3 A_4 A_5 \frac{dA_6}{d\omega_2} A_7 A_8.$$

For example, the speed of the nodal point C_2 is established from (6) using equalities (7), (8).

$$\begin{bmatrix} \dot{x}_{C2}^0 & \dot{y}_{C2}^0 & \dot{z}_{C2}^0 & 0 \end{bmatrix}^T = \dot{T} \begin{bmatrix} x_{C2}^1 & y_{C2}^1 & z_{C2}^1 & 1 \end{bmatrix}^T \quad (9)$$

Similarly, speeds of any points of the WB can be established in the fixed coordinate system, including nodal points B_2, A_2.

A change in the position of the upper platform of the manipulator causes a change in the position, lengths of actuators, the direction of reaction forces vectors, impacting the upper platform from the actuators. To establish these values, the inverse kinematics problem is solved for the manipulator. The main point of this task is that at a known position of the moving upper platform on coordinates of nodal points (A_2, B_2, C_2) the orientation and lengths of actuators are established, hence the direction of actuators reaction forces acting along the axis of actuators. As an example, we will solve the inverse problem for the actuator. The coordinates of the node C_2 ($x_{C2}^0 \quad y_{C2}^0 \quad z_{C2}^0$) are known from equality (5), the coordinates of the node C1 (0 0 0) are equal to zero. The current length of the actuator 6 is equal to

$$l_{6t} = \sqrt{(x_{C2}^0)^2 + (y_{C2}^0)^2 + (z_{C2}^0)^2}\,. \quad (10)$$

Anglesα_6,β_6, $\gamma_6$between theaxisoftheactuator 6 and axes C_1X_0, C_1Y_0, C_1Z_0 fixed coordinate systems are established as

$$\alpha_6 = arc\cos(x_{C2}^0 / l_{6t}), \quad \beta_6 = arc\cos(y_{C2}^0 / l_{6t}) \quad \gamma_6 = arc\cos(z_{C2}^0 / l_{6t})\,. \quad (11)$$

Changing the length of the actuator 6

$$\Delta l_6 = l_{6t} - l_6\,. \quad (12)$$

4 Dynamic Analysis of the Manipulator Converter

If we take into account that the sail is influenced by a resultant of aerodynamic forces of wind and gravitational force of the working body, then during cyclic variation of the wind speed and the change of the wind direction, the working body's movement will be spatially closed. With this movement of the upper platform, the lengths of all six telescopic joints will vary. Shown in Figure 4, results of solving a kinematic task for a parallel manipulator in MatLab by

the algorithm proposed in [7] demonstrate that during the reverse motion of the movable platform all six rods receive the displacement increment.

To estimate the power of the WPI during the engineering or in operation, data are required on the forces in actuators and the rods motion speed. Force of the cylinder rod along with the motion allow to choose the design of pneumatic and hydraulic cylinders. To determine the forces on the rods, a force analysis of the MC is performed using the kinetostatics method.

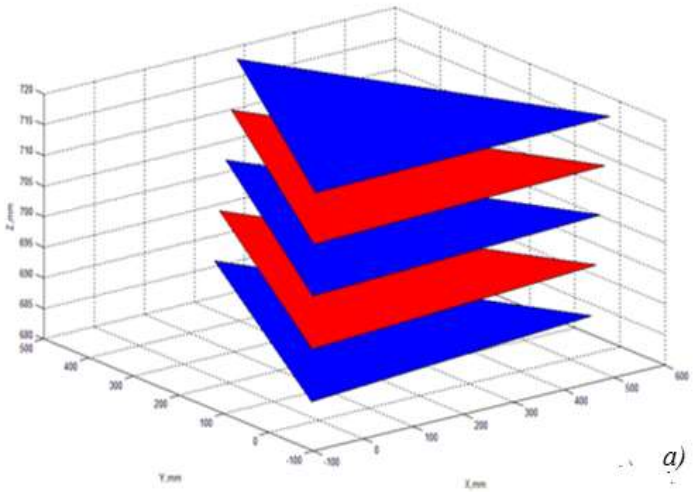


Fig. 4 The diagram of location of the working body's upper platform

The initial data is the main vector of the reaction force $\vec{R} = \vec{R}_Z + \vec{R}_Y$ applied at the point O_1 and a vector of a couple of forces $\vec{M}_R$ resulting from the dynamic analysis of the working body under the action of maximum wind speed of 15-20 m/s and with the inertia forces taken into account. It should be noted that the force $\overline{R}$ is located (Figure 5) in the plane Q, and the moment of couple $\vec{M}_R$ is directed perpendicularly to the plane Q which coincides with the moving plane $O_1Y_1Z_1$. Thus, with a given wind direction, the values $\overline{R}$ and $\vec{M}_R$ are known. In addition to these forces, the unknown forces N_3, N_4, N_5, N_6, N_7, N_8 directed along the rods' axes influence platform 2. The directions of the action lines for each position are determined by solving an inverse problem of kinematics.

We chose a fixed coordinate system $O_0 X_0 Y_0 Z_0$ connected with the lower platform of the manipulator converter with the beginning in the node C_1. Since $\overline{R}$ and $\vec{M}_R$ comprise, the inertia forces the kinetostatics equation for the platform has the following form.

$$\sum_{i=3}^{i=8} N_{iz} + R_z = 0; \sum_{i=3}^{i=8} N_{iy} + R_y = 0;$$
$$\sum_{i=3}^{i=8} N_{ix} = 0; \sum_{i=3}^{i=8} M_x(\mathbf{N}_i) + M_{Rx} = 0; \qquad (14)$$
$$\sum_{i=3}^{i=8} M_y(\mathbf{N}_i) + M_{Ry} = 0; \sum_{i=3}^{i=8} M_z(\mathbf{N}_i) = 0.$$

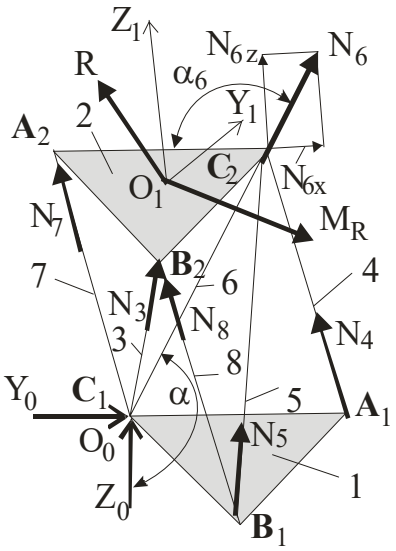


Fig. 5 Scheme for the MC force analysis

The obtained values of forces on the MC rod are used later in choosing pneumatic-hydraulic cylinders for the WPI design engineering.

5 SCADA system for supervision and data acquisition of adaptive control system.

To monitor the complex of wind farms, it is proposed to develop a SCADA system to display the operation of the complex. Figure 6 shows the state of one wind power station in the normal state of operation.

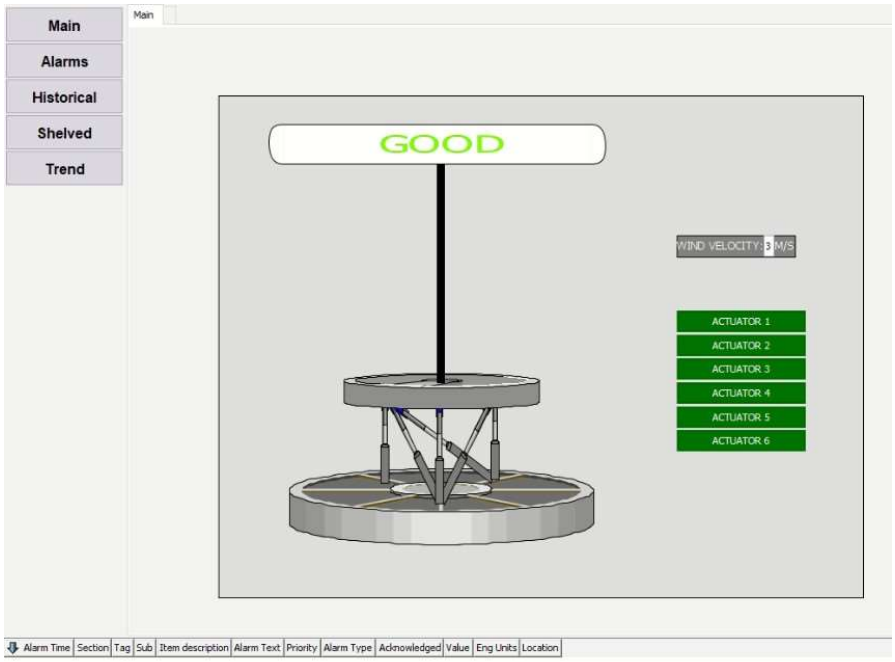


Fig. 6 The normal state of wind power station

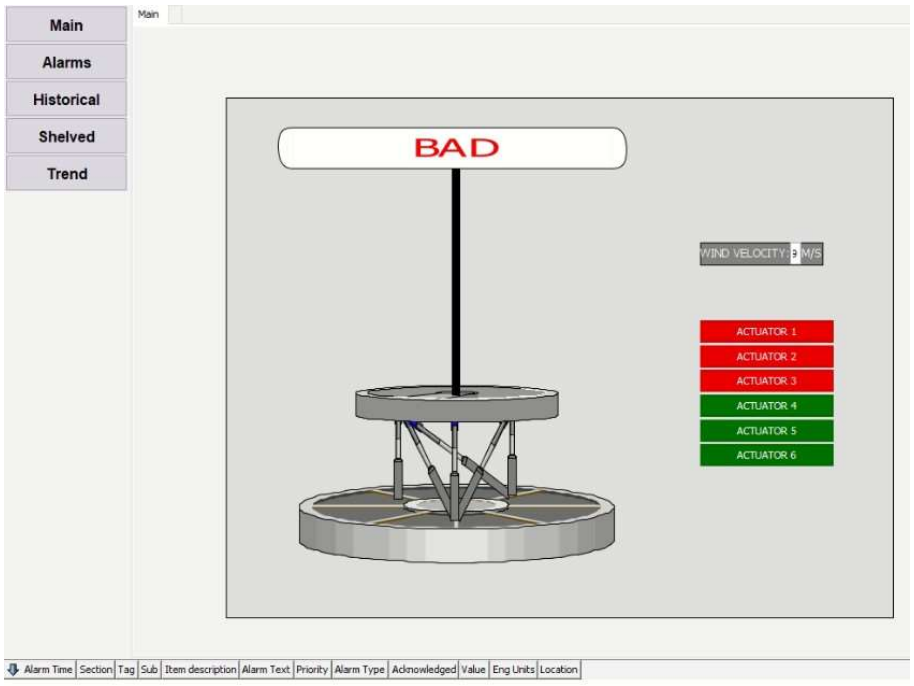


Fig. 7 The emergency state of the wind power station

Figure 7 shows the response to an increase in wind speed, the emergency state of the wind power station, and the system's response to load damping.

This system helps to track the state of the power station on historical trends, which will facilitate the training of the adaptive system to respond to changes in external influences. It is assumed that 5 emergency events will be enough to fully train the system to respond to external disturbances.

Conclusion

In conclusion, we would like to note that this article contains the date required for creating and implementing the new automatically controlled sail wind power station having the opportunity to generate electric power with low wind speed. The article justifies the sail parameters for the new WPS by the method of computer modeling. To design and research the sail WPS, the mathematical apparatus based on state equations. The paper describes the working demonstration sample of the new sail WPS with working body performing cyclic movements. The offered sail wind power station due to its capabilities in efficient energy conversion can significantly expand the application field of the wind sources of renewable energy.

References

1. Quaschning V. Regenerative Energiesysteme / V. Quaschning // - Hanser Munich: Berlin., 2022. – Vol. 11, N. 9. - P. 444.
2. Al-Majed S. I., Fujigaki T. Advances in Wind Power. - http://dx.doi.org/10.5772/3376
3. Al-Majed S.I. and Fujigaki T. *Wind power generation*: An *overview.* - Wind power generation: An overview | IEEE Conference Publication | IEEE Xplore

4. Labaied H., Aouini A. Saphon Energy is a cleantech start-up specialized in R&D in Wind Energy. - http://www.saphonenergy.com/
5. K.S. Sholanov. Platform robot manipulator. - WO2015016692 PLATFORM ROBOT MANIPULATOR (wipo.int)
6. K.S. Sholanov. Power plants (variants) on the basis of parallel manipulator. - WO2018147716 ELECTRIC POWER PLANT (VARIANTS) BASED ON A PARALLEL MANIPULATOR (wipo.int)
7. K.S. Sholanov, K.A. Abzhaparov, Zh.T. Zhumasheva, M. Ceccarelli. A new parallel manipulator hydraulically actuated / K.S. Sholanov, K.A. Abzhaparov, Zh.T. Zhumasheva, M. Ceccarelli // International Journal of Mechanics and Control – Libreria Editrice Universitaria: Torino 2016. - Vol. 17, N. 1. - P.4 9-57.

MORPHOLOGICAL FEATURE EXTRACTION FOR LANGUAGE IDENTIFICATION: A COMPREHENSIVE STUDY

Esbolat Sapargali, Assel Jaxylykova
email: e_ sapargali@kbtu.kz, aselya17.89@mail.ru
Faculty of Information Technology, Kazakh-British Technical University, Almaty, Kazakhstan
Institute of Information and Computational Technologies, Almaty, Kazakhstan

Abstract. *The provided article offers a comprehensive exploration of morphological feature extraction for language identification. Language identification is a fundamental task in natural language processing (NLP) that enables language-specific processing and analysis. Extracting morphological features plays a crucial role in language identification by capturing linguistic properties inherent in words. This article presents an in-depth exploration of the significance of morphological feature extraction in language identification. We discuss the interdependencies between Pointwise Mutual Information (PMI) calculation, Naive Discriminative Learning (NDL) algorithm, and heatmap visualization for identifying essential bigrams and trigrams in natural languages. Furthermore, we showcase the potential application of these extracted features in recognizing unknown handwriting through pattern matching. This research may help identify common morphological features in Romance and IndoEuropean languages and shed light on the linguistic characteristics of the enigmatic Voynich Manuscript. Keywords: Voynich manuscript Pointwise Mutual Information · Naive Discriminative Learning · morpheme.*

1 Introduction

The Voynich Manuscript, an enigmatic and mysterious XV century illustrated handwritten code, has puzzled researchers, cryptographers, and linguists for centuries. [1] This ancient manuscript, filled with intricate illustrations and an indecipherable text, has been the subject of numerous studies and investigations in an attempt to unravel its secrets and unlock its hidden meaning. Despite the efforts of countless scholars, the true nature and origin of the Voynich Manuscript remain elusive. The significance of the Voynich Manuscript lies not only in its mysterious content but also in its potential to shed light on historical, linguistic, and cultural aspects of the past. Understanding the language and symbolism encoded within its pages could provide valuable insights into medieval knowledge, science, and art.

In recent years, computational linguistics and statistical analysis have emerged as powerful tools for studying ancient languages and deciphering ancient texts. [2] By applying these modern methodologies to the Voynich Manuscript, researchers have sought to identify linguistic patterns, explore potential language families, and assess the validity of different decipherment hypotheses. In particular, studies focusing on shared bigrams and trigrams between the Voynich language and other known languages have shown promise in revealing the manuscript's underlying linguistic structure. The literature review draws on research articles on the Voynich manuscript Scopus for 2022-2023 to examine various aspects of the manuscript, gradually moving from historical and contextual studies [3] to palaeography[4], cipher analysis and finally to computational linguistics [5] and statistical analysis. [6] Historical and Contextual

Studies provide a foundational understanding of the manuscript's origins, cultural context, and potential purposes. Researchers delve into its illustrations, botanical and zoological depictions, and historical connections to gain insights into its creators and intended audience. Paleography and Scribal Analysis focus on the manuscript's writing style, paleographic features, and scribes. Scholars aim to identify distinctive handwriting styles, determine the number of scribes involved, and potentially link the text to known historical figures. Encipherment Analysis and Historical Ciphers investigate the possibility of the Voynich Manuscript being an encrypted text. Scholars explore historical ciphers and cryptographic methods, analyzing the text's properties to identify underlying cipher systems or cryptographic keys. Text Analysis of Gynaecological and Sexological Manuscripts examines the connections between the Voynich Manuscript and medical texts of the time. Researchers explore similarities and influences to shed light on the possible medical knowledge and practices of the era. Finally, the category of Computational Linguistics and Statistical Analysis employs advanced techniques to study the Voynich Manuscript's language. Our article, which utilizes PMI and NDL metrics to study shared bigrams and trigrams, falls under this category. By arranging the literature review in this order, we establish a smooth transition from historical context and paleographic analysis to encipherment studies, and then to the relevance and significance of computational linguistics and statistical analysis, where our research contributes to deciphering the Voynich Manuscript.

2 Background

In this study, we intend to find the interdependence between computing different methods (including NDL) to identify important bigrams and trigrams in natural languages by searching for optimal hyperparameters. In other words, to conduct a comprehensive study of morphological feature extraction for language identifi- Morphological Feature Extraction for Language Identification 3 cation of arbitrary ancient texts of an unknown Voynich manuscript in order to get an idea of the direction in which our hypothesis should be investigated. Our study suggests that the manuscript may have been encrypted. This assumption is gradually revealed with detail and nuance through past work, starting with R. Zandbergen, where it is described that the solution area may be a meaningful text in a language that has the same properties as the manuscript, or it may be the result of a more complex process that changed the properties of Latin, German, Italian, French and English to those of the Voynich text. [7] But Bennett notes that the entropy values of Voynich's text were significantly lower than those of most comparable works of that era written in natural language. Characters are consistently replaced by others, then counted, which means that the frequency values do not change. This cannot be said of other types of mapping of source characters to target characters. If the manuscript was the product of a simple substitution cipher of a meaningful open text in some known language, then the number of characters and entropy would be similar to other texts in the same language. [8] There is speculation according to Nick Pelling's article that Voynich's manuscript may have used additional characters that are misleading, making it difficult to decipher to modern times (degenerate or redundant coding). The presence of many closely grouped forms in Voynich's manuscript suggests that the cipher system may have used a detailed cipher, where individual letters in the plaintext are ciphered into groups of forged letters. [9] In their study, C. Papadimitriou et al. used the concept of entropy to analyse binary simplifications, i.e. the authors conducted a numerical experiment in which they generated binary character sequences from written texts in natural language, and demonstrated the sensitivity of their approach to language and text genre through statistical measures. [10] Kolmogorov entropy is the most traditional complexity measures which are often based on subjective or reductionist feature selection, this measure is more objective and holistic at the same time it is inherently based on

a radical use of text. Ravi Singh and James Miller present Kolmogorov complexity as a finite compressor which, for any arbitrary string (or Voynich text), produces a minimal description of that string given some form of descriptive language. Hence, practical approaches to arbitrary string compression can be seen as an approximation to an optimal, though unattainable, compressor. In fact, it does not depend on form-function pairs, since the algorithm has no knowledge of the texts to which it applies. [11] Katharina Ehret suggests in her paper that the compression method is an excellent candidate for a universally applicable measure of natural language complexity. This gives additional credibility to the consideration of differences and similarities of works in unknown languages apart from the Shannon method. The Kolmogorov complexity is the minimum length of the description in bits of string: this is the lower bound on compressibility. And at the time, Shannon entropy was interpreted as the number of bits needed to optimally encode a string from a stochastic source [12].

Text conversion methods of the work use binary translation. This gives an unbiased analysis regardless of the number of words in the alphabet (also different transliterations of characters from the Voynich manuscript). Algorithm methods for morphological features from our study previously mentioned only Pointwise Mutual Information in the Reddy and Knight paper. There it was used to measure the average pointwise mutual information over all pairs of words w1, w2 that occur more than once at distance d apart. [13] In conclusion, the NDL method is a powerful tool for analysing and investigating the Voynich manuscript, and its application leads to the discovery of new data and perspectives to unravel this enigmatic writing. It involves counting the repetitions of each pair and trigram in the training text and updating the classes, features and trigrams accordingly. The NDL algorithm allows the extraction of essential morphological features that contribute to language identification.

3 Methodology

In this study, we use the Naive Discriminative Learning (NDL) algorithm, which provides a framework for learning and predicting probabilities for letter pairs and trigrams and a method for calculating pointwise mutual information (PMI) to quantify the strength of association between consecutive pairs of letters in the Voynich manuscript, in medieval manuscripts in Romance and Indo-European languages including English, Latin, Italian, Spanish and French. The formula for PMI (Pointwise Mutual Information) for a bigram (x, y) in the text is as follows:

$$PMI(x,y) = log_2\left(\frac{P(x,y)}{P(x) * P(y)}\right)$$

where, P(x, y) is the probability of the bigram (x, y) occurring in the text. It is calculated as the ratio of the number of occurrences of this bigram to the total number of letters in the text; P(x) is the probability of letter x appearing in the text. It is calculated as the ratio of the number of occurrences of letter x to the total number of letters in the text; P(y) is the probability of letter y appearing in the text. It is calculated as the ratio of the number of occurrences of letter y to the total number of letters in the text. The value of PMI for the bigram (x, y) allows us to assess how strongly the occurrences of letters x and y are correlated in the text. If PMI is positive, it indicates a positive correlation between letters x and y, meaning they frequently occur together. If PMI is negative, it indicates a negative correlation between letters x and y, meaning their joint occurrences are less likely than expected if they were independent. To visualize the associations between letter pairs, we use a heatmap. The heatmap represents the PMI values as a color-coded matrix, where higher PMI values indicate stronger associations between letter pairs. By visualizing the

Morphological Feature Extraction for Language Identification

PMI matrix using the matplotlib library, we gain insights into the prominent associations within the text. The NDL algorithm involves counting the occurrences of each pair and trigram in the training text, and updating the class, feature, and trigram counts accordingly. To handle unseen pairs or trigrams, we apply smoothing techniques using alpha and beta values. The NDL algorithm enables the extraction of significant morphological features that contribute to language identification. Here's a description of calculating the probability for a pair:

$$\text{probability} = \frac{\text{feature count} + \alpha}{\text{class count} + \beta \times \text{total class count}}$$

where, feature count - the number of occurrences of the given pair in the text (counted during training); class count - total number of pairs in the text (counted during training); total class count - total number of pairs (sum of all class counts); alpha and beta - hyperparameters of the algorithm. The working process of the NDL method for analyzing the Voynich manuscript, comparing it with PMI (Pointwise Mutual Information), can be described as follows: Training: The NDL method is first provided with a training text, which is a part of the Voynich manuscript. The NDL method breaks the training text into bigrams (sequences of two characters) and counts the frequency of occurrence of each bigram in the text. NDL then updates the class and feature counters to account for the bigrams from the training text. Prediction: Once trained, the NDL method can be used to predict the probability of trigrams, i.e., sequences of three characters. Bigrams exceeding a certain threshold in terms of NDL probability (more than 0.09) are selected. The NDL method breaks the test text (the rest of the Voynich manuscript) into trigrams and calculates the probability of each trigram using the trained model. Trigrams for which the NDL method calculates a probability above a certain threshold depending on the language (above 0.05 for Voynich) are considered authenticated morphemes. Comparison with PMI: PMI (Pointwise Mutual Information) is a measure used to determine the statistical relationship between two events. In the context of text analysis, PMI is used to identify frequent and related character sequences. In this case, PMI is used to identify common bigrams and trigrams between the Voynich manuscript and other natural languages. PMI is then compared with the results of the NDL method for the Voynich manuscript to see if there are overlaps between the NDL morphemes and the bigrams/trigrams of PMI. We applied the grid search technique to find the optimal values for the smoothing parameters alpha and beta in the Naive Discriminative Learning (NDL) algorithm. The grid search involved systematically testing different combinations of alpha and beta values from a predefined range. We used a set of alpha values and beta values from 0.1 to 1.0 to perform the search. The goal was to identify the alpha and beta values that maximize the average probability of the predicted trigrams on the test text.

Algorithm 1: Grid Search Algorithm

Input: α list: List of alpha values
β list: List of beta values
train text: Training text
pair list: List of pairs of letters
Output: best alpha: Best alpha value
best beta: Best beta value
best avg prob: Best average probability
1 **begin**

```
best alpha ← None;
best beta ← None;
best avg prob ← −1;
foreach α in α list do
  foreach β in β list do
    ndl ← N aiveDiscriminativeLearning(α = α, β = β);
    ndl.train(train text);
    sel tri probs ← [];
    add letter ← 'abcdefghijklmnopqrstuvwxyz' ;
    foreach bigram in pair list do
      foreach letter in add letter do
        trigram ← bigram + letter;
        prob ← ndl.predict(trigram).get(trigram, 0);
        sel tri probs.append((trigram, prob));
      end
    end
    avg prob ← Σ_{trigram,prob∈sel tri probs} prob / len(sel tri probs);
    if avg prob > best avg prob then
      best alpha, best beta, best avg prob ← α, β, avg prob;
      end
    end
  end
  return best alpha, best beta, best avg prob;
end
```

The grid search allowed us to fine-tune the NDL algorithm and optimize its performance on the Voynich Manuscript. By finding the best alpha and beta values, we improved the accuracy of the NDL algorithm in predicting significant morphological features in the manuscript. By evaluating the metrics for Naive Discriminative Learning algorithm is usually done based on the quality of the model prediction (such as accuracy, f1-measure, error matrix, Receiver Operating Characteristic) and comparing its results with other methods or baselines what used in our study. The grid search metric evaluation involves testing the NDL algorithm on known Romance and Indo-European languages, including English, Latin, Italian, Spanish, and French [14].

Morphological Feature Extraction for Language Identification 7 This algorithm performs a grid search for the best values of hyperparameters (alpha and beta) for the Naive Discriminative Learning algorithm (code of Algorithm 1). The code was in the python language, which can be found in the open source github repository 3 . It iterates through the given lists of alpha and beta values and trains the model with each combination of alpha and beta using the training text data. Then, it evaluates the model's performance by computing the average probability for a set of trigrams formed by appending each letter of the alphabet to each pair of letters in the pair list.[15] After iterating through all combinations of alpha and beta, the algorithm selects the best combination that maximises the average probability on the set of trigrams. It returns the best values of alpha and beta, along with with the corresponding best average probability as the result. The heatmap visualization technique was not only applied to the Voynich Manuscript but also extended to medieval manuscripts in the aforementioned Indo-European languages. By comparing the heatmaps of different manuscripts, we can identify

similarities and differences in the associations between letter pairs and explore potential linguistic connections between the Voynich Manuscript and the known languages.

4 Result

To compare the probabilities of trigrams in the Voynich manuscript with other natural languages, the probabilities of trigrams for each language were calculated using the NaiveDiscriminativeLearning model and additionally PMI to find commonality and certainty of bigrams and trigrams. More detailed results for all languages are available in the open github repository (”voynich mopheme”) mentioned in the previous chapter. Bigrams and trigrams were obtained for all languages of the Romance and Indo-European families under study. Commonalities of bigrams and trigrams by NDL and PMI and their heat map on the figure 1 truncated matrix for the Voynich language. And the details can be found in appendix A. Determined via gridsearch the hyperparameters alpha and betta with optimal values of 1 and 0.1 respectively. And also we identified thresholds after bigrams calculation with average values of 0.09 for languages. Probabilities of trigrams in the Voynich manuscript have commonality and differ from probabilities of trigrams in other some natural languages, English, Spanish, Latin, Italian and French. This may indicate the uniqueness and distinctive features of the language of the Voynich manuscript and other languages. According to Voynich ’edy’, ’che’, ’qok’, ’she’, ’iin’, ’hed’, ’oke’, ’aii’, ’kee’, ’oka’, ’kai’, ’eed’, ’eey’, ’ain’ showed high NDL scores. At that time in English, ’the’, ’and’, ’all’, ’ing’, ’her’ were highlighted in the tops (above the threshold of 0.05). Common bigrams and trigrams were generated by NDL and PMI from the Voynich manuscript (Table 1).

Table 1.

Common Bigrams and Trigrams in Voynich Manuscript based on PMI and NDL

Bigrams	Trigrams
or	hey
ey	kee
ed	she
he	qot
ke	che
ok	ked
ar	ote
te	aii
ee	oke
ol	eed
ii	kai
da	ota
ot	hed
sh	eey
ka	iin
in	edy
ai	oka
ta	dai
al	ain
ch	qok
dy	kal
qo	-

Trigrams in the Voynich manuscript have higher probabilities.The Voynich manuscript contains special structures and morphemes that may be peculiar to this unknown and some natural languages (Spanish and English) or writing system. The trigrams with high probabilities in the Voynich manuscript have partially explicit semantic correspondences in other languages. This emphasizes the complexity and uniqueness of the Voynich manuscript, which probably represents an unusual linguistic form. French and Italian (similar distributions in the results), and then below that Latin. Overall, comparing the probabilities of trigrams in the Voynich manuscript with other natural languages allows us to see their unique features and significant differences. Each of these languages has its own characteristics in morpheme usage and word structures that help us better understand the uniqueness and mystery of the Voynich manuscript.

5 Conclusion

In this paper, a comprehensive study of morphological feature extraction for language identification has been carried out. The interdependence between point

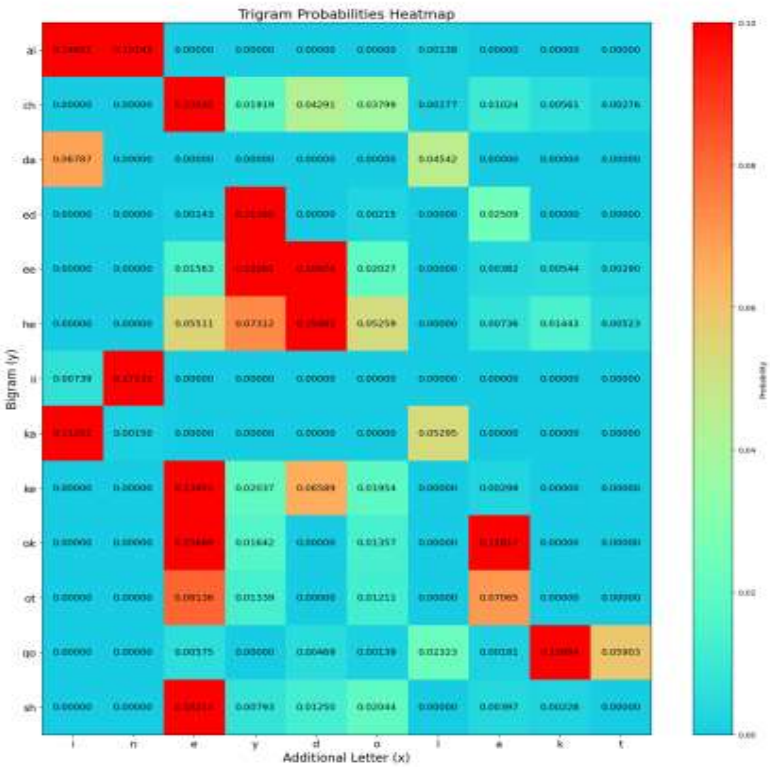


Fig. 1. An example of a truncated heatmap of a pair from a Voynich text

wise mutual information (PMI) computation, naive discriminative learning (NDL) algorithm and heatmap matrix visualisation was examined to identify important bigrams and trigrams in natural languages by searching for optimal alpha and betta hyperparameters (GridSearch) from various possible combinations of morphological features. Morphological feature extraction plays a crucial role in language identification and the analysis of ancient

manuscripts. These morphological features find practical application in recognizing unknown handwriting through pattern matching, showcasing their versatility and potential in diverse NLP domains. Continued research and development in morphological feature extraction will further advance language identification techniques, enhancing the accuracy and effectiveness of NLP applications. Furthermore, this research holds promise in uncovering common morphological features in Romance and Indo-European languages and contributing to the analysis of the enigmatic Voynich Manuscript. The study of the Voynich manuscript using the NDL method represents progress in unraveling this enigmatic writing. The NDL method automates text analysis and identifies authenticated morphemes, which can help linguists and researchers in their search for structures and features of the Voynich language. Although the Voynich manuscript remains unsolved, the NDL results provide new data and hypotheses that can help in understanding this document and its possible origins. Compared to other languages, the Voynich manuscript represents a unique language that may have its own rules and features.

In conclusion, the NDL method is a powerful tool for analyzing and investigating the Voynich manuscript, and its application leads to the discovery of new data and perspectives to unravel this mysterious writing. It may stimulate further research and approaches to understanding the language and content of the Voynich manuscript, opening new horizons in linguistics and history.

6 Acknowledgment

We gratefully acknowledge the financial support of the Ministry of Education and Sciences,Republic of Kazakhstan (Grant num. AP09260670 “Development of methods and algorithms for augmentation of input data for modifying vector embeddings of words”).

References

1. Voynich manuscript. beinecke rare book & manuscript library, yale university. - https://beinecke.library.yale.edu/collections/highlights/voynich-manuscript.
2. Claire Bowern and Daniel Gaskell. Enciphered after all? word-level text metrics are compatible with some types of encipherment, 2022.
3. Demaitre L. Domesticity in Middle Dutch ‘Secrets of Men and Women’ //Social history of medicine. – 2001. – T. 14. – №. 1. – C. 1-25.
4. Davis L. F. How Many Glyphs and How Many Scribes? Digital Paleography and the Voynich Manuscript //Manuscript Studies. – 2021. – T. 5. – №. 1. – C. 6.
5. Kevin F., Colin L., Lonneke P. Demystifying the scribes behind the voynich manuscript using computational linguistic techniques // In CEUR Workshop Proceedings, Virtual, CEUR-WS; Malta, - 2022.- Art.No. 3313.
6. Matlach V., Janeˇckov´a B., Dost´al D. The voynich manuscript: Symbol roles revisited // PLOS ONE, 17:e0260948, 01 2022.
7. Kałużna A., Syguła J. Breaking" The Flo (wer) urishing Code" of The Voynich Manuscript. – 2019.
8. Abou-Assaleh T. et al. Detection of New Malicious Code Using N-grams Signatures //PST. – 2004. – C. 193-196.
9. Kałużna A., Syguła J. Breaking" The Flo (wer) urishing Code" of The Voynich Manuscript. – 2019.
10. Kalimeri M. et al. Entropy analysis of word-length series of natural language texts: Effects of text language and genre //International Journal of Bifurcation and Chaos. – 2012. – T. 22. – №. 09. – C. 1250223.
11. Singh R. I., Miller J. Empirical knowledge discovery by triangulation in computer science //Advances in Computers. – Elsevier, 2010. – T. 80. – C. 163-190.

12. Ehret K. Kolmogorov complexity as a universal measure of language complexity //Proceedings of the First Shared Task on Measuring Language Complexity. – 2018. – С. 8-14.
13. Reddy S., Knight K. What we know about the Voynich manuscript //Proceedings of the 5th ACL-HLT workshop on language technology for cultural heritage, social sciences, and humanities. – 2011. – С. 78-86.
14. Akhmetov I. et al. Highly language-independent word lemmatization using a machine-learning classifier //Computación y Sistemas. – 2020. – Т. 24. – №. 3. – С. 1353-1364.
15. Yang L., Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice //Neurocomputing. – 2020. – Т. 415. – С. 295-316.

ОЦЕНКА ИЗМЕНЕНИЯ ПРОСТРАНСТВЕННО-ВРЕМЕННОГО СОСТОЯНИЯ ОКРУЖАЮЩЕЙ СРЕДЫ РЕГИОНА

Амиргалиев Е.Н, Козбакова А., Мустафин С.А.
email:mustafinsal@mail.ru
Институт информационных и вычислительных технологий
КН МОН РК Казахстан

Аннотация. *В настоящее время одним из важных направлений деятельности человека является исследование состояний природных и техногенных объектов (экологически загрязненных территорий, инженерных и гражданских сооружений, сейсмически активных участков земной поверхности и т. д.) с целью обеспечения определения состояния окружающей среды региона, безопасности и предупреждения чрезвычайных ситуаций. Природные и техногенные катастрофы, происходящие в мире, обусловливают необходимость разработки новых приемов и методов исследования и оценки состояния таких объектов. При решении задачи определения состояний объектов по мониторинговым данным региона основной целью является описание его пространственно-временного состояния, которое в полной мере характеризуется положением в пространстве и во времени. В статье рассмотрен вопрос изменения пространственно-временного состояния окружающей среды по его частичным описаниям.*

Ключевые слова: пространственно-временное состояние, инженерные и гражданские сооружения, природные объекты, мониторинг, контроль, динамический многопараметрический объект, отклонение, оценка изменения состояния окружающей среды.

Материал работы относится к проблемам выбора и сравнения состояний динамического объекта для осуществления комплексного контроля состояния объекта окружающей среды по разнородной информации и может быть применен в системах диагностики и управления для получения достоверной информации о изменения состояния объекта исследования.

Состояние окружающей среды определяется множеством его свойств, каждое из которых выражено в качественных и/или количественных показателях по региону, отнесенных к некоторому фиксированному моменту времени. Свойства окружающей среды изменяются в результате внутренних процессов и/или взаимодействия с внешними факторами. Окружающая среда обладает множеством свойств. При описании состояния окружающей среды региона необходимо сформулировать цель определения состояния и, в зависимости от поставленной цели, из множества свойств (показателей, характеристик) окружающей среды выбрать экспертно те, которые необходимы для его описания поставленной цели [1].

При решении задачи определения состояний окружающей среды по его описаниям основной целью является описание его пространственно-временного состояния окружающей среды, которое в полной мере характеризуется положением в пространстве и во времени.

Обеспечение оценкой состояния окружающей среды связано с решением совокупности задач, относящихся к различным областям деятельности - задачи политического,

социально-экономического, управленческого, технологического характера и т.д. Далеко не последнюю роль в этом списке занимает задача обеспечения безопасности техногенных систем, решение которой дает информацию о пространственно-временном состоянии окружающей среды региона. Как известно абсолютную безопасность техногенной деятельности обеспечить невозможно и поэтому появляется необходимость оценивать меру изменения состояния окружающей среды, т.е. совместное оценивание последствий в различных областям среды. Этим обосновывается важность темы исследований.

В предлагаемой работе рассмотрена проблема изменения состояния окружающей среды в пространстве ее показателей во времени. Цель работы направлена на создание системы мониторинга и контроля процессов окружающей среды региона, позволяющей вести постоянный удаленный мониторинг и помощь в подготовке и принятии управленческих решений для своевременного предупреждения нежелательных ситуаций и их возникновения, последствий, снижение затрат на восстановление среды. Отмечено, что каждый показатель окружающей среды региона в отдельности определяет состояние частично, только одну ее сторону описания, и не может быть глобальной оценкой состояния среды всего региона.

Несмотря на множество методов оценивания изменений состояний объектов на временном интервале, каждый из методов приводит к различным результатам. Это можно объяснить тем, что математическая модель, лежащая в основе того или иного явления, предполагается абсолютно достоверной, а ее свойства при выборе уровня адекватности реальности окончательного решения предоставляется экспертам разработки. Поэтому расширение множества методов оценки изменения окружающей среды на временном интервале является необходимым и актуальным [4-5].

Один из необходимых элементов разработки методов оценки изменения окружающей среды на временном интервале приходится на выбор показателей окружающей среды, измеренных в разных шкалах, и введение скалярного интерпретирующего функционала состояния окружающей среды региона. Что касается выбора показателей окружающей среды региона в данный интервал времени, то его осуществляют опытные специалисты эксперты, экологи.

Среди множества задач оценки окружающей среды выделяют задачу определения пространственно-временного состояния систем. Для ее решения необходим систематический контроль состояний объектов. Существуют разнообразные методы контроля, каждый из которых имеет свои условия применения и цели. Для контроля пространственно-временного состояния объектов необходимы данные об их свойствах во времени. К ним относятся и положение в пространстве и другие свойства, характеризующие взаимное расположение множества точек объекта относительно внешней среды и относительно друг друга. Выполнить непосредственное измерение таких свойств чаще всего затруднено и поэтому для их определения применяют методы математического моделирования [1]. Исходными данными для моделирования служат временные ряды координат множества точек исследуемого объекта, полученные по результатам повторных циклов наблюдений. Анализируя результаты моделирования свойств объекта, можно судить о его состоянии, оценивать изменения этого состояния и принимать необходимые мероприятия для снижения риска возникновения нежелательных аварийных состояний, сопровождающихся не только значительным материальным ущербом, но и человеческими жертвами и связанных с полным или частичным разрушением объекта. Нарушение безопасности жизни рассматриваемого региона

можно было бы предвидеть при постоянном и своевременном выполнении работ по наблюдению и анализу состояния региона [4,5].

Отметим, что среди множества методов описания признаков объектов отдают предпочтение параметрическим методам, которые позволяют моделировать ситуации с той или иной системой показателей на временном интервале изменения свойств объекта.

В связи с появлением новых возможностей получения и обработки пространственной информации об объектах, основанных на использовании современных, аппаратных технологий, стало возможным определять, и как следствие, анализировать изменение пространственного положения объекта в целом, используя формальные процедуры анализа, а не выводить суждение об изменении состояния окружающей среды региона по отдельным выбранным показателям или их части.

Моделирование целостной картины изменения пространственно-временного состояния окружающей среды региона на временном интервале, основанное на формальных методах и процедурах, является по сей день востребованным научным направлением.

Исходными данными для моделирования служат временные ряды показателей исследуемого объекта. Для контроля пространственно-временного состояния окружающей среды региона необходимы данные об их свойствах, как функциях времени. К ним относятся и положение окружающей среды региона как объекта в пространстве выбранных показателях. Выполнить непосредственное измерение свойств среды даже современными техническими средствами не всегда удаётся и поэтому для определения целостной картины изменения пространственно-временного состояния окружающей среды региона необходимо применять подход, используемый в распознавании образов и нашедший применение в различных областях человеческой деятельности. Необходимость разработки новых подходов к решению так поставленных задач является ответом на возникшие задачи на практике.

Использование математического моделирования с привлечением методов интеллектуальной обработки данных позволяет решать задачи, как показывает практика, давая зачастую лучший результат, нежели решения, основанные на классических сложных математических моделях [1,2].

Основными недостатками применяемых методов для оценки изменения ОС являются односторонность описания объекта рассмотрения и необходимость построения индивидуальных градировочных зависимостей по результатам исследований окружающей среды. Цель предлагаемого нами подхода – построение метода сравнения состояний окружающей среды региона за счет использования структуры данных по состояниям окружающей среды во времени. Технический результат предлагаемого нами заключается в повышении достоверности оценки изменения состояния объекта исследования с оперативным определением относительной величины и характера изменения его состояния.

Задача модели – повышение достоверности и оперативности оценки изменения состояния окружающей среды региона.

Содержательная запись предлагаемого метода. Методика оценки изменений состояния сложного динамического объекта по измерениям параметров объекта в моменты времени состоит, по крайней мере, из четырех этапов и заключается в их исполнении [2-4]:

1. выбор параметров состояний окружающей среды региона в момент времени t;
2. измерение параметров состояний окружающей среды региона в моменты времени t_1 и t_2 соответственно;

3. нормализация значений параметров состояния объекта окружающей среды по интервалам изменения выбранных параметров описания среды региона;

4. оценка изменения состояний среды региона согласно следующему решающему правилу:

оценка изменения между состояниями объекта окружающей среды равна средней от суммы приведенных отклонений по выбранным параметрам изучаемой среды.

Таким образом, изложенное и приведенная последовательность этапов обработки данных окружающей среды региона показывает возможность моделирования изменений пространственно-временного состояния окружающей среды по данным мониторинговым наблюдениям по региону.

Благодарности

Работа выполнена в рамках ПЦФ № BR10965172 «Космический мониторинг и ГИС для количественной оценки засоленности почв и деградации сельскохозяйственных угодий Юга Казахстана».

Библиография

1. Дородницын А.А. Проблема математического моделирования в описательных науках // Кибернетика. 1983.- 4, 6–10.
2. Журавлев Ю. И. Избранные труды // М.: Магистр, 1998. - 568 с.
3. Вентцель Е.С. Исследование операций // М.: Наука, 1975. - 610 с.
4. Байконуров О.А. Классификация и выбор методов подземной разработки месторождений // Алма-Ата: Наука, 1969. - 606 с.
5. Аленичев В. М. Системная оценка изменений технологических систем при недропользовании // Вестник ГИАБ. — 2020. — № 5.

УДК 004.8, 004.9

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ИССЛЕДОВАНИЯ ЗОН ПЛАСТОВОГО ОКИСЛЕНИЯ НА УРАНОВЫХ МЕСТОРОЖДЕНИЯХ

Я.И. Кучин[1,2] , Н.Р. Юничева[1,3] А. Сымагулов[1,2], Е. Мухамедиева[1]
[1]Институт информационных и вычислительных технологий МНВО РК, Алматы
[2] КазНИТУ им. К.И. Сатпаева, Алматы,
[3]Алматинский Университет энергетики и связи им. Г. Даукеева, Алматы
E-mail: ykuchin@mail.ru, naduni@mail.ru

Аннотация. При добыче урана методом подземного скважинного выщелачивания (ПСВ) в Казахстане, запасы урана определяются путем пересчета зарегистрированного естественного гамма-излучения радия и других продуктов распада урана в содержание урана. В зонах пластового окисления (ЗПО), при наличии ярко выраженных гамма-аномалий, уран полностью отсутствует, поэтому их необходимо учитывать при подсчёте запасов урана. Рассмотрение зон пластового окисления (ЗПО) при планировании добычи урана методом подземного скважинного выщелачивания является одной из актуальных задач при добыче урана. Этот факт позволит подтвердить расчетные запасы урана и сэкономить крупные материальные ресурсы, так как из обработки будут исключены пустые добычные блоки. В статье рассматривается возможность выделения ЗПО методами машинного обучения на основе данных каротажа. Для этого скважины были разделены на три класса в зависимости от характеристик ЗПО. RandomForestClassifier показал среднюю точность выделения ЗПО 0.55, однако значения сильно варьировались в зависимости от класса скважин.

Введение. Особенности формирования месторождений предопределяют их радиологическую обстановку и элементный состав. При этом, как правило, в рудной залежи наблюдается недостаток радия в сравнении с равновесным состоянием, а в ее обрамлении радиоактивное равновесие нарушено в сторону избытка радия, что является следствием образования т. н. «остаточных» и «диффузионных» ореолов радия. Это связано с тем, что, месторождения пластово-инфильтрационного типа формируются в осадочных проницаемых толщах горных пород на границе окислительно-восстановительного барьера. А поскольку в окислительной и восстановительной обстановках поведение подвижных форм урана и радия существенно различаются, в различных морфологических элементах рудных тел в результате процессов «выноса-привноса» «материнского» урана и «дочернего» радия возникают геохимические зоны, где соотношения массовых долей радия и урана отличаются от значений, соответствующих состоянию радиоактивного равновесия между ними.

Состояние радиоактивного равновесия между радием и ураном принято характеризовать коэффициентом нарушения радиоактивного равновесия (или просто коэффициентом радиоактивного равновесия) $K_{рр}$, который равен отношению массовых долей радия и урана.

Таким образом, значение величины $К_{рр}$ = 1 соответствует наличию радиоактивного равновесия, а отличие значений $К_{рр}$ от единицы фиксирует наличие систем, не достигших равновесия либо претерпевших нарушения их замкнутости.

В настоящее время содержание урана определяется путем деления содержания радия, полученного в результате интерпретации гамма-каротажа (ГК), на Крр [1].

При ГК регистрируется естественное гамма-излучение радия и других продуктов распада урана. Поскольку в зонах пластового окисления Крр=∞, т.е. при наличии ярко выраженных гамма-аномалий, уран полностью отсутствует, их необходимо учитывать при интерпретации гамма-каротажа. ЗПО можно выделять при анализе керна на этапе разведки, а затем строить геологические разрезы и протягивать их по геологическим разрезам.

Одним из способов выделения ЗПО является проведение каротажа нейтронов деления (КНД), позволяющим определять содержание урана в естественном залегании, исключая ошибки, связанные с изменением коэффициента радиоактивного равновесия в процессе выщелачивания урана, а также дает возможность оценивать геотехнологические параметры (влажность, глинистость, пористость). Также этот метод позволяет непосредственно определять содержания урана, минуя стадию пересчета содержания радия в содержание урана через Крр. В таких случаях ЗПО выделяются там, где рудным интервалам по гамма-каротажу не соответствуют рудные интервалы по КНД.

Пример переинтерпретации ГК после выделения ЗПО по результатам КНД приведен на рисунке 1.

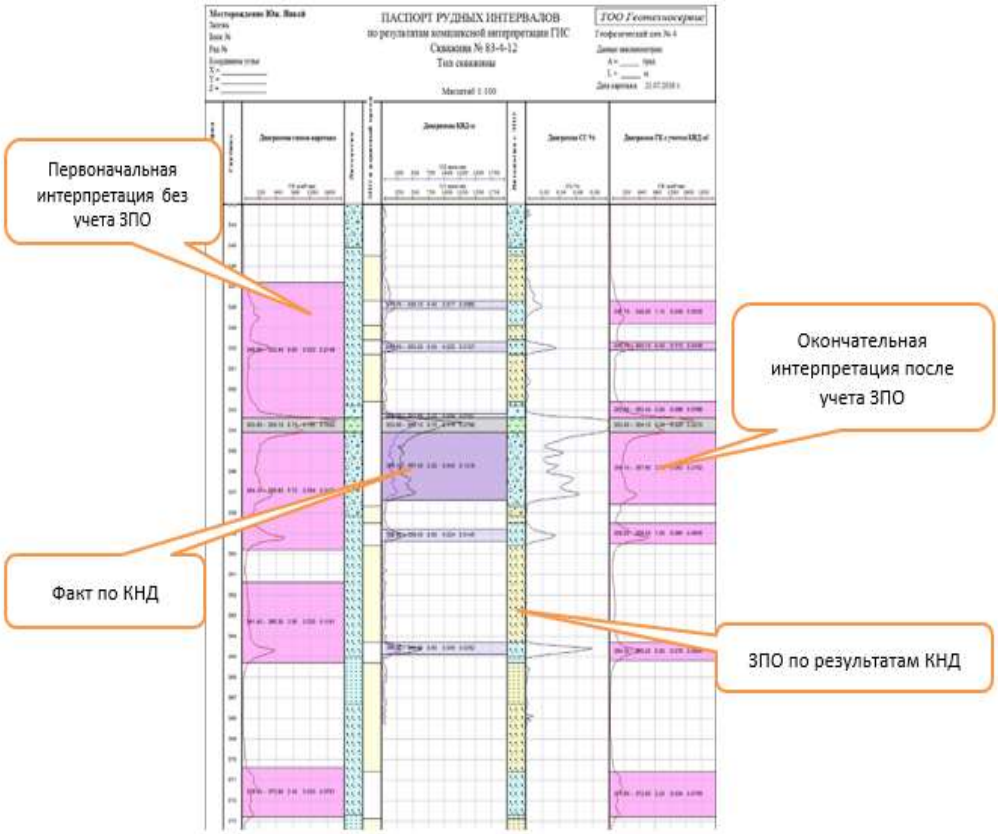


Рисунок 1. Пример переинтерпретации ГК после выделения ЗПО по результатам КНД.

На рисунке видно, что фактические рудные интервалы по КНД оказались значительно меньше рассчитанных по радию. Однако проведение КНД стоит дорого, скорость каротажа составляет не более 50 м/ч (скорость проведения ГК составляет до 1000 м/ч), но главное ограничение состоит в том, что ресурс используемой трубки генератора нейтронов крайне ограничен. Поэтому КНД заказывают не на всех месторождениях и в небольших объемах (5-10%) от общего количества скважин.

Методы. Неучет ЗПО является одной из основных причин неподтверждения запасов и часто приводит к существенным материальным потерям, когда целые блоки оказываются пустыми. В то же время на данный момент нет быстрого и надежного способа выделения ЗПО для корректной интерпретации ГК. Создание формализованного алгоритма выделения ЗПО также представляется невозможным.

В этой связи одним из перспективных направлений решения данной проблемы может стать машинное обучение (МО), когда предсказание расположения ЗПО будет осуществляться путем анализа всей совокупности имеющейся геологической и геофизической информации, путем выявления закономерностей, не замечаемых людьми. Авторам не удалось найти публикации на данную тему, хотя в целом, машинное обучение широко применяется в области анализа геологических данных [2-15].

Для оценки возможности применения МО для выявления ЗПО предложено провести целый ряд исследований, возможная схема которых приведена на рисунке 2.

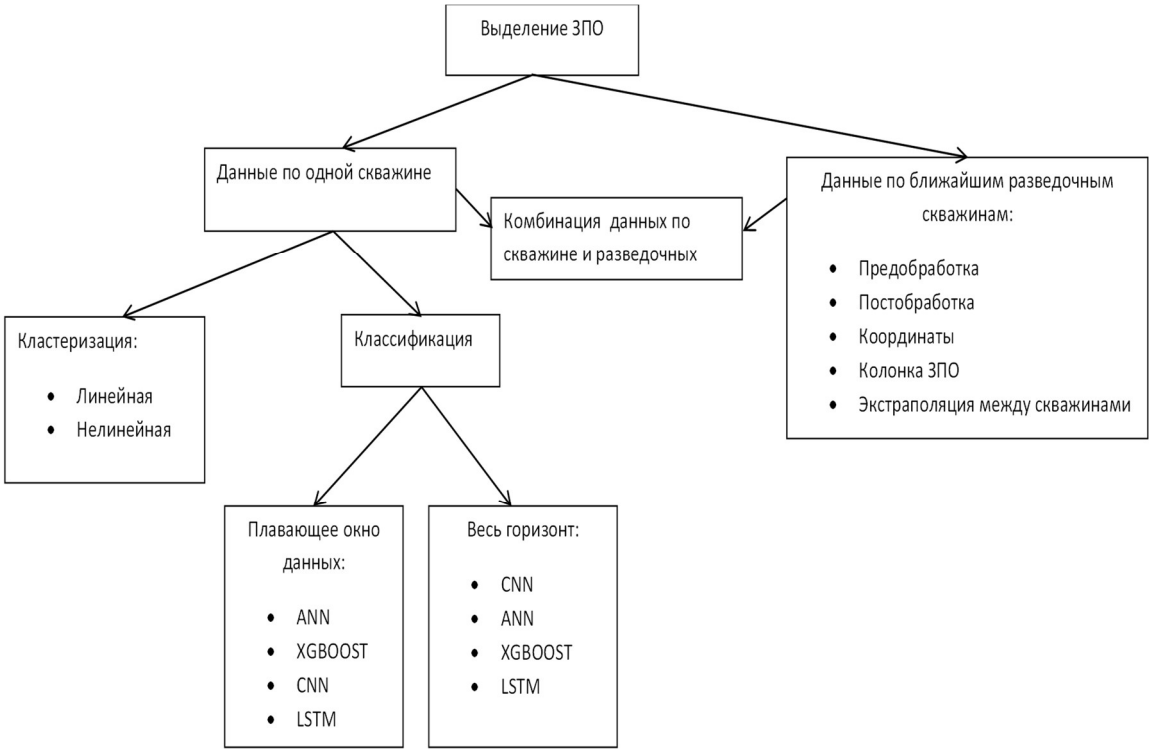


Рисунок 2. Схема исследования применимости методов МО для выделения ЗПО

Таким образом, методологическая схема исследования состоит из следующих этапов:

– Создание специального набора данных с разделением скважин на три класса: 42 скважины без ЗПО (LOW_ZPO), 84 скважины с долей ЗПО 5-50% от рудовмещающего горизонта (MEDIUM ZPO) и 42 скважины с долей ЗПО более 50% рудовмещающего горизонта (HI_ZPO).

В качестве входных параметров использовались данные гамма-каротажа, электрокаротажа (КС, ПС), литология (поскольку ЗПО может находиться только в проницаемых породах), глубина и координаты устья скважины, а также данные по ближайшей соседней скважине.

–Разработка методов предобработки данных;

–Применение методов машинного обучения.

В качестве классификатора использовался RandomForestClassifier (RFC), который представляет собой ансамбль деревьев решений [16] (рис. 3). Это один из самых гибких и простых в использовании алгоритмов. RFC создает деревья решений для случайно выбранных семплов данных, получает прогноз от каждого дерева и выбирает наилучшее решение посредством голосования. Он также предоставляет довольно эффективный критерий важности показателей (признаков).

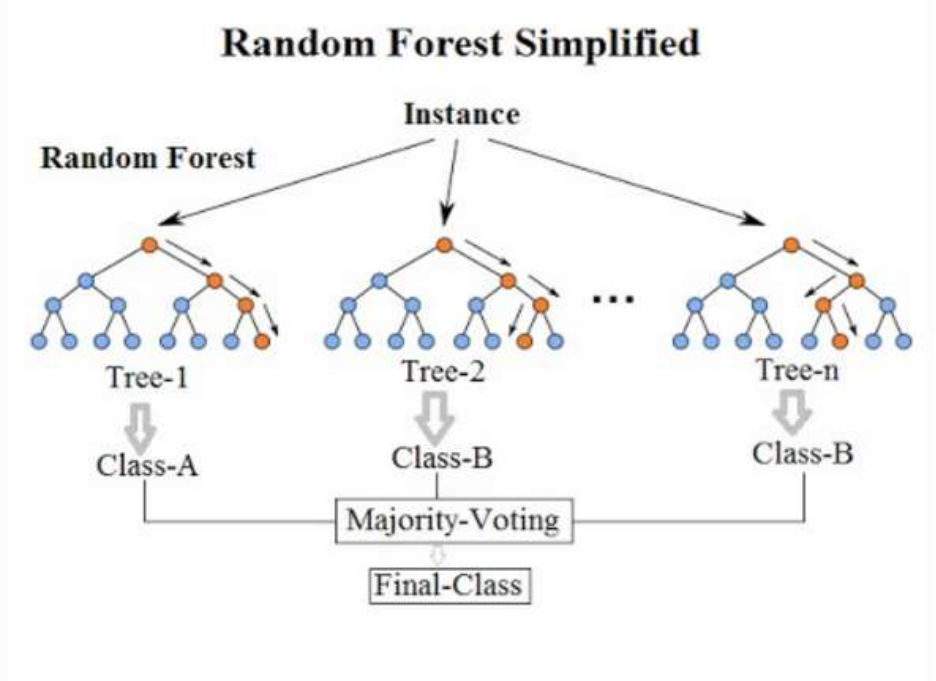


Рисунок 3. Классификатор RandomForestClassifier

Полученные результаты. На данный момент проведены предварительные эксперименты применения модели RandomForestClassifier на плавающем окне данных. В ходе экспериментов, точность определения ЗПО сильно варьировалась в зависимости от класса скважины, но в среднем составила 0.55, что позволяет сделать вывод о принципиальной возможности решения задачи выделения ЗПО методами МО. Для оценки работы классификатора был разработан простейший визуализатор, который показывает фактические (синие) и предсказанные (красные) трех классов: 1-проницаемые породы, 2-непроницаемые породы, 8 – ЗПО, при этом ось Y представляет собой глубину в пределах рудовмещающего горизонта с шагом 10см. Примеры результатов выделения ЗПО с помощью вышеупомянутой модели на разных классах скважин (LOW_ZPO, MEDIUM_ZPO и HI_ZPO) приведены на рис. 4, 5 и 6.

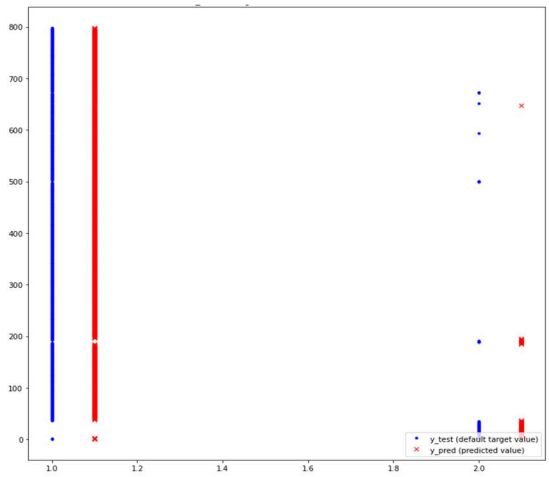


Рисунок 4. Пример фактических и предсказанных классов пород для скважины LOW_ZPO

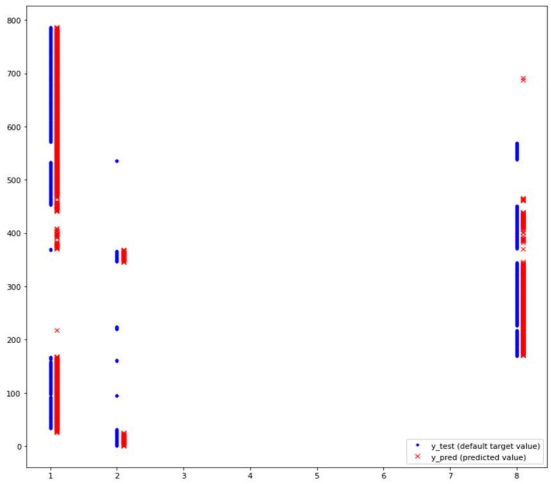


Рисунок 5. Пример фактических и предсказанных классов пород для скважины MEDIUM_ZPO

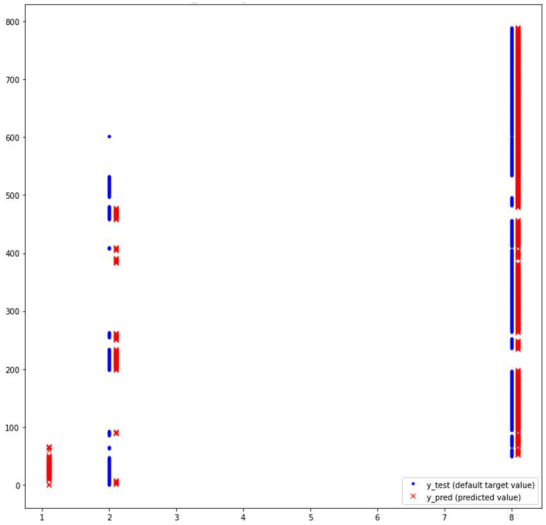


Рисунок 6. Пример фактических и предсказанных классов пород для скважины HI_ZPO

Заключение. Таким образом, неточности в определении литологического состава, фильтрационных характеристик и ЗПО приводят к ошибкам в технологическом процессе установки фильтров, ошибкам в определении запасов руды и в конечном счете существенным финансовым потерям. Существующие методики решения перечисленных проблем разработаны начиная с 70-х годов прошлого века, не поддаются существенной модернизации и требуют большое количество высокопрофессионального ручного труда.

Таким образом, для преодоления вышеперечисленных трудностей, в данной статье исследован вопрос возможности выделения зон пластового окисления методами машинного обучения.

Для проверки справедливости данной гипотезы создан специальный набор данных с разделением скважин на три класса: скважины без ЗПО, скважины с долей ЗПО 5-50% от рудовмещающего горизонта и скважины с долей ЗПО более 50% рудовмещающего горизонта. В качестве входных параметров использовались данные гамма-каротажа, электрокаротажа (КС, ПС), литология (поскольку ЗПО может находиться только в проницаемых породах), глубина и координаты устья скважины, а также данные по ближайшей соседней скважине. Далее проведена процедура разработки методов предобработки данных и применены методы машинного обучения.

В результате экспериментов можно сделать вывод о возможности применения методов машинного обучения для выделения ЗПО Тем не менее, учитывая важность и ответственность выделения ЗПО при подсчете запасов, для оценки возможности промышленного применения алгоритма выделения ЗПО методами МО, необходимо провести все исследования согласно схеме исследования применимости методов МО для выделения ЗПО (рисунок 2).

Благодарности. Работа выполнена при финансовой поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан Республики Казахстан (грант: № AP14869110 "Повышение точности решения задач интерпретации данных геофизических исследований скважин на урановых месторождениях с использованием методов машинного обучения.

Литература

1. Инструкция (методические рекомендации) по подземному подземному выщелачиванию урана. -2006. - 315 с.
2. Амиргалиев Е., Искаков С., Кучин Я., Мухамедиев Р. Методы машинного обучения в задачах распознавания пород на урановых месторождениях // Известия НАН РК: Серия физико-математических наук. -2013. - № 3. – С.82-88.
3. Крекнелл М. Машинное обучение для геологического картирования: Алгоритмы и приложения". Докторская диссертация, Школа наук о Земле, Университет Тасмании, Хобарт, Австралия. -2014.
4. Кумар С, Чаттерджи С, Ооммен Т, Гуха А. Автоматизированное литологическое картирование путем интеграции методов спектрального усиления и алгоритмов машинного обучения с использованием гиперспектральных данных AVIRIS-NG в золотоносных гранитно-зеленокаменных породах в Хатти, Индия // Междунар. Журнал прикладных наблюдений Земли. Геоинф. – 2020. - № 86, статья 102006.
5. Кучин Я., Якунин К, Мухамедиева Е, Мухамедиев Р. Проект по созданию классификатора литологических типов для урановых месторождений Казахстана // Физика: Серия конференций. – 2020. - № 1405(1). – С. 012001.
6. Кучин Я., Мухамедиев Р., Якунин К. Качество классификации данных при непоследовательных экспертных оценках // Облако науки. -2019. - № 6(1). -С.109-126.
7. Мухамедиев Р., Амиргалиев Е., Искаков С., Кучин Я., Мухамедиева Е. (2014) Интеграция результатов алгоритмов распознавания на урановых месторождениях. Журнал передового вычислительного интеллекта и интеллектуальной информатики, 18(3):347-352.
8. Мухамедиев Р.И. и др. Оценка фильтрационных свойств вмещающих пород в урановых месторождениях песчаникового типа с использованием методов машинного обучения // IEEE Access. – 2022. - №10. – С.18855-18872, DOI: 10.1109/ACCESS.2022.3149625.
9. Мухамедиев Р.И., Кучин Я.А., Якунин К.О., Мухамедиева Е.Л., Костарев С.В. Предварительные результаты оценки литологических классификаторов для урановых месторождений инфильтрационного типа // Облако науки. -2020. -№7(2). -С.258-272.

10. Мухамедиев Р.И., Сымагулов А., Кучин Ю., Якунин К. и Елис М. (2021) От классического машинного обучения к глубоким нейронным сетям: Упрощенный наукометрический обзор // Прикладн. науки. – 2021. -№ 11(12). -С.554.
11. Нгуен Г. Среды и библиотеки машинного обучения и глубокого обучения для крупномасштабного интеллектуального анализа данных: опрос // Обзор искусственного интеллекта. -2019. -№ 52(1). -С.77-124.
12. Сан Т., Х Ли, К Ву, Ф Ч, З Жу, З Ху Моделирование перспективности полезных ископаемых на основе данных с использованием методов машинного обучения и глубокого обучения: Исследование на примере южной провинции Цзянси, Китай // Минералы. -2020. -№ 10(2). - С.102.
13. Чэнсян Д, Хэпин П, Синан Ф, Конате А, Руйдун Ц. Машина опорных векторов как альтернативный метод для классификации литологии кристаллических пород //Журнал геофизики и инженерии. -2017. -№ 14(2). – С. 341-349.
14. Кучин Я., Мухамедиев Р.,. Якунин К., Грундспенкис Дж. Сымагулов А. Оценка влияния экспертной маркировки обучающих данных на качество автоматической классификации литологических групп с помощью искусственных нейронных сетей // Прикладные компьютерные системы. -2020. -№ 25(2). -С145-152, https://doi.org/10.2478/acss-2020-0016.
15. Кучин Ю.И., Мухамедиев Р.И., Якунин К.О. Один из методов генерации синтетических данных для оценки верхней границы производительности алгоритмов машинного обучения // Cogent Engineering. – 2020. - № 7(1). -С.1718821.
16. Хо Т. Метод случайного леса. – Матер. 3-й Междунар. Конф. по анализу и распознаванию документов IEEE". -1995. -№ 1. -С.278-282.

ИССЛЕДОВАНИЕ ИСПОЛЬЗОВАНИЯ НЕЙРОННЫХ СЕТЕЙ И ПРИЛОЖЕНИЙ С ИИ ДЛЯ АВТОМАТИЗАЦИИ ПРОЦЕССОВ НАПИСАНИЯ ТЕКСТОВОГО ОПИСАНИЯ ПРОДУКТОВ

Киселева О.В., Майнцер Д.А., Молдабеков Ә. Т.
e-mail: o.kisseleva@turan-edu.kz, d.maintser@turan-edu.kz, 22232108@turan-edu.kz
Учреждение «Университет Туран»

Аннотация. *В современном мире, где электронная коммерция становится все более популярной, важным фактором успешного онлайн-бизнеса является представление товаров в интернет-магазинах. Один из крупнейших онлайн-магазинов мира, Amazon, имеет множество товаров в своем каталоге, которые необходимо описывать для привлечения внимания и увеличения продаж.*

Создание приложения с искусственным интеллектом для автоматического написания описаний продуктов на Amazon является одной из инновационных разработок, которая до сих пор не получила широкого распространения на рынке электронной коммерции и редко используется в данной сфере. ИИ в последние годы стал одним из ключевых направлений в развитии технологий и науки. В данной статье представлен проект по созданию приложения с внутренней интеграцией искусственного интеллекта для автоматического написания описаний товаров, которые размещаются в интернет-магазине Amazon.

Цель исследования заключается в разработке и реализации приложения с ИИ, которая будет способна генерировать описания товаров на основе их названия и преимуществ, учитывая особенности бизнес-предметной области Amazon, а также получение удовлетворительных результатов конверсии, в сравнении с ручным написанием информации человеком.

Для достижения поставленной цели были решены следующие задачи: анализ существующих методов и алгоритмов для интеграции ИИ; использования существующих моделей ИИ для написания описаний продуктов; изучения алгоритмов создания и функционирования искусственных интеллектов; создание интерфейса для взаимодействия с пользователем.

Теоретическая и методологическая основа данного исследования основывается на анализе существующих методов генерации текста, анализе бизнес-предметной области Amazon и изучении модели, использующей глубокое машинное обучение - модели языковых представлений (language models) и алгоритмы обработки естественного языка.

Для реализации была использована модель GPT (Generative Pre-trained Transformer), разработанная компанией OpenAI, которая представляет собой одну из наиболее продвинутых моделей языковых представлений на сегодняшний день. Она была обучена на огромном объеме текстовых данных, включая тексты из интернета, книги, научные статьи и другие источники информации.

Также в данной работе было использовано несколько подходов для создания веб-приложения, а именно разработка с дополнительным обучением ИИ модели на основе GPT и разработка с использованием открытого ключа API от компании OpenAI.

В процессе реализации первого вышеупомянутого подхода, после этапа проектирования модели, она была дообучена на специфическом наборе данных, состоящем из описаний продуктов на Amazon. Для этого был использован подход fine-tuning, который заключается в дополнительном обучении модели на массивном наборе данных, специфичном для конкретной задачи. Это позволяет генерировать описания товаров, которые соответствуют стилю и формату, используемому на Amazon.

Также в рамках данного исследования был создан локальный веб-интерфейс для более удобного и понятного использования ИИ, для написания которого был использован язык программирования Python, библиотека Streamlit для создания веб-приложения и другие языковые компонующие.

Конечный интерфейс имеет простой дизайн и состоит из нескольких страниц. На главной странице пользователь может ввести название товара и его преимущества, после чего приложение генерирует описание товара на основе нейронной сети.

Также в интерфейсе присутствует возможность сохранения сгенерированного описания в базу данных и просмотра всех сохраненных описаний на отдельной странице.

В первом подходе для улучшения качества генерации описания товара были использованы различные методы обработки естественного языка, такие как лемматизация, удаление стоп-слов и применение частеречной разметки.

Практическая значимость заключается в том, что разработка искусственного интеллекта может значительно снизить затраты на ручную работу, повысить качество и количество продаж, а также ускорить процесс и сократить время для создания описаний товаров. Данная разработка может стать основой для создания более сложных ИИ для других задач. Также данное исследование имеет практическую значимость не только на платформе Amazon, но и на других платформах электронной коммерции.

Актуальность исследования заключается в том, что растущее количество продавцов на платформе Amazon и необходимость повышения эффективности их работы, ставят перед бизнес-сообществом новые вызовы и задачи. В частности, одной из таких задач является увеличение конкурентоспособности продуктов за счет улучшения описаний продуктов. Однако, данная задача является трудоемким процессом, который занимает значительное количество времени и ресурсов.

В связи с этим, создание веб-приложения с интеграцией искусственного интеллекта для автоматизации процесса написания описаний продуктов на Amazon становится актуальной задачей. Использование такого инструмента не только сокращает время и ресурсы, затрачиваемые на написание описания продуктов, но и повышает качество текста, что позволяет улучшить конкурентоспособность продуктов и повысить удовлетворенность потребителей.

Одним из ключевых преимуществ данной разработки является ее способность адаптироваться к различным категориям продуктов на платформе Amazon. Благодаря этому система сможет генерировать качественные описания продуктов в любой категории, что даст возможность использовать ее для любого типа продаж.

Кроме того, система будет иметь возможность обучаться на основе данных о продажах и предпочтениях потребителей, что позволит ей становиться все более точной и эффективной в генерировании описаний продуктов.

Модель ИИ GBT2 была обучена на основе информации о продуктах, которые имеют успешную репутацию на рынке. Для анализа и сбора данных используется сервис

Hellium10. Hellium10 представляет собой комплексный инструмент для продавцов на платформе Amazon, который предоставляет различные функции, такие как исследование ключевых слов, анализ конкурентов, оценка потенциала продаж и многое другое.

Использование Hellium10 позволяет получить более точные и актуальные данные о продуктах с высокой конверсией, что в свою очередь улучшает качество обучения модели.

С помощью Hellium10 можно выявить успешные продукты, их характеристики и особенности описания. Эти данные затем используются для формирования обучающей выборки для модели, что обеспечивает ее обучение на актуальных и проверенных примерах успешных товаров. Такой подход обеспечивает более эффективное обучение и позволяет создать модель, способную генерировать описания товаров, которые могут повысить конверсию и привлечь больше покупателей.

Приложение с искусственным интеллектом, оно было протестировано на реальных данных в компании Kiki-green. Данная компания является одним из конкурентноспособных в сфере продажи пищевых добавок, бадов, витаминов на маркетплейсе Amazon. Kiki-green на сегодняшний день оценивается в 1.700.000$.

В процессе внедрения данного исследования, было получено официальное разрешение от руководство компании Kiki-green, для внедрения и использования разработанного проекта для изменения описания уже существующих продуктов.

Продукт KIKI Green Sea Moss Powder with Elderberry был впервые опубликован в октябре 2022 года. Изменение описания продукта было произведено 1 февраля 2023 года посредством использование разработанного веб-приложения. Данный период был выбран посредством анализа рынка, так как в период с декабря по январь, является сначала повышающим конверсию, а после понижающим, ввиду праздников и иных событий. Февраль месяц является стабильным, что позволяет наглядно увидеть результат конверсии. Пример, как выглядит официальная страница продукта, отражена на рисунке 1. Результат генерации описания в пользовательском интерфейсе отражен на рисунке 2.

До изменения описания продукта KIKI Green Sea Moss Powder with Elderberry, наблюдался убывающий тренд по продажам. За январь месяц конверсия составила 9.593$. Показатели на 1 февраля 2023 года отражены на рисунке 3.

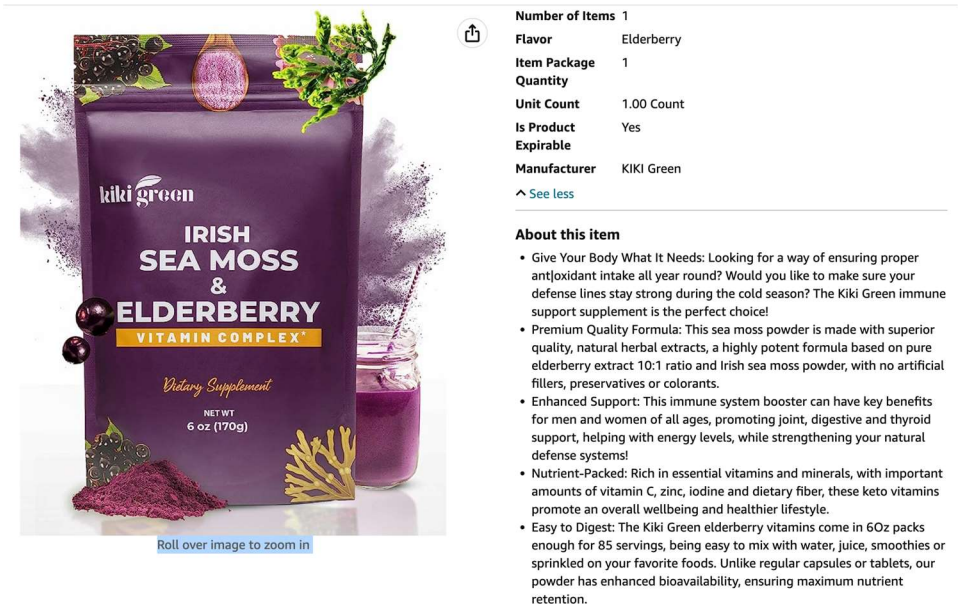


Рис 1. Официальный сайт продукта KIKI Green Sea Moss Powder with Elderberry на Amazon

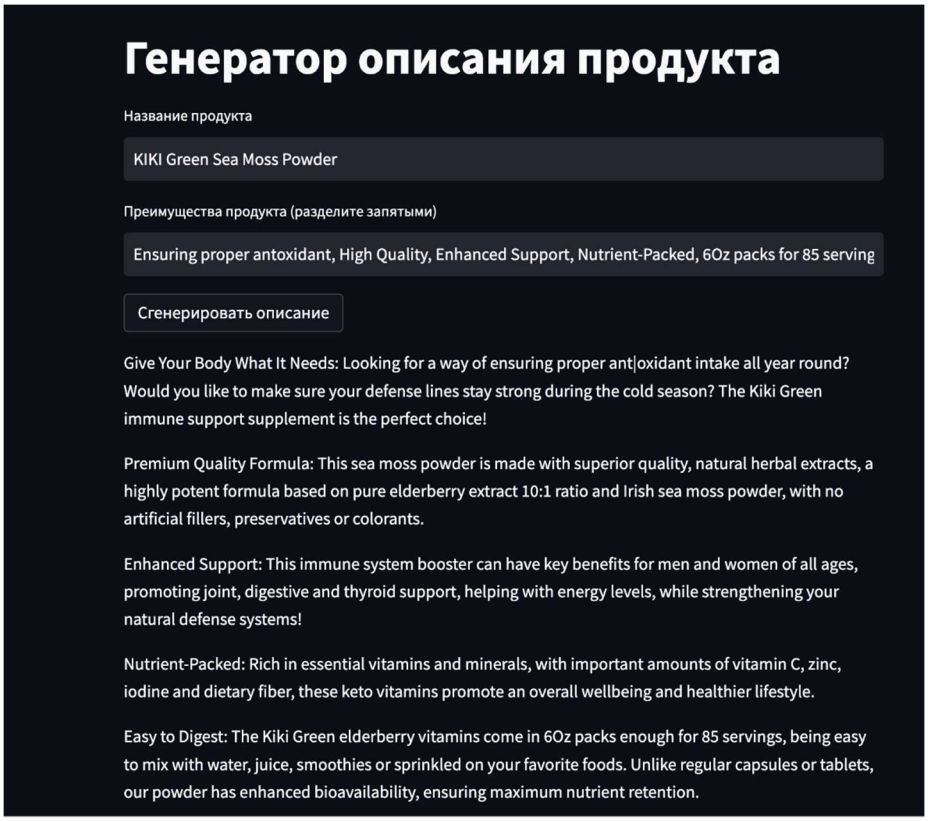


Рис. 2 Результат генерации описания для продукта KIKI Green Sea Moss Powder with Elderberry (составлено автором)

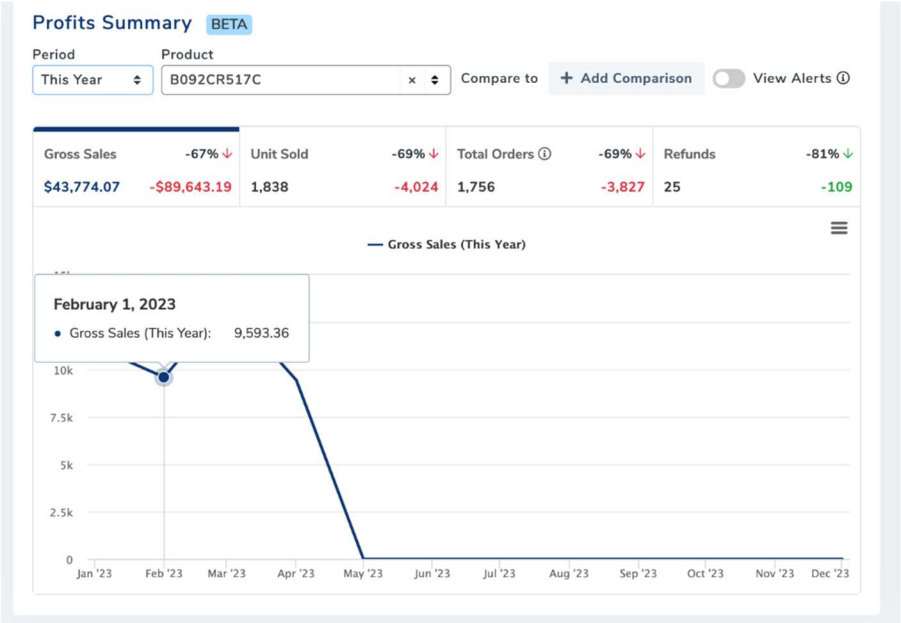


Рис. 3 Статистические показатели конверсии продукта KIKI Green Sea Moss Powder with Elderberry за январь месяц 2023 года

После изменений в конце февраля, в начале марта конверсия выросла на 3.948$, что означает ее повышение на 41%.

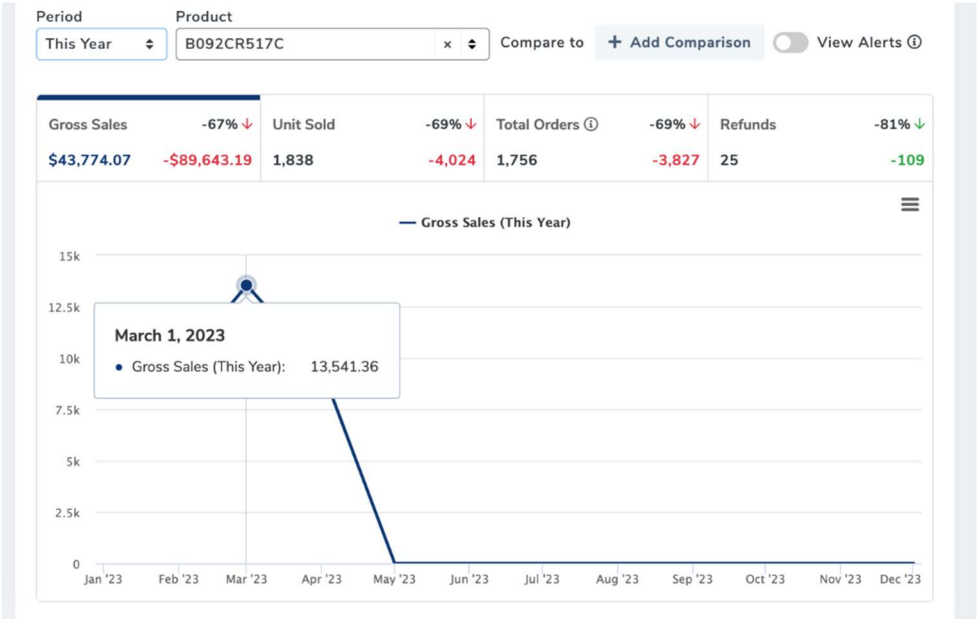


Рис. 4 Статистические показатели конверсии продукта KIKI Green Sea Moss Powder with Elderberry за февраль месяц 2023 года

Вторым продуктом является KIKI Green Irish Sea Moss Powder 8 Oz, который был впервые опубликован в сентябре 2022 года. Изменение описания продукта было произведено 1 февраля 2023 года. Данный период был выбран посредством анализа рынка, так как в период с декабрь по январь, является сначала повышающим конверсию, а после понижающий, ввиду праздников и иных событий. Февраль месяц является стабильным, что позволяет наглядно увидеть результат конверсии.

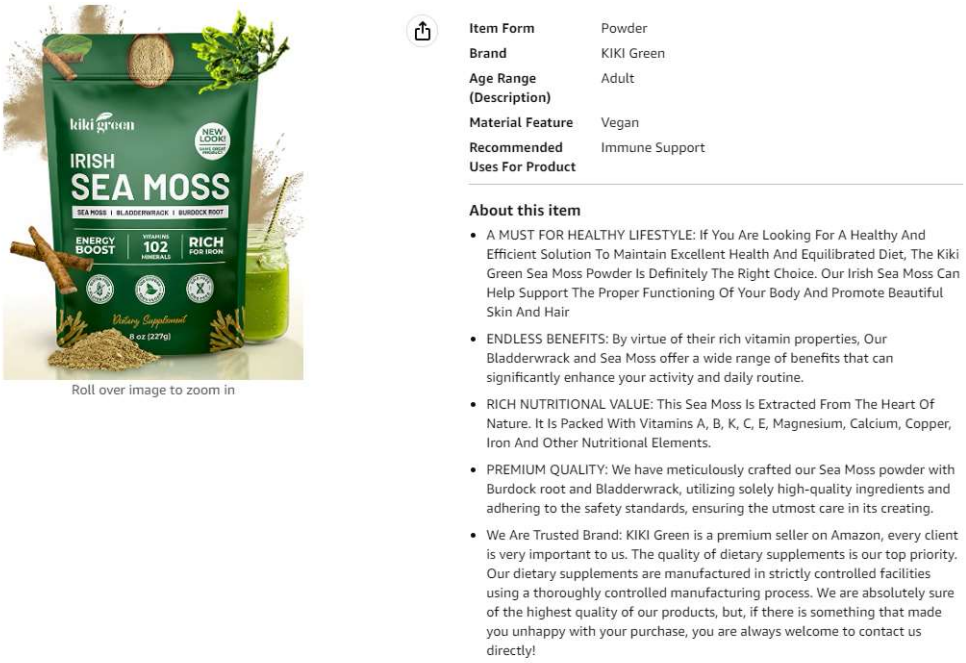


Рис. 5 Официальный сайт продукта KIKI Green Irish Sea Moss Powder на Amazon

Рис. 6 Результат генерации описания для продукта KIKI Green Irish Sea Moss Powder 8 Oz (составлено автором)

До изменения описания продукта KIKI Green Irish Sea Moss Powder 8 Oz, также наблюдался убывающий тренд по продажам. За январь месяц конверсия составила 9.203$.

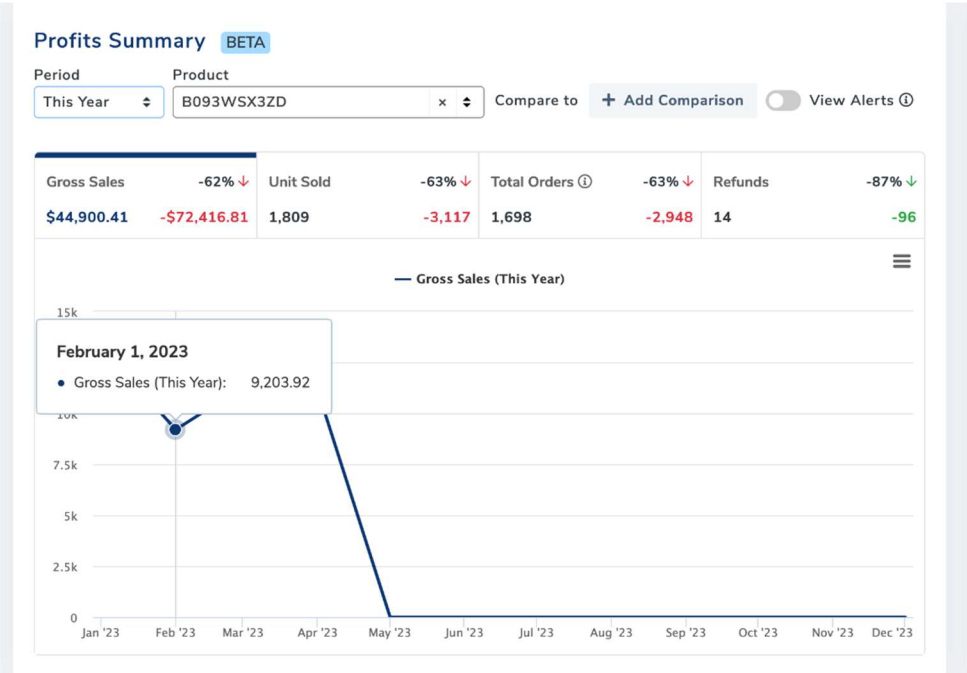


Рис. 7 – Статистические показатели конверсии продукта KIKI Green Irish Sea Moss Powder 8 Oz за явнварь месяц 2023 года

После изменений в конце февраля, в начале марта конверсия выросла на 2.108$, что означает ее повышение на 22%.

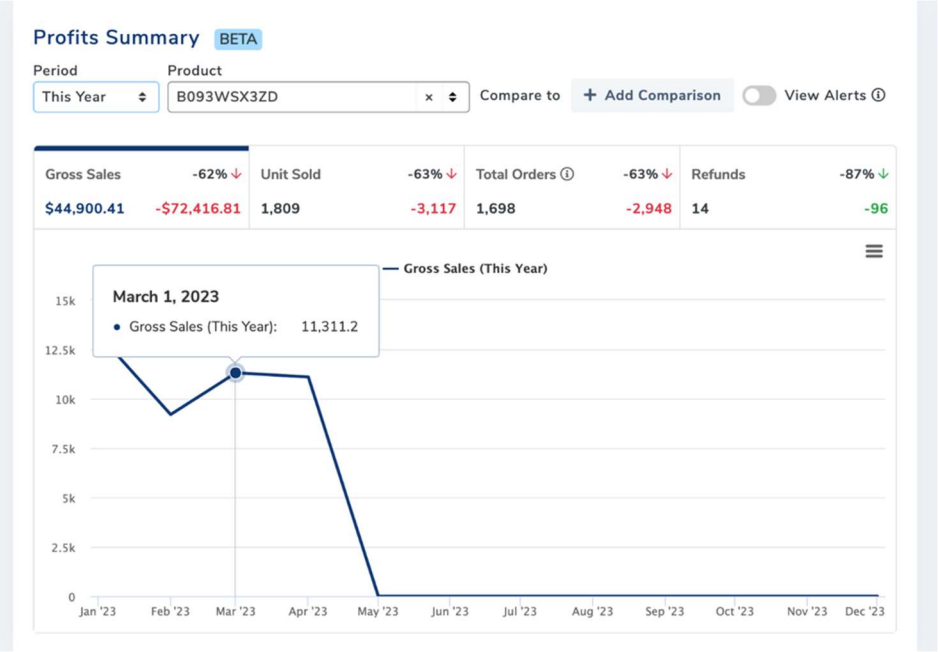


Рис. 8 Статистические показатели конверсии продукта KIKI Green Irish Sea Moss Powder 8 Oz за февраль месяц 2023 года

Использование приложения с ИИ для написания описания продуктов Amazon привело к положительным результатам. Конверсия продукта увеличилась от 20% до 40%, что привело к увеличению количества продаж. Кроме того, реальные покупатели начали оставлять положительные отзывы о продукте, а их количество достигло 30.

Таким образом, данный проект может быть использован для создания уникальных описаний товаров и услуг, что приводит к увеличению конверсии и продаж. Использование ИИ также существенно сокращает время, необходимое для написания описания, что позволяет сосредоточиться на других задачах и увеличить эффективность работы.

Использование нейронных сетей и моделей на основе искусственного интеллекта, таких как GPT-2, является перспективным направлением в области автоматизации процессов написания текстов для различных целей. В данной работе были использованы различные инструменты и библиотеки, такие как Selenium, GPT2Tokenizer, TFGPT2LMHeadModel и др., для создания модели, обучения ее на доступных данных и последующего использования для генерации текстового описания продуктов.

Было проведено исследование и сбор данных о продуктах успешных компаний с высокой конверсией на Amazon с помощью сервиса Hellium10. После этого была использована модель ИИ на основе GPT-2, которая была переобучена на полученных данных. Были выполнены все поставленные задачи, а полученные результаты позволили значительно упростить процесс написания описаний товаров и повысить их качество.

Одним из ключевых достижений данного исследования является создание системы, которая позволяет автоматически генерировать описания товаров на основе названия продукта и его преимуществ. Такой подход является весьма эффективным, поскольку позволяет сэкономить время на написание описаний и улучшить их качество.

В рамках исследования были проведены работы по сбору и предобработке данных, интеграция и дополнительное обучению модели искусственного интеллекта, а также разработке интерфейса для взаимодействия с пользователем. Были использованы современные технологии и инструменты для обработки и анализа данных.

На первом этапе было проведено исследование предметной области Amazon, были выделены все аспекты, относящиеся к данной бизнес области. Также был проанализирован процесс создания названия и формирования описания для продукта, составлены требования и выделена релевантность и целесообразность реализации.

На втором этапе были выделены этапы проектирования, определены инструменты, языки программирования, библиотеки и методы. Также проведен анализ современных технологий для сбора данных, обработки информации, обучения модели искусственного интеллекта GBT2, создания простого и понятного пользовательского интерфейса.

В третьем этапе, основываясь на ранее полученной теоретический информации, был реализован этап проектирования приложения с использованием модели искусственного интеллекта и создание программы для сбора и обработки информации. Также в ходе исследования был выбран более простой способ создания веб-приложения с использованием API ключа и раздел процесса написания пользовательского интерфейса.

На четвертом этапе было спроектировано веб-приложение посредством использования более простого способа инициализации API ключа компании OpenAI.

В пятом, заключающем этапе проведено внедрение и использование приложения с ИИ в компании занимающейся продажей пищевых, добавок и витаминов. Тесты проводились на примере реальной компании Kiki-green, занимающейся продажей пищевых добавок на маркетплейсе Amazon. Было решено использовать проектируемое приложения для изменения описания уже существующих продуктов с помощью ИИ. В результате использования проекта было получено крайне положительный результат, который выразился в повышении конверсии и получении около 30 положительных отзывов от реальных покупателей.

Таким образом, данное исследование доказывает перспективность использования нейронных сетей и приложений с ИИ для автоматизации процессов написания текстового описания продуктов, что может привести к улучшению качества продукции и повышению ее конкурентоспособности на рынке.

Литература

1. Анализ покупательского поведения на Amazon // Интерфакс. – 2020. – 3 ноября.
2. Комплексная поддержка разработки и внедрения искусственного интеллекта в бизнесе // Бизнес Online. – 2020. – 15 декабря.
3. Инновационная система на Amazon помогает повышать качество описаний товаров // РИА «Новости». – 2022. – 20 июня.
4. Технологии искусственного интеллекта в коммерческой деятельности: применение на платформе Amazon // Независимое военное обозрение. – 2021. – 5 марта.
5. Кейс-стади: Как Amazon использует машинное обучение для улучшения рекомендаций продуктов // TechCrunch. – 2018. – 12 июля.
6. Технологии глубокого обучения в маркетинге // Коммерсантъ. – 2019. – 20 февраля.
7. Сбор и подготовка данных для машинного обучения // ai-news.ru. – 2019. – https://ai-news.ru/2019/05/sbor-i-podgotovka-dannyh-dlya-mashinnogo-obucheniya/ (был доступен 1 апреля 2023 года).
8. Методы машинного обучения в Python: от простого к сложному // habr.com. – 2021. – https://habr.com/ru/company/ruvds/blog/448310/ (был доступен 1 апреля 2023 года).

9. Как обучить нейронную сеть на Python // geekbrains.ru. – 2021. – https://geekbrains.ru/posts/kak-obuchit-nejronnuyu-set-na-python (был доступен 1 апреля 2023 года).
10. Искусственный интеллект и машинное обучение: возможности и риски // it-rating.ru. – 2020. – https://it-rating.ru/articles/iskusstvennyj-intellekt-i-mashinnoe-obuchenie-vozmozhnosti-i-riski/ (был доступен 1 апреля 2023 года).
11. TensorFlow: от идеи до реализации // dataisfun.ru. – 2020. – https://dataisfun.ru/tensorflow-ot-idei-do-realizatsii/ (был доступен 1 апреля 2023 года).
12. Машинное обучение для написания описаний товаров на Amazon // DataRoot University. – 2022. // https://dataroot-university.com/ru/blog/amazon-product-description-generator-machine-learning (был доступен 1 апреля 2023 года)
13. Как обучить нейронную сеть для генерации описаний товаров на Amazon // Medium. – 2019. // https://medium.com/@ilyazhirkov/how-to-train-neural-network-to-generate-product-descriptions-8a3c400916b6 (был доступен 1 апреля 2023 года).
14. Официальный сайт продукта KIKI Green Sea Moss Powder with Elderberry на Amazon // https://www.amazon.com/KIKI-Green-Powder-Elderberry-Immune/dp/B092CR517C/ref=sr_1_1?keywords=kiki+green&qid=1684855661&refrref=1&sr=8-1 (был доступен 1 апреля 2023 года).
15. Официальный сайт продукта KIKI Green Irish Sea Moss Powder на Amazon // https://www.amazon.com/KIKI-Green-Irish-Moss-Powder/dp/B093WSX3ZD/ref=sr_1_3?keywords=kiki+green&qid=1684855661&refresh=1&sr=8-3 (был доступен 1 апреля 2023 года).
16. Официальный сайт Hellium 10 со статистикой продаж и конверсии // https://members.helium10.com/profits/dashboard?sellerName=A1SM4ZAB0SFO8K&marketplace=ATVPDKIKX0DER&platform=amazon&accountId=1538720021 (был доступен 1 апреля 2023 года).

ОНТОЛОГИЯЛЫҚ ИНЖЕНЕРИНГ НЕГІЗІНДЕ ӘСКЕРІ-ҒАРЫШ ҚЫЗМЕТІ САЛАСЫНДАҒЫ СТАНДАРТТАРДЫ ӘЗІРЛЕУ ЖӘНЕ ҮЙЛЕСТІРУ ҮШІН ЗИЯЛДЫ БАСҚАРУ ЖҮЙЕСІ

[1]Утегенова А.У., [1]Ермолдина Г.Т., [1,2]Бапышев А.М., [1]Науменко В.В.
e-mail: ako-bapyshev@mail.ru
[1]ИНТ-САТ Алатау, Алматы қ., Қазақстан
[2]«Инфракос» РМК Алматы қ. филиалы, Алматы қ., Қазақстан

Аңдатпа. *Барлық салаларды технологиялық жаңғырту дамушы елдер үшін басты басымдық болып табылады. Халықаралық қатынастардың кеңеюі мемлекеттердің белгілі бір салада халықаралық тәжірибені енгізуге дайындығына тікелей байланысты. Осы бағыттардың бірі әскери-ғарыш қызметі саласындағы халықаралық стандарттарды үйлестіру қажеттілігі болып табылады. Мақалада нормативтік құжаттардың мәтіндік ақпаратының үлкен көлемін өңдеу мәселелерін шешуге жаңа көзқарас ұсынылады. Мақалада сипатталған міндеттер зерттеу процедураларының тізбегінің бірегей және ерекше реттілігін білдіреді - зерттеуден, тапсырмаларды қоюдан және ресімдеуден қолданбалы шешімді әзірлеуге дейін. Құжаттардың әдістемелік сараптамасын ішінара автоматтандырудан тұратын Ұлттық стандарттау жоспарына сәйкес стандарттарды үйлестіру үшін процестерді автоматтандыру саласында онтологиялық инженеринг алғаш рет қолданылатын болады. Тәжірибе шеңберінде қолданыстағы нормативтік құжаттардың репозиторийлеріне сілтеме жасамай, тәуелсіз құрал енгізілді. Тәсіл мәтіндік ақпараттан сапалы деректерді алу әдістеріне, интеллектуалды басқару әдістеріне, стандарттауға қатысты жаңа ғылыми және ғылыми-техникалық мәселелерді қоюға, процестерді бақылау және басқару үлгілеріне негізделген.*

Соңғы жылдардағы жаңа технологиялар мен алгоритмдердің қарқынды дамуы көптеген процестерді автоматтандыруға қажетті құралдардың дамуына ықпал етті. Ең алдымен, біз қолданыстағы стандарттар мен нормативтік құжаттарда белгіленген ақпараттың сәйкестігін ескере отырып, ұлттық стандарттарды басқарудың жұмыс процесін оңтайландыратын стандарттарды әзірлеушілер мен сарапшыларға арналған құралдар жинағын құру туралы айтып отырмыз. Екіншіден, әзірлеуге қатысушылардың барлығын және негізгі тұтынушыларды қадағалау қажет. Ал, мемлекеттік тілдегі ұлттық әскери эталондардың терминологиялық бірлігін ұйымдастыруға, сондай-ақ өлшем бірлігін ескере отырып, көмекші құралдарды әзірлеу ең елеулі жұмыс болып табылады.

Бұл зерттеудің мақсаты – бірыңғай репозиторийді құрудың оңтайлы әдістерінің технологиялық интеграциясын пайдалана отырып, стандартты әзірлеудің операциялық циклінің процестерін басқару, бақылау және бақылаудың интеллектуалды автоматтандырылған жүйесін жасау.

Осы мақсатқа жету үшін келесі зерттеу міндеттері орындалды:

- Терминологиялық бірліктегі қайшылықтарды немесе сәйкессіздіктерді анықтау мүмкіндігімен, субъектілер арасындағы байланысты жоғалтпай, үйлестірілген және жаңадан жасалған стандарттар анықтамаларының бірмәнді тұжырымын қалыптастыру үшін онтологиялық инженеринг әдістемесіне сәйкес жүйелік тәсілді пайдалана отырып модельді әзірлеу;

- формальды атрибуттың талаптарына сәйкес келетін құжат ретінде стандарттың біртұтас ұсынылуы тұжырымдамасын әзірлеу кезінде стандартты әзірлеудің операциялық циклінің процестерінде интеллектуалды басқару әдістерін қолдану.

Жұмыста мыналар ұсынылған: стандарттарды жіктеу моделі, нысанды алу модулі, санаттау модулі, тезаурустарды құру модулі, мониторинг және талдау модулі. Есептеулер формализация әдістерін, мақсаттық функцияны белгілеу, ұқсастық өлшемін белгілеу және құжаттың индикативтік сипаттамасын анықтау арқылы жүргізілді. Жұмыста қолдану аймағының теориялық құрамдас бөлігі, дизайн бөлігі және жүйе функционалдығының сипаттамасы көрсетіледі. Алынған нәтижелерді аэроғарыш өнеркәсібі саласындағы стандарттар мен ережелер негізінде осы мәселені шешуге құзыретті сарапшылар тобы сынады.

Түйін сөздер: *ұлттық стандарттар, аэроғарыш қызметі, стандарттарды үйлестіру, кілт сөздерді анықтау алгоритмі, стандарттарды санаттау.*

Кіріспе. Барлық салаларды технологиялық жаңғырту – мемлекеттік маңызы бар басты басымдық. Қазақстанның қорғаныс өнеркәсібін дамыту және басқа елдермен бірлескен кәсіпорындарды дамытуға дайындығы әскери-өнеркәсіптік кешендерді реформалауды жүзеге асыруға мүмкіндік береді. Әскери-өнеркәсіптік кешендердің өсіп келе жатқан қажеттілігіне байланысты осы саланың нормативтік талаптары мен заңнамалық құрамдас бөлігіне тоқталу қажет. Халықаралық тәжірибе саласындағы зерттеулер ұлттық стандарттарды үйлестіруге және әзірлеуге дейін қысқарады. Бірақ халықаралық стандарттардың барлық ашықтығымен Қазақстан аумағындағы қолданыстағы стандарттар талаптарын ескеру қажет. Бұл тәжірибе ғылыми-техникалық ақпараттың берілуіне, әртүрлі елдерде өндірілетін тауарлардың сапа деңгейінің жақындасуына, күрделі өнім элементтерінің өзара алмасуын қамтамасыз етуге және халықаралық қатынастарды дамытуға сұраныстың өсуіне байланысты міндетті болып табылады. Тәжірибе көрсеткендей, өнеркәсіп объектілерін салу және іске қосу заңнамалық құрамдас бөлікті, нормативтік құжаттарды және ұлттық стандарттарды әзірлеумен бірдей дерлік уақытты алады. Осыған байланысты барлық мүдделі тараптардың талаптарын ескере отырып, барлық кезеңдерді мұқият зерделеу қажет. Әскери-өнеркәсіптік кешендердің көпшілігінің екі жақты мақсаты бар, сондықтан стандарттарды әзірлеу кезінде тек әскери стандарттарды ғана емес, сонымен бірге азаматтық стандарттарды да ескеру қажет.

Соңғы жылдардағы жаңа технологиялар мен алгоритмдердің қарқынды дамуы көптеген процестерді автоматтандыруға қажетті құралдардың дамуына ықпал етті. Ең алдымен, біз қолданыстағы стандарттар мен нормативтік құжаттарда белгіленген ақпараттың сәйкестігін ескере отырып, ұлттық стандарттарды басқарудың жұмыс процесін оңтайландыратын стандарттарды әзірлеушілер мен сарапшыларға арналған құралдар жинағын құру туралы айтып отырмыз. Екіншіден, әзірлеуге қатысушылардың барлығын және негізгі тұтынушыларды қадағалау қажет. Ал, мемлекеттік тілдегі ұлттық әскери эталондардың терминологиялық бірлігін ұйымдастыруға, сондай-ақ өлшем бірлігін ескере отырып, көмекші құралдарды әзірлеу ең елеулі жұмыс болып табылады.

Азаматтық-әскери интеграцияның заңнамалық және нормативтік құқықтық актілерін әзірлеу және қайта қарауды басқару жүйелерін талдау

Азаматтық-әскери интеграция ережелері мен стандарттарды басқару жүйелеріне арналған негізгі технологияларды жобалау кезінде ақпараттық жүйелерді пайдалану проблемалары бойынша зерттеулер Хуан, Т еңбектерінде көрсетілген. Автор заңдар мен

азаматтық-әскери интеграция нормативтік құқықтық актілердің жобаларын әзірлеу және қайта қарау процесін басқару мәселелерін ашады. Автордың пікірінше, жүйені пайдалану стандарттау сапасын және азаматтық-әскери интеграция ережелері мен стандарттарын басқарудың уақтылылығын жақсартады, ережелер мен стандарттар туралы деректердің дәлдігі мен қадағалануын арттырады. [1] Х.Э. Рюдің елеулі үлестері мәлім. Авторлық зерттеулер стандарттарды үйлестіру және асимметриялық сәйкестік технологиялары бойынша айтарлықтай күшті және жақсы дәлелдерді ұсынады. Зерттеудің мақсаты – технологиялық асимметриялық елдер арасындағы стандартты үйлестірудің әл-ауқатының салдарын зерттеу және технологиялық прогрестің орташа деңгейі бар елдер үшін оңтайлы үйлестіру стратегияларын анықтау [2]. Чуднова, О.А., Поцулин, А.Д., Шилов, А.С. ұсынған талаптарға қатысты ресейлік стандарттарды жаңартуды талдау халықаралық серіктестермен өз нарықтарын кеңейтуге және бизнеспен айналысуға ниет білдірген отандық кәсіпорындардың әлемдік нарыққа интеграциялануы кезінде стандарттарды үйлестіру мәселелерін еске салды. Зерттеуде ҰЛСТ классификаторы негізінде белгілі бір саладағы ресейлік стандарттарды салыстырмалы талдау нәтижелерін ұсынады, сонымен қатар қазіргі заманғы экономикалық жағдайларда стандарттарды әзірлеу және үйлестіру эволюциясын көрсетеді. Авторлар стандарттарды үйлестіру деңгейін өлшеу критерийін енгізеді [3].

Жоспарланған зерттеулердің маңыздылығын негіздейтін ғылыми-технологиялық қажеттіліктер мәтіндік ақпаратты тану саласындағы Р.Чайсричароен, В.Шримахараж, С.Чейзинг және К.Паманидің негізгі терминдерді анықтауға негізделген зерттеулері сияқты құқықтық категорияларға қатысты салалық стандарттарға жаңа бағыттарды сипаттайды [4]. Ұлттық стандарттарды негізгі салыстыру деректерінің сәйкессіздігін бағалау алгоритмін метрологиялық аттестациялау зерттеушілер тобының Бурмистрова, Н.А., Викторов, И.В., Чуновкина, А.Г зерттеулерінде берілген [5].

Білім беру саласындағы зерттеулерге, оқу бағдарламаларын және оқу нәтижелерін әзірлеу үшін ұлттық стандарттар қолданыла ма, ерекше назар аудару қажет. Ұсынылған әдістер пәндер бойынша әзірленген құжаттаманы сынақтан өткізуге және нәтижелердің талдауын алуға мүмкіндік береді [6-7]. Түйінді сөздерді шығару алгоритмдерін салыстыру әлемдік қауымдастықтың ғылыми қызығушылығын тудырады [8-12].

Бұл зерттеудің мақсаты – бірыңғай репозиторийді құрудың оңтайлы әдістерінің технологиялық интеграциясын пайдалана отырып, стандартты әзірлеудің операциялық циклінің процестерін басқару, бақылау және тексерудің интеллектуалды автоматтандырылған жүйесін жасау.

Осы мақсатқа жету үшін келесі зерттеу міндеттері орындалды:

- Терминологиялық бірліктегі қарама-қайшылықтарды немесе сәйкессіздіктерді анықтау мүмкіндігімен субъектілер арасындағы байланысты жоғалтпай, үйлестірілген және жаңадан жасалған стандарттар анықтамаларының бірмәнді тұжырымын қалыптастыру үшін онтологиялық инженеринг әдістемесіне сәйкес жүйелі тәсілді қолдану арқылы модель әзірленді;
- формальды атрибут талаптарына жауап беретін құжат ретінде стандарттың біртұтас ұсынылуы тұжырымдамасын әзірлеу кезінде стандартты әзірлеудің операциялық циклінің процестерінде интеллектуалды бақылау әдістері қолданылды.

Материалдар мен тәсілдер. Жобаның негізгі идеясы – ұлттық стандарттарды әзірлеу кезінде процестерді бақылау және бақылаудың бірыңғай жүйесін, жетілдірілген механизмін әзірлеу. Жоба зерттеу процедураларының тізбегінің бірегей және ерекше тізбегін қамтиды - зерттеуден, тапсырмаларды қоюдан және рәсімдеуден қолданбалы шешімді әзірлеуге дейін. ҚазИнСт-ке тапсыру үшін құжаттарды әдістемелік сараптауды

ішінара автоматтандырудан тұратын Ұлттық стандарттау жоспарына сәйкес стандарттарды үйлестіру үшін процестерді автоматтандыру саласында алғаш рет онтологиялық инженерингісі қолданылатын болады. Осы жобада ұсынылған шешімдер әлі іске асырылған жоқ және Қазақстанда баламасы жоқ, міндеттерді іске асыру кезінде әскери стандарттау саласындағы ағымдағы өзгерістерге серпінді жауап бере отырып, ұлттық стандарттарды әзірлеушілерді қолдау мақсаттарына қол жеткізілді.

Тапсырмаларды орындау кезінде онтологиялық инженеринг метадеректерді, цифрлық ізді, мәтінді автоматтандыру әдістерін пайдалану үлгілері мен әдістерін салыстырмалы талдау нәтижелері бойынша Ұлттық стандарттау жоспарына сәйкес стандарттарды үйлестіру үшін стандартты әзірлеудің операциялық циклінің процестерінде белгілеу және интеллектуалды басқару әдістерінің процестерді автоматтандыру саласында қолданылады.

Стандарттардың онтологиялық көрінісі стандарттарды сала бойынша жіктеу міндетін жүзеге асырады, яғни әрбір құжат үшін оның белгілі бір салаға жату дәрежесі анықталады. Категориялау кезінде екі деңгей енгізіледі: бірінші деңгейде стандарттың классификациялық топқа жататындығы, одан кейін екінші деңгейде стандарттың мазмұны жағынан жақын стандарттарға жататындығы анықталады.

Әрбір құжат үшін құжаттың санатқа жататындық дәрежесі анықталады, ол сәйкес келетін түйінді сөздердің санымен қалыптасады (мүшелік дәрежесі неғұрлым жоғары болса, құжат соғұрлым осы санатқа сәйкес келеді).

Жіктеу жіктеу тобының ішінде бірнеше топтар үшін бір стандартты анықтау мүмкіндігін ескереді.

Құжат ұқсастығының формальды өлшемі құжат кілт сөздерінің іздеу үлгілерін салыстыру арқылы ұйымдастырылады. Құжаттың түйінді сөздерінің бір кездері қалыптасқан бейнесі, сондай-ақ мүмкіндіктер кеңістігінің масштабы қолданыстағы стандарттарды қолданудың барлық мерзімінде өзгермейді.

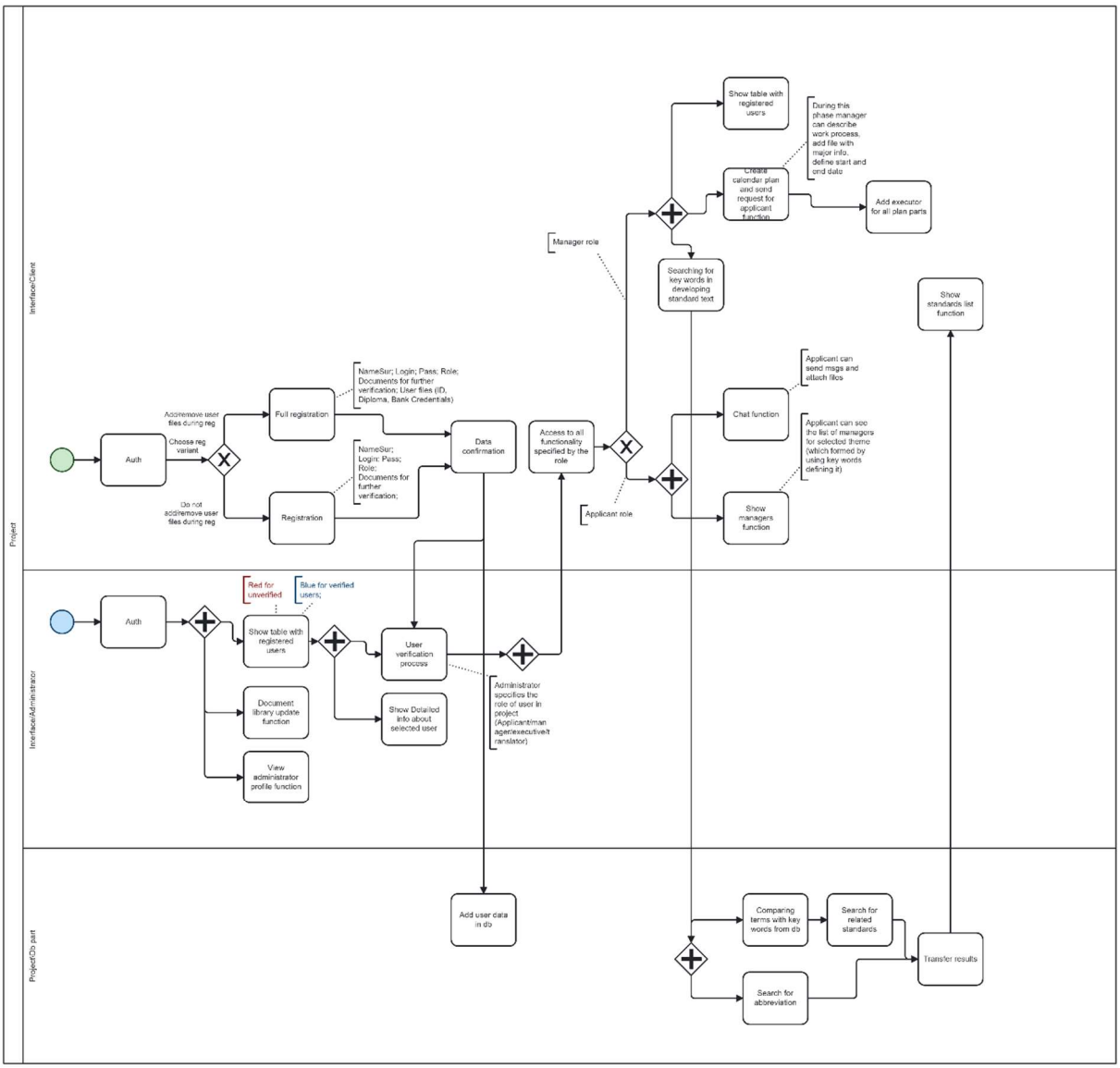
Мазмұны жағынан жақын стандарттарға жататын стандарт тізбесін қалыптастыру кезінде құжаттарды іріктеу басымдылық бойынша (құжаттардың мазмұны неғұрлым жақын болса, оның басымдылығы төменірек) реттелген. Қаралатын құжаттардың санына ешқандай шектеулер жоқ.

Ұқсас жіктеу әдісі пайдаланушылар үшін сала бойынша санаттау үшін қолданылады. Сарапшылар тобының жетекшілері стандарттарды үйлестіруге арналған өтінімдерді сарапшының түйінді сөздері мен жарияланымдары туралы ақпарат беретін бейіні бойынша сәйкес келетіндерін ғана көреді.

Құжат классификацияларының соңында стандарттардың ұйымдасқан классификациялық тобы үшін түйінді сөздердің тезаурусы қалыптасады.

Жұмыста мыналар ұсынылған: стандарттарды жіктеу моделі, нысанды алу модулі, санаттау модулі, тезаурустарды құру модулі, мониторинг және талдау модулі. Есептеулер формализация әдістерін, мақсаттық функцияны белгілеу, ұқсастық өлшемін белгілеу және құжаттың индикативтік сипаттамасын анықтау арқылы жүргізілді. Жұмыста қолдану аймағының теориялық құрамдас бөлігі, дизайн бөлігі және жүйе функционалдығының сипаттамасы көрсетіледі. Алынған нәтижелерді аэроғарыш өнеркәсібіндегі тәжірибе негізінде осы мәселені шешуде құзыретті сарапшылар тобы сынады.

1-суретте BPMN-сызба процесінің диаграммасы көрсетілген. Бизнес-процестерді модельдеу нақты сипаттама мен техникалық функционалдылық арасындағы ымыраға келуге мүмкіндік береді.

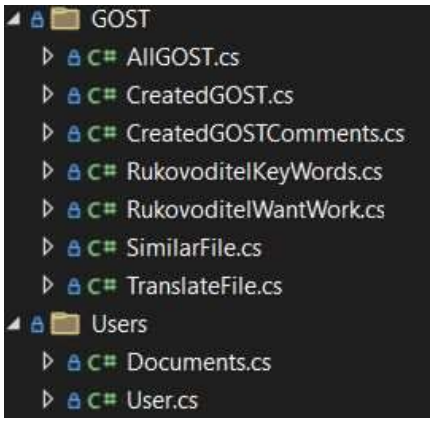


Сурет 1- BPMN-сызба процесі

Нәтижелер мен талқылаулар. Әрбір зерттеу шеңберінде құжатты талдау құжаттың өзінің нысаны мен мазмұнына, зерттеудің мақсаты мен шарттарына, зерттеушінің тәжірибесінің байлығына және шығармашылық интуициясына байланысты дербес шығармашылық процесс болып табылады. Егер бір-бірінен тәуелсіз алынған деректердің екі қатары сәйкес келетіні анықталса, бұл ақпарат жеткілікті сенімді және тұрақты екенін білдіреді.

Файлдың программалық кодпен әрекеттесуі (2-сурет).

1. Файл үлгіге таратылады;
2. Әрі қарай үлгілерден алынған код қажетті ақпаратты алады;
3. Өңдеуден кейін оны серверден интерфейсте клиентке көрсетеді.



Сурет 2 – Мәліметтер базасының кестелері

Модель деректер қорындағы кестелерді сипаттайды (2-сурет)

- Allgost әкімші мен өтініш беруші жүктеген файлдардың тізімін жүргізеді;
- Created Gost стандарттарды құруға арналған сұраныстар тізімін сақтайды, мұнда барлық әрекет орындау кезінде орын алады;
- Стандартты құру кезінде менеджерлердің, орындаушылардың, аудармашылардың, әкімшілер мен өтініш берушінің Created GOST Comments хабарламалары;
- RukovoditelKeyWords басшыға арналған кілт сөздерді сақтайды, олар өтініш берушіге ұсыну үшін таңдалады;
- RukovoditelWantWork күнтізбелік жоспары, мұнда менеджер орындаушыларға тапсырма береді және жұмысқа өтінім қалдырады;
- SimilarFile ұқсас терминдердің ең көп саны бар файлдарды сақтайды;
- TranslateFile күнтізбелік жоспарға ұқсас, бірақ аудармашылар үшін орындаушылар жұмысты аяқтағаннан кейін беріледі;
- Құжаттарда дипломдар, сертификаттар және т.б. оны тексеру үшін жаңа пайдаланушы;
- Пайдаланушылардың пайдаланушы тізімі.

Әрбір зерттеу шеңберінде құжатты талдау құжаттың өзінің нысаны мен мазмұнына, зерттеудің мақсаты мен шарттарына, зерттеушінің тәжірибесінің байлығына және шығармашылық интуициясына байланысты дербес шығармашылық процесс болып табылады. Егер бір-бірінен тәуелсіз алынған деректердің екі қатары сәйкес келетіні анықталса, бұл ақпарат жеткілікті сенімді және тұрақты екенін білдіреді.

Пайдаланушы аутентификациясы бар негізгі кіру терезесі тек тексерілген пайдаланушыларға жоба ресурстарын пайдалануға рұқсат береді.

Тіркеу терезесі толық аты-жөні, логин, пароль, жобаны орындаушының рөлі, кейінгі тексеру үшін құжаттар сияқты пайдаланушы деректерін енгізуге мүмкіндік береді. Пайдаланушы файлдарын (сертификаттар, дипломдар, банктік деректемелер және т.б.) қосу және оларды жою функциясы да қол жетімді. Тіркеу кезінде бұл қадамды өткізіп жіберуге және тіркеуден кейін жасауға болады.

Тіркелу кезінде жүйенің пайдаланушысы өзі енгізген деректерді қарайды. Сондай-ақ бұл терезеде пайдаланушының күйі, тексеру туралы ақпаратты көруге болады.

Пайдаланушылардың ортамен жұмысы әкімшінің негізгі беті арқылы ұйымдастырылады, ол тіркелген пайдаланушылар туралы кестені қамтиды, пайдаланушы туралы

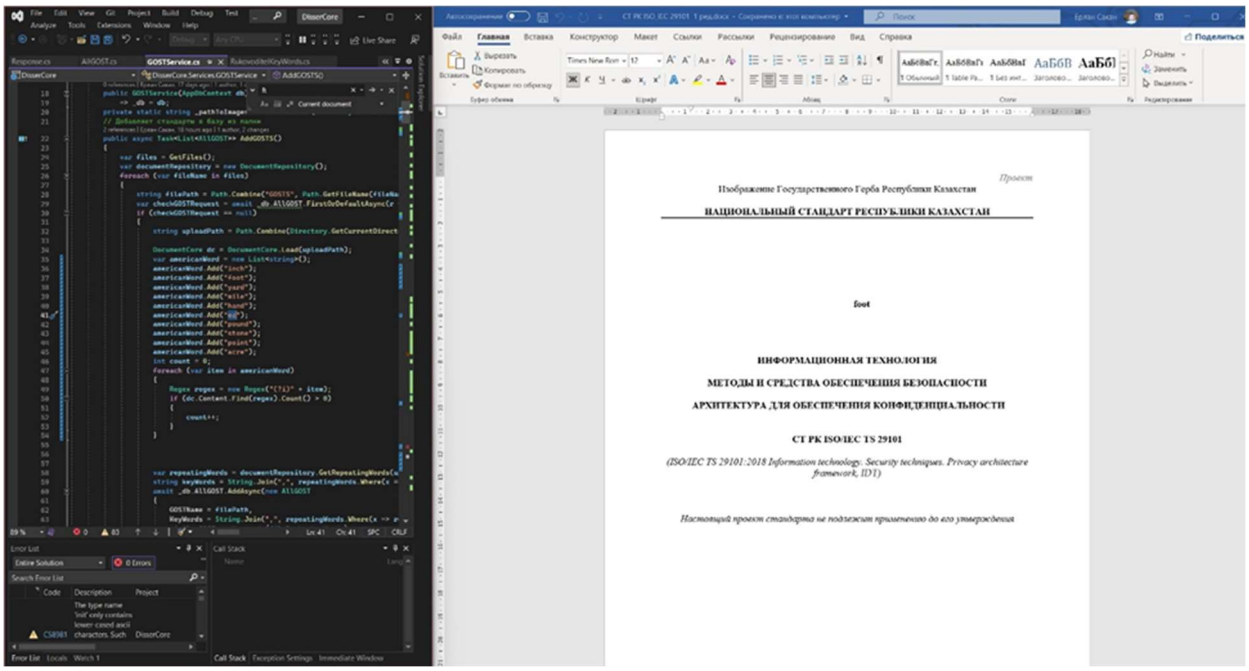
мәліметтер түймесі бар. Жоғарғы мәзір бес қойындыны көрсетеді: Профиль, Пайдаланушы тізімі, Стандарттар тізімі, Стандарттарды әзірлеушілер тізімі. Кітапхананы толықтыру функциясы бар. Пайдаланушыны растау ақпараты жасыл түспен, тексерілмеген жүйе пайдаланушылары қызыл түспен белгіленеді.

Тексерілмеген пайдаланушылардың аттары қызыл түспен бөлектелген. Тексеру кезінде әкімші жобадағы пайдаланушының рөлін көрсетеді. Қалқымалы мәзірде төрт рөл бар: өтініш беруші, жетекші, орындаушы, аудармашы. Тексерудің соңында пайдаланушының толық аты-жөнінің түсі көк түске өзгереді. Әкімші менеджерді тағайындайды және пайдаланушыны тексереді.

Менеджер профилінде тіркелген пайдаланушылар кесте түрінде көрсетіледі, пайдаланушы туралы ақпаратты нақтылау үшін егжей-тегжейлі түйме бар.

Халықаралық стандарттарды үйлестіру кезінде өлшем бірліктерін егжей-тегжейлі зерделеу және дұрыс түрлендіру қажет. Жүйе қолданыстағы өлшем бірліктерінің үйлестірілген эталонының мәтінінде автоматты іздеуді қамтамасыз етеді, әзірлеушілердің назарын аудару және метрикалық жүйелерді, натурал бірліктер жүйелерін аударуды нақтылау үшін кітапхананы жаңа өлшем бірліктерімен толықтыру функциясы қарастырылған. өлшеу және дәстүрлі өлшем жүйелері.

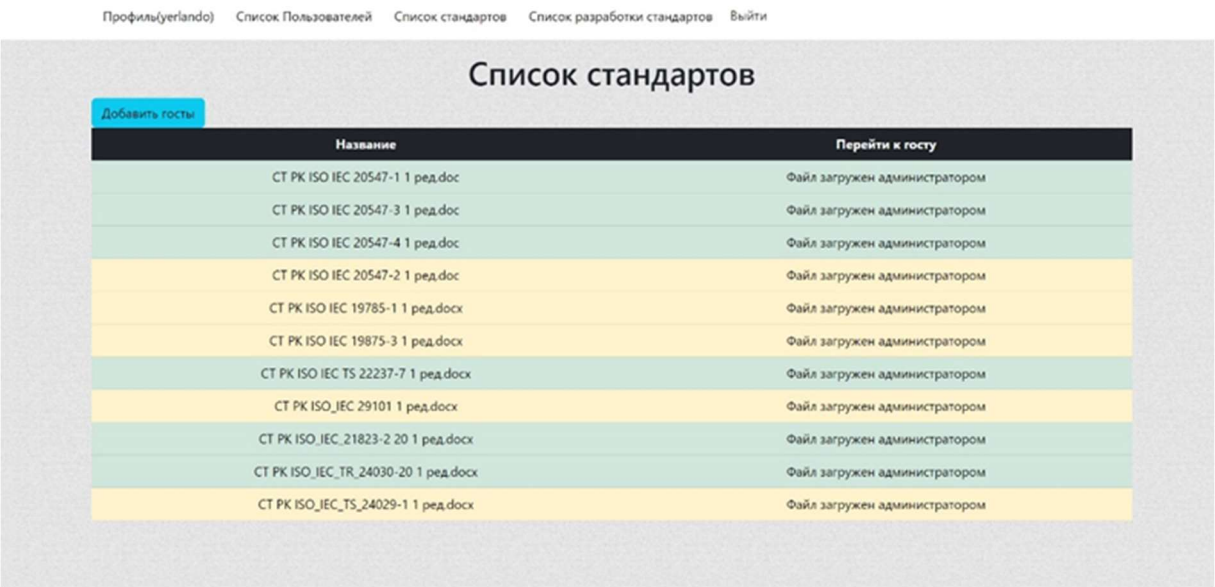
Үйлестірілген стандарттармен 3 түрі анықталған, бірдей (ID), модификацияланған (MOD) және баламалы емес стандарттар (NEQ). Ұсынылған жүйе оны үйлестірудің барлық түрлеріне қолдануға мүмкіндік береді. Құрылыс стандарттарының ережелерін және олардың мазмұнына қойылатын жалпы талаптарды ұйымдастыру модуліне ерекше назар аудару керек. Модуль стандарттардың құрылысына, көрсетілуіне, дизайнына және мазмұнына қойылатын жалпы талаптарға жауап беретін жартылай автоматты режимде шаблон жасауға мүмкіндік береді (3-сурет).



Сурет 3 - Автоматты пішімдеу

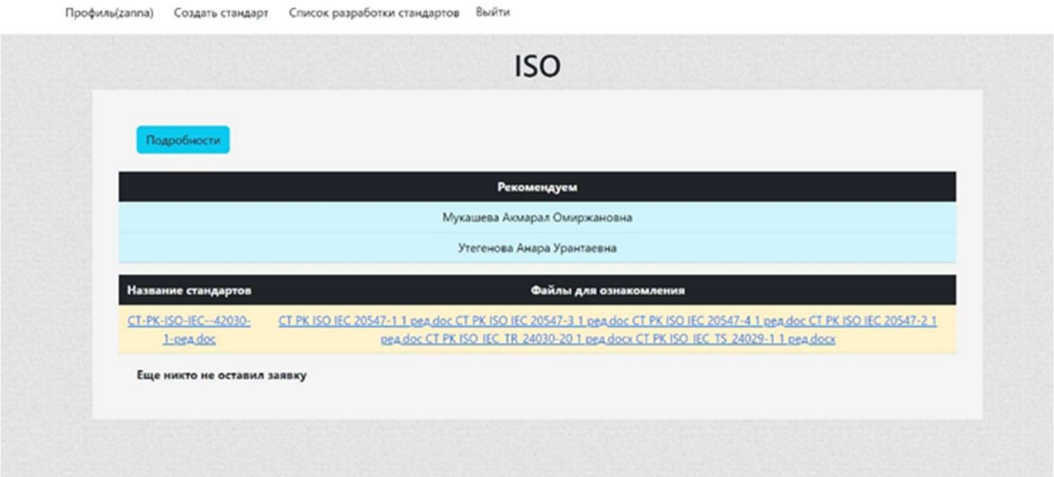
Мысал ретінде жүйе қолданатын негізгі стандарттар Қазақстан стандарттау және метрология институтының https://www.ksm.kz/public-discussion/standards/discussion-of-

draft-of-st-rk сайтынан алынды. Бұл ресурста талқылауға арналған стандарттар жобалары бар, бұл репозиторийдің стандарттар базасын қалыптастыруға мүмкіндік берді (4-сурет).



Сурет 4 – Бірліктерді табу

Өңдеу нәтижелері бойынша менеджерлердің шолу үшін стандартты қарау мүмкіндігі бар. Бұл кезеңде талапкер белгілі бір басшының іс-әрекетінің бағытын анықтайтын түйінді сөздер арқылы қалыптасатын ұсынылған тақырып бойынша ағымдағы көшбасшылармен таныса алады. Негізгі модуль әзірленген және мазмұнды сандық талдаудан, түйінді сөздерді қалыптастырудан тұрады. Негізгі құжаттан және оларды басқа құжаттармен салыстыра отырып, құжат мәтініндегі бөлінген негізгі терминдер базаны құрайды және жүйеде сақталады. Бұл тізім тек осы саланың сарапшысы болып табылатын басшыларға ғана көрінеді, менеджерлердің деректер базасында салалар бойынша сарапшыларды таңдауға арналған түйінді сөздер бар (5-сурет).



Сурет 5 – Стандарт беті

```
            var repeatingWords = documentRepository.GetRepeatingWords(uploadpath, 3);
            string keyWords = String.Join(",", repeatingWords.Where(x => repeatingWords.Values.OrderByDescending(i => i).Take(50).Contains(x.Value)).Select(p => p.Key));
            AddinAllGOSTS.GOSTName = filepath;
            AddinAllGOSTS.KeyWords = keyWords;
            AddinAllGOSTS.similarGOSTs = Keywords(filepath, keyWords);
            AddinAllGOSTS.IsAmerican = count > 0 ? true : false;
        }
    }
    return AddinAllGOSTS;
```

Сурет 6 – Түйінді сөз алгоритмі

Қорытынды. Стандарттау саласында онтологияны қолдану келесі артықшылықтарды береді: терминдердің қайталануын жою, терминдердегі қайшылықтарды жою, бұрыннан енгізілген стандарттарды қайталау, ұйымдар немесе басқа пәндік салалар үшін стандарттарды масштабтау, бар онтологиялардан жасырын білімдерді ашу.

Ұсынылған модульдерді тестілеу кезінде мәтінді қолмен өңдеу нәтижелерімен сәйкестіктердің 93% алынды. Сәйкестендіріліп жатқан құжатқа қатысы жоқ құжаттардың көпшілігі тексеріліп, сарапшыларға зерттеуге ұсынылған құжаттар тізімі ұсынылды. Осылайша, бір орташа өлшемді құжатты өңдеуге 1500 миллисекунд уақыт жұмсалды. Сонымен бірге алынған онтологиялар есептеліп, қайта пайдалану критерийлерін ескере отырып әзірленеді. Алынған көрсеткіштерді ескере отырып, қолданылатын құралды әртүрлі салалар мен бизнес модельдер үшін пайдалануға болады деп болжауға болады. Жоспарларға мемлекеттік жүйелерге арналған ұсыныстарды ескере отырып, веб-қосымша интерфейсін әзірлеу кіреді.

Ризашылық. Бұл зерттеуді Қазақстан Республикасы Білім және ғылым министрлігінің Ғылым комитеті (грант №AP09058441 «Мемлекетаралық және ұлттық стандарттау негізінде онтологиялық инженерия бойынша стандарттарды әзірлеу мен үйлестіруді басқарудың интеллектуалды жүйесін құру») жүргізілді және қаржыландырылды.

Пайдаланылған әдебиеттер тізімі

1. Huang, T.Design and Key Technology Research of Civil-Military Integration Regulations and Standards Management System(2021) Smart Innovation, Systems and Technologies, . DOI: 10.1007/978-981-15-5073-7_21
2. Ryu, H. E. (2021). Standards Harmonization and Asymmetric Compliance Technology. Journal of Korea Trade, 25(3), 1-20. doi:10.35611/jkt.2021.25.3.1
3. Chudnova, O. A., Potsulin, A. D., & Shilov, A. S. (2018, Oct 02-04). Analysis of Updating of the Russian Standards Concerning Requirements Harmonization. Paper presented at the International Scientific Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), Vladivostok, RUSSIA doi:10.2991/iscfec-18.2019.268
4. R. Chaisricharoen, W. Srimaharaj, S. Chaising and K. Pamanee, «Classification Approach for Industry Standards Categorization» 2022 Joint International Conference, on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), 2022, pp. 308-313, doi: 10.1109/ECTIDAMTNCON53731.2022.9720348.
5. Burmistrova, N.A., Viktorov, I.V., Chunovkina, A.G., Metrological Certification of the Algorithm for Estimating Inconsistent Data of Key Comparisons of National Standards, (2023) Measurement Techniques, Volume 65, Issue 7, October 2022, Pages 508-514, DOI: 10.1007/s11018-023-02111-1

6. Tiffanie Ford-Baxter, Kendall Faulkner, Jennifer Masunaga, Situating Information Literacy: A Case Study Exploring Faculty Knowledge of National Disciplinary Standards and Local Program Learning Outcomes, The Journal of Academic Librarianship, Volume 48, Issue 3, 2022, ISSN 0099-1333, https://doi.org/10.1016/j.acalib.2022.102523.
7. Barthelemy, F., Cochez, M., Dimitriadis, I., Karim, N., Loutas, N., Magnisalis, I., Comet, L. M., Peristeras, V., & Wyns, B. (2022). Towards a standard-based open data ecosystem: analysis of DCAT-AP use at national and European level. *Electronic Government*, *18*(2), 137-180. https://doi.org/10.1504/EG.2022.121856
8. Eliseo Reategui, Marcio Bigolin, Michel Carniato, and Rafael Antunes dos Santos. 2022. Evaluating the Performance of SOBEK Text Mining Keyword Extraction Algorithm. In Machine Learning and Knowledge Extraction: 6th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2022, Vienna, Austria, August 23–26, 2022, Proceedings. Springer-Verlag, Berlin, Heidelberg, 233–243. https://doi.org/10.1007/978-3-031-14463-9_15
9. Mehrabi, Elaheh & Mohebi, Azadeh & Ahmadi, Abbas. (2020). Improved Keyword Extraction for Persian Academic Texts Using RAKE Algorithm, Case Study: Persian Thesis and Dissertations. Iranian Journal of Information Processing Management. doi: 37. 10.52547/jipm.37.1.197.
10. M. S. U. Miah, J. Sulaiman, S. Azad, K. Z. Zamli and R. Jose, "Comparison of document similarity algorithms in extracting document keywords from an academic paper," 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM), Pekan, Malaysia, 2021, pp. 631-636, doi: 10.1109/ICSECS52883.2021.00121.
11. Kungurtsev O. et al. Development of information technology of term extraction from documents in natural language //Eastern-European Journal of Enterprise Technologies. – 2018. – Т. 6. – №. 2. – С. 44-51.
12. Lytvyn V. et al. Development of a method for determining the keywords in the slavic language texts based on the technology of web mining //Восточно-Европейский журнал передовых технологий. – 2017. – №. 2 (2). – С. 14-23.

СЕКЦИЯ 4

Ақпараттық қауіпсіздік және деректерді қорғау: Ақпаратты қорғауға арналған бағдарламалық және техникалық құралдар. Криптография. Күрделі жүйелердің ақпараттық қауіпсіздігін қамтамасыз етудің математикалық әдістері.

Информационная безопасность и защита данных: Программно-технические средства защиты информации. Криптография. Математические методы обеспечения информационной безопасности сложных систем.

Information security and data protection: Software and hardware means of information protection. Cryptography. Mathematical methods of ensuring information security of complex systems.

АҚПАРАТТЫҚ ҚАУІПСІЗДІК ҮШІН ЖАСАНДЫ ИНТЕЛЛЕКТ ӘДІСІН ҚОЛДАНУ

Тынымбаев Ә. Б., Байшоланова К. С., Кубаев К. Е.
E-mail: abilkaiyr.tynymbayev@kaznu.edu.kz
Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан

Аңдатпа: *Жасанды интеллект – қазіргі заманауи және жылдам интеллектуалды жүйелерді цифрлық ортада дамытумен айналысатын жаңа салалардың бірі. Жасанды интеллектінің модельдері мен әдістері - интеллект жүйелер арқылы жаңа білімді ұсынуды зерттейді.*

Мақалада ақпараттық жүйелердің қауіпсіздігін қамтамасыз ету үшін тиімді жасанды интеллект әдісін құруға бағытталған зерттеулер қарастырылады. Осы зерттеудің бір бөлігі ретінде машиналық оқыту, деректерді талдау және аномалияны анықтау сияқты әртүрлі жасанды интеллект алгоритмдері зерттеліп, ең тиімді әдіс ретінде шешім ағаштары әдісі таңдалады. Бұл зерттеудің нәтижесі ақпаратты неғұрлым тиімді қорғауды қамтамасыз ету үшін ақпараттық қауіпсіздік жүйелерінде қолданылуы мүмкін жаңа әдісті ұсыну болып табылады.

Түйін сөздер: *ақпараттық қауіпсіздік, машиналық оқыту, жасанды интеллект, шешім ағаштары, Шенон энтропия, киберқауіпсіздік.*

Кіріспе. Технологияға тәуелділіктің өсуі тиімді ақпараттық қауіпсіздік жүйелеріне қажеттіліктің артуына әкелді. Көптеген құпия ақпарат электронды түрде сақталатындықтан және тасымалданатындықтан, қазіргі кезде қауіпсіздікті бұзу және кибершабуылдар ықтималдығы айтарлықтай өсті. Бұл мәселені шешу үшін жасанды интеллект қолданатын әдістерді әзірлеуболашағы зор, өзекті шешімдердің біріне айналды.

Жасанды интеллектті ақпаратты қауіпсіздіктің әртүрлі аспектілеріне қолдануға болады, мысалы оларға: шабуылды анықтау, қауіпті болжау және алдын алу мен деректерді талдау жатады. Оның алгоритмдерін пайдалану дәстүрлі қауіпсіздік әдістеріне қарағанда тиімдірек және қауіпсіздік қатерлерін де нақты анықтауға және азайтуға көмектеседі. Сонымен қатар, жасанды интеллект алгоритмдері бұрынғы қауіпсіздік оқиғаларынан сабақ алып, уақыт өте келе олардың жұмысын жақсарта алады.

Осыған байланысты бұл зерттеудің мақсаты - ақпараттық қауіпсіздік жүйелерінде қолдануға арналған жасанды интеллект әдісін әзірлеу болып табылады. Зерттеу барысында ақпараттық қауіпсіздік саласында қолданылуы мүмкін жасанды интеллекттің әртүрлі алгоритмдері мен әдістерін талданып, олардың тиімділігі мен шектеулері бағаланды. Зерттеу нәтижесі – ұйымдар өздерінің ақпараттық қауіпсіздігін жақсарту үшін қабылдай алатын кешенді және практикалық шешімді ұсыну.

Зерттеудің методологиялық базасын отандық және шет елдік ғалымдардың, атап айтқанда Х. Мехрпуян, Н. Каправелос, Дж. Маркофф, және т.б. осы тақырыпқа сәйкес еңбектері құрайды [1-6].

Зерттеуде қолданылған әдістер:

– теориялық әдістерден теориялық талдау және синтездеу, эмпирикалық әдіс, бағдарламалық модельдеу, жүйелік талдау және зерделеу, сондай-ақ жасанды интеллект әдістерін талдау, логикалық қорытынды әдістері қолданылды.

Зерттеудің ғылыми жаңалығы – Ақпараттық қауіпсіздік жүйелерінде жасанды интеллект әдісі зерттеліп, сол ортадағы тиімді алгоритмдер таңдалды.

Бұл алгоритмдер машиналық оқыту, терең оқыту және компьютерлік көру сияқты жасанды интеллект әдістерінің ауқымын көрсетеді.

Жасанды интеллект алгоритмдерін қолданысқа енгізуде таңдалған алгоритмдер тиісті деректер жиынын пайдалану арқылы зертханалық ортада жүзеге асырылады. Бұл алгоритмдердің өнімділігін дәлдік, негізділік, еске түсіру және F1 ұпайлары бойынша бағалауға болады. Нақты айтқанда, өнімділікті бағалау: жасанды интеллект алгоритмдерінің өнімділігі жалған оң жылдамдық, жалған теріс жылдамдық және жалпы дәлдік сияқты тиісті көрсеткіштер арқылы бағаланады. Бұл көрсеткіштер қауіпсіздік қатерлерін анықтау және жоюдағы алгоритмдердің тиімділігін жан-жақты бағалауды қамтамасыз етеді. Зерттеу алгоритмдерін жан-жақты бағалауды қамтамасыз ету үшін сандық және сапалық әдістердің комбинациясын пайдалануға болады. Зертханалық орта бағдарламалық және аппараттық ресурстарды біріктіру арқылы құрылатындықтан, сәйкес қауіпсіздік деректер жиыны жалпыға ортақ деректер жиынынан және модельденген қауіпсіздік сценарийлерінен алынады.

Сонымен жасанды интеллект әдісін әзірлеу - өнімділікті бағалау нәтижелері бойынша біріктірілген әдіс ретінде әзірленеді. Бұл әдіс жасанды интеллект негізіндегі ақпараттық қауіпсіздікті қамтамасыз етудегі озық тәжірибелерді, сондай-ақ бағаланатын әртүрлі алгоритмдердің күшті және әлсіз жақтарын ескереді. Ақпараттық қауіпсіздік жүйелері үшін жасанды интеллект әдістерін әзірлеу кезінде келесі (1-сурет) алгоритмдерді қолдануға болады [7]:

1-сурет. Алгоритмдер құрылымы

Осылардың ішінен машиналық оқыту алгоритмдеріннің шешім ағаштары модельін таңдауға болады. Себебі ол есептердің нақты шешімдерін иерархиялық сипатта алуға мүмкіндік береді. Ол индукция, яғни иерархиялық шешімдер ағашын тұрғызу және қию үрдістерінен тұрады. Соңғы үрдісте машинаға қайта оқытылатын шешімдерді ықшамдайды.

Индукция үрдісінде жақсы белгілер іріктеліп, олар ағаш түйіндерін білдіреді.

Нақты баяндағанда, Python тіліндегі scikit-learn кітапханасын қолданатын шешім ағаштары былай келтіріледі:

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import tree
from sklearn.metrics import accuracy_score

# Деректер жинағын жүктеп алыңыз
data = pd.read_csv("/content/sample_data/incidents.csv")

# Деректер жиынтығын нысандар мен мақсатты айнымалыларға бөліңіз
X = data.iloc[:, :-1]
y = data.iloc[:, -1]

# Деректерді оқу және тест жиынтықтарына бөліңіз
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

# Шешім ағашының үлгісін жасаңыз
model = tree.DecisionTreeClassifier()

# Оқу деректері негізінде модельді оқытыңыз
model.fit(X_train, y_train)

# Сынақ деректері негізінде болжам жасаңыз
y_pred = model.predict(X_test)

# Модельдің дәлдігін бағалаңыз
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy: ", accuracy)
```

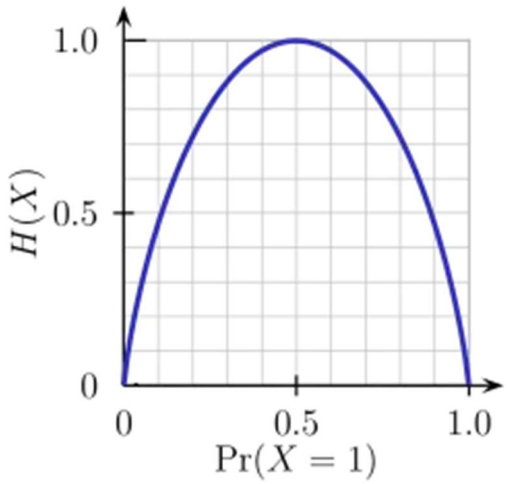
Код деректерді оқыту және тестілеу жиынтықтарына бөледі, шешім ағашының моделін жасайды, модельді оқыту деректеріне үйретеді, сынақ деректеріне болжам жасайды және модельдің дәлдігін бағалайды.

Ақпараттық қауіпсіздіктегі шешім ағашы алгоритмі шарттар жиынтығына немесе кірістерге негізделген шешім қабылдау үшін ағаш үлгісін пайдаланады. Шешім ағашын құру үшін қолданылатын S Шеннон энтропиясын келесідей (1) өрнегімен көрсетуге болады [8,9]:

$$H(S) = \sum_{x \in X} p(x) log_2 \frac{1}{p(x)} \quad (1)$$

мұндағы S – ағымдағы жағдай жиынтығы.

Бұл байланысты былай көрсетуге де болады:



2-сурет. Ықтималдығы екі шешімге қатысты энтропия мысалы

Энтропия деректердің тазалығы немесе кездейсоқтығының өлшемі болып табылады. Шешім ағашы алгоритмінің мақсаты әрбір функция үшін Gain (S, x) мәнін азайту болып табылады, осылайша алынған ағаш үлгілерді сынып белгілеріне қарай тиімді бөле алады. Ең үлкен күшейту мәні бар объект ағаштың тамыры ретінде таңдалады және процесс ағаштың максималды тереңдігі немесе түйіндегі үлгілердің ең аз саны сияқты тоқтату шарты орындалғанға дейін әрбір бұтақ үшін қайталанады.

Шешім ағашы алгоритмі – бұл ақпараттық қауіпсіздік саласында, әсіресе интрузияларды анықтау, спамды сүзу және басқа жіктеу тапсырмаларында машиналық оқытудың кеңінен қолданылатын әдісі. Алгоритм ең маңызды қасиетке негізделген кірістерді кішігірім топтарға бөлу арқылы жұмыс істейді, содан кейін шешімдер мен олардың нәтижелерінің ағаш құрылымын жасайды. Ондағы мәліметтер берілген параметрлерге сәйкес бөлініп отырады да, ағаш бұрақтары қалыптаса бастайды. Әрбір түйіндерде шешім қабылданып, мәліметтер әрі қарай тағы тармақталады. Соңғы түйінде нәтиже алынады. Түпкі мақсат – ағашты айналып өту арқылы көрінбейтін үлгілер класының белгісін дәл болжау.

Мұнда шешім ағашының алгоритмі ақпараттық қауіпсіздікте қалай жұмыс істейтінін көрсететін мысал қарастырылады. Дереккөздің IP мекенжайы, алушының IP мекенжайы және порт нөмірі негізінде желі қосылымының қалыпты немесе қалыпты емес екенін болжағымыз келеді делік. Шешім ағашын құру үшін алдымен деректерді ең біртекті топтарға бөлетін ең маңызды функцияны анықтаймыз. Бұл жағдайда, айталық, ең маңызды белгі - бастапқы IP мекенжайы.

Бастапқы IP мекенжайының мәндеріне сүйене отырып, біз әр тармақ әртүрлі нәтижелерге әкелетін шешім тармақтарын жасаймыз. Мысалы, егер бастапқы IP мекенжайы сенімді жерде болса, біз байланыс қалыпты деп есептейміз. Екінші жағынан, егер бастапқы IP мекен-жайы сенімсіз жерде болса, біз байланыс қалыпты емес деп есептейміз.

Содан кейін біз бұл процесті тағайындалған IP мекенжайы және порт нөмірі сияқты қалған функциялар үшін алгоритм ағаштың максималды тереңдігі немесе соңғы түйінге

үлгілердің ең аз саны сияқты тоқтату критерийіне жеткенше қайталаймыз. Соңында бізде кіріс белгілеріне негізделген шешімдер мен олардың нәтижелерін көрсететін ағаш құрылымы болады.

Болжау шешімдерінің ағашын пайдалану үшін біз кіріс белгілерінің мәндеріне негізделген ағашты аралап, тиісті нәтижеге қол жеткіземіз. Мысалы, егер көздің IP-мекен-жайы сенімді жерде болса, тағайындалған IP-мекен-жайы сенімді жерде болса және порт нөмірі әдетте зиянды әрекеттер үшін пайдаланылмаса, алгоритм қосылым қалыпты деп болжайды:

```
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
import matplotlib.pyplot as plt
from sklearn import tree

# үлгі деректер
data = {'Source IP': ['Trusted', 'Unknown', 'Unknown', 'Trusted', 'Trusted'],
        'Destination IP': ['Trusted', 'Trusted', 'Unknown', 'Unknown', 'Trusted'],
        'Port': [80, 8080, 22, 443, 21],
        'Connection': ['Normal', 'Abnormal', 'Normal', 'Abnormal', 'Normal']}

# деректер жақтауын жасаңыз
df = pd.DataFrame(data)

# белгілер матрицасы
X = df.iloc[:, :-1]

# мақсатты вектор
y = df.iloc[:, -1]

# категориялық айнымалыларды бір ыстық кодтау
X = pd.get_dummies(X)

# деректерді оқу және тест жиынтықтарына бөлу
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

# шешім ағашының үлгісін жасаңыз
model = DecisionTreeClassifier()
model.fit(X_train, y_train)

# сынақ деректері негізінде болжау
y_pred = model.predict(X_test)
# дәлдікті бағалаңыз
print("Accuracy:", accuracy_score(y_test, y_pred))
```

Кодтың нәтижесі:

Accuracy: 1.0

Бұл код sklearn кітапханасынан Decision Tree Classifier пайдаланып, шешім ағашының үлгісін құрылады. Содан кейін модель оқыту деректеріне бейімделеді және сынақ деректерін болжау үшін қолданылады. Әрі қарай, бағдарламада модельдің дәлдігі бағаланады және басып шығарылады.

Пікірталас. Ақпараттық қауіпсіздік жүйелерінде жасанды интеллектті пайдалану жылдам дамып келе жатқан сала болып табылады, өйткені қауіпсіздікті қамтамасыз ету үшін неғұрлым жетілдірілген және күрделі шешімдерге сұраныс артып келеді. Бұл зерттеуде біз ақпараттық қауіпсіздік жүйелерінде тиімді қолдануға болатын жасанды интеллект әдісін әзірлеуге тырыстық.

Біздің зерттеу нәтижелеріміз машиналық оқыту және терең оқыту сияқты әртүрлі жасанды интеллект алгоритмдерін ақпараттық қауіпсіздік жүйелерінде ықтимал болатын қауіпсіздік қауіптерін анықтау және олардың алдын алу үшін тиімді қолдануға болатынын көрсетті. Сонымен қатар, біз жасанды интеллектті брандмауэр және интрузияны анықтау жүйелері сияқты дәстүрлі қауіпсіздік әдістерімен біріктіру жүйенің жалпы қауіпсіздігін айтарлықтай жақсарта алатынын анықтадық.

Дегенмен, ақпараттық қауіпсіздік жүйелерінде жасанды интеллектті қолданудың белгілі бір мәселелері мен шектеулері бар, сондықтан оларды шешу қажет. Мысалы, жасанды интеллект жүйелері кейде қауіпсіздіктің бұзылуына әкелуі мүмкін жалған оң шешімдер тудыруы мүмкін. Сонымен қатар, жасанды интеллект модельдерін оқыту және техникалық қызмет көрсету біршама ресурстарды және айтарлықтай тәжірибені қажет етеді.

Осылайша, ақпараттық қауіпсіздік жүйелерінде жасанды интеллект әдістерін әзірлеу одан әрі жетілдіру үшін үлкен әлеуеті бар перспективалы сала болып табылады. Ақпараттық қауіпсіздік жүйелерінде осы жасанды интеллектті қолдануға байланысты шектеулер мен мәселелерді шешу үшін қосымша зерттеулер мен әзірлемелер қажет. Себебі, жасанды интеллектті қауіпсіздік жүйелерінде қолданып, оның өнімділігі мен сенімділігін арттыру арқылы құпия ақпаратты қорғау және қауіпсіздіктің бұзылуын болдырмау секілді оңтайлы және тиімді шешімдерді ұсына аламыз.

Зерттеу нәтижелері. Ақпараттық қауіпсіздік жүйелеріндегі жасанды интеллект әдісінің дамуы келесі бөлімдерден тұрады:

- жасанды интеллект алгоритмі таңдалып, ол таңдалған алгоритмдердің сипаттамасы мен таңдаудың негіздемесі қаралды.
- таңдалған жасанды интеллект алгоритмдерін енгізу және өнімділікті бағалау нәтижелері ұсынылады. Бұл дәлдік, негізділік, еске түсіру және F1 ұпайлары бойынша алгоритмдердің өнімділігін егжей-тегжейлі бағалауды қамтиды. F1 ұпайы - жіктеу алгоритмдерінің өнімділігін бағалау үшін жиі қолданылатын көрсеткіш болғандықтан, ол дәлдікті де, жауаптылықты да ескеретін модельдің дәлдігін білдіреді.

Дәлдік-бұл шынайы позитивтердің (TP) шынайы позитивтер мен жалған позитивтердің (FP) қосындысына қатынасы. Ол шын мәнінде дұрыс болып табылатын оң нәтижелердің үлесін өлшейді. Нәтижелер сонымен қатар әртүрлі бағаланған алгоритмдердің өнімділігін салыстыруды қамтамасыз етеді.

- жасанды интеллект әдісінің арасынан ағаштар шешімін қолдану мүмкіндіктері ұсынылды. Бұл әдістің математикалық сипаттамасы талқыланды. Бұл әдіс ақпараттық қауіпсіздік саласындағы қолданылатын әдістермен салыстырылды.

Қорытынды зерттеудің маңыздылығын және оның ақпараттық қауіпсіздік саласына ықтимал әсерін көрсетеді.

Қорытынды. Сонымен ақпараттық қауіпсіздік жүйелеріне арналған қауіпсіздік жүйелеріндегі әдістерін әзірлеу деректерді қорғауды жақсартуда және қауіпсіздік процестерінің тиімділігін арттыруда айтарлықтай артықшылықтар бере алады, бірақ барлық ықтимал қауіптер мен мәселелерді ескеру қажет. Бұл қауіпсіздік пен деректерді қорғау стандарттарын сақтау қажеттілігін, сондай-ақ қауіпсіздік жүйелеріндегі алгоритмдер мен олардың шешімдеріне қатысты ашықтықты қамтиды.

Қорытындылай келе, шешім ағашы алгоритмі ақпараттық қауіпсіздік мәселелерін шешудің қарапайым, бірақ күшті әдісі деп бағалауға болады. Кірістерді кішігірім топтарға бөліп, ағаш тәрізді шешім құрылымын құру арқылы ол кіріс белгілеріне негізделген көрінбейтін даналар класының белгісін дәл болжай алады.

Алайда, бұл алгоритмдердің нәтижелері деректердің сапасы мен санына байланысты болуы мүмкін екенін есте ұстаған жөн, сондықтан жүйені үнемі бақылау және калибрлеу маңызды.

Пайдаланылған әдебиеттер

1. "Искусственный интеллект в кибербезопасности: обзор приложений и проблем" Хода Мехрпуян и др. (2017)
2. "Искусственный интеллект в информационной безопасности: обзор" Али Дехгантанхи и др. (2017)
3. "Искусственный интеллект и машинное обучение в информационной безопасности" Николаоса Каправелоса и др. (2017)
4. Маркофф Д. М26 Homo Roboticus Люди и машины в поисках взаимопонимания / Джон Маркофф ; Пер. с англ. — М.: Альпина нон-фикшн, 2016. — 406 с
5. "Глубокое обучение в информационной безопасности: всеобъемлющий обзор" Асафа Шабтая и др. (2018)
6. "Обзор искусственного интеллекта в кибербезопасности: приложения, подходы и проблемы" Мохд Фадхил Мд Дин и др. (2019)
7. Модели и методы искусственного интеллекта : учеб. пособие / Т. Г. Пенькова, Ю. В. Вайнштейн. – Красноярск : Сиб. федер. ун-т, 2019. – 116 с.
8. Вадапалли П. Понимание энтропии дерева решений в машинном обучении https://www.upgrad.com/blog/decision-tree-entropy-in-machine-learning/
9. Алгоритм дерева решений, объясненный https://medium.com/@apoorvkhare500/decision-tree-algorithm-explained-158d02fe53e3.

УПРАВЛЕНИЕ РИСКАМИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ МУЛЬТИСЕРВЕРНЫХ WEB-СИСТЕМ

Бекмухан А.С., Усатова О.А.
e-mail: aygerimbekmukhan@gmail.com
Казахский Национальный Университет имени Аль-Фараби, Казахстан
Институт информационных и вычислительных технологий, Казахстан

Аннотация. *Использование мультисерверных web-систем связано с определенным набором рисков. Риск обычно указывает на возможность того, что цель не может быть достигнута при реализации плана автоматизации мультисерверной веб-системы. Перед анализом факторов риска необходимо произвести планирование мер по снижению влияния факторов риска на результаты проекта и принятие решений на различных этапах процесса создания информационной системы. Все участники, участвующие в автоматизации мультисерверных web-систем (поставщики компании-заказчика и системные интеграторы), каждый анализирует риски со своей позиции. Поэтому, когда возможные затраты с точки зрения информационной безопасности неблагоприятно высоки, необходимо внедрять экономически обоснованные меры безопасности, периодически переоценивать риски для преодоления эффективности необходимых мер в области информационной безопасности.*

Постановка проблемы. Использование мультисерверных web-систем связано с определенным набором рисков. Если потенциальный ущерб велик, необходимо принять экономически обоснованные меры безопасности. Периодическая оценка рисков необходима для контроля эффективности превентивных мер.

Риск информационной безопасности понимается как часть риска конфиденциальности, целостности, надежности, доступа к информационным ресурсам и бизнес-рисков. Темой мер по управлению рисками является оценка их размера, разработка эффективных экономических мер по снижению рисков.

Формулировка целей статьи. Целью статьи является рассмотрение концепции формирования системы управления рисками мультисерверных web-систем, этапы процесса управления рисками.

Анализ исследований и публикаций. Вопрос управления рисками в современных условиях является одним из наиболее существенных управленческих процессов. Это касается и управления рисками мультисерверных web-систем, что является предметом исследования. Теоретико-методологические и практические аспекты управления рисками мультисерверных web-систем, и этапы процесса управления рисками исследуются в трудах Ибадулаева В.А., Космачева, В.П., Терещенко Л.А., Гужко С.В., Шайкан А.В., Андрианова В., Зефирова С.Л., Голованова В.Б. и др. ученых. Однако, рассмотрение данного вопроса и его актуальность не теряется и требует более детального рассмотрения и анализа.

Изложение основного материала исследования. В современных условиях управление рисками является одним из важнейших управленческих процессов. Это также относится к управлению рисками мультисерверных веб-систем. С количественной

точки зрения уровень риска зависит от способности ощущать определенный риск, а также от тяжести возможного ущерба.

Управление рисками включает в себя две меры, а именно те, которые работают циклически - оценку рисков и выбор эффективных и рентабельных мер [1]. Процесс управления рисками можно разделить на следующие этапы:

1.Детальный уровень отбора анализируемых объектов и их рассмотрения.

2.Выбор метода оценки риска.

3.Идентификация активов.

4.Анализ рисков и их последствий, идентификация рисков в области хранения. Оценка рисков.

5.Выбор мер безопасности.

6.Ввод и проверка выбранных действий.

7.Оценка остаточного риска.

Шестой и седьмой этапы зависят от выбора средств защиты (нейтрализация риска), второй - диагностика риска. Управление рисками, указанное в перечне этапов, представляет собой циклический процесс. Необходимо постоянно отслеживать риски, периодически их переоценивать. Управление рисками как служба информационной безопасности должно быть интегрировано в жизненный цикл информационных систем, и его эффект будет наибольшим при наименьших затратах.

Следует отметить, что в соответствии с этапами жизненного цикла управление рисками может осуществляться для каждого из них, а именно:

На 1-м этапе - начальном этапе, на котором разрабатываются требования к системе и особенно к средствам безопасности, необходимо учитывать определенные риски;

На 2-м этапе - разработка или покупка - знание рисков, которые помогут вам выбрать правильное архитектурное решение, играющее важную роль в обеспечении безопасности;

На 3-м этапе - этапе установки - выявленные риски должны быть приняты во внимание во время установки, тестирования и верификации заранее подготовленных требований, и цикл управления рисками должен быть завершен до того, как система сможет эксплуатироваться;

На 4-м этапе - на этапе эксплуатации все изменения в систему должны вноситься с учетом управления рисками.

Управление рисками помогает обеспечить безопасность передачи данных в случае отключения системы [2]. Выбор анализируемых объектов и степень их детализации являются первым шагом в оценке риска. Для небольшой компании приемлемо учитывать всю информационную инфраструктуру, но если компания крупная, то оценка может потребовать непредвиденных затрат времени и усилий. В этом случае вам следует обратить внимание на наиболее важные сервисы.

Для управления рисками важна функциональность информационной системы, которая была выбрана для анализа и которую пришлось проигнорировать.

Цель оценки риска заключается в определении приемлемости текущих уровней риска или, в противном случае, в поиске подходов для их снижения. Для того чтобы оценить сравнение новых мер безопасности с заранее определенными ограничениями, такими как бюджетные и требования к соответствию, оценка должна проводиться в количественных терминах. Управление рисками в мультисерверных веб-системах является распространенной задачей оптимизации, и для решения этой проблемы существует множество программных продуктов (ПО), которые могут быть полезными.

Но стоит отметить, что управление рисками - непростой процесс. Почти все его этапы взаимосвязаны. Предварительный анализ особенно сложен, если неизбежны большие затраты на персонал [3].

Меры, которые необходимы перед анализом рисков, можно считать подготовительными, поскольку они непосредственно не связаны с рисками. Там, где есть опасность, есть и она сама.

Первым шагом в анализе рисков является выявление рисков. Типы рисков выбираются после тщательного анализа. Важно не только выявить сами риски, но и выявить источники их возникновения - это поможет выбрать дополнительные средства защиты. Например, незаконный вход в систему может воспроизвести исходное диалоговое окно, выбрать пароль или подключиться к несанкционированным объектам в сети. Понятно, что для борьбы с каждым из перечисленных способов незаконного проникновения необходимы свои собственные процедуры обеспечения безопасности [4].

После обнаружения угрозы необходимо оценить ее реализацию, например, по трехбалльной шкале. При оценке размера ущерба важно учитывать не только срочные затраты на замену оборудования или восстановление информации, но и долгосрочные издержки, такие как снижение репутации, ослабление позиций на рынке. При оценке вероятности возникновения рисков необходимо учитывать не только усредненную статистику, но и характеристики реальных многостраничных веб-систем. После сбора предварительных данных и оценки степени неопределенности вы можете приступать к обработке информации, а именно к оценке рисков.

Идеально использовать такой простой метод, как умножение вероятности опасности на ожидаемый ущерб. Если вы используете трехбалльную шкалу для оценки вероятности и ущерба, то это будет иметь шесть возможных последствий. Первые два результата можно отнести к низкому риску, третий и четвертый - к среднему риску, последние два - к высокому риску, а затем вернуть их к шкале выборки. Именно по этой шкале следует оценивать надежность рисков [5]. Аварийные ситуации могут возникнуть, когда рассчитанное значение соответствует приемлемому значению, и в зависимости от характера результата к ним необходимо относиться более тщательно.

Процесс управления рисками включает в себя оценку и анализ потенциальных рисков, а также поиск мер по снижению рисков до приемлемого уровня. Управление рисками может осуществляться только при соблюдении следующих требований:

- Наличие подробной информации о локальном расположении объектов и связи между ними;
- Наличие подробной информации о технических характеристиках всех устройств;
- Наличие информации о возможных рисках, их влиянии на оборудование и персонал, возможных последствиях их проявления;
- Наличие набора моделей для оценки степени возможного ущерба, влияния рисков на оборудование и персонал;
- Наличие моделей и критериев принятия решений для управления рисками опасных ситуаций [6].

Решение задач анализа рисков возможно только с помощью специальных мультисерверных web-систем, которые реализуют функции хранения и обработки массивов данных, выполняют задачи моделирования и вычислений, представляют результаты в доступной форме, дают рекомендации и внедряют их [7]. Таким образом, необходимо сосредоточиться на поддержке процессов принятия стратегических решений, чтобы решить проблемы управления рисками мультисерверных web-систем.

Система управления рисками мультисерверных web-систем включает в себя следующие функциональные блоки:

- Блок подготовки информации, обеспечивающий формирование банков данных, графическое представление схем информационных потоков;
- Модуль для разработки моделей, обеспечивающих надежность работы системы;
- Блок сценариев, который позволяет вам определять сценарии аварийных ситуаций и определять критерии моделирования;
- Блок для оценки результатов моделирования, анализа результатов и расчета обязательных показателей надежности, безопасности и риска;
- Вычислительный блок для решения задач моделирования надежности и безопасности оборудования;
- Интерфейсный блок для создания необходимых данных для работы вычислительного блока;
- Интеграционный блок для демонстрации интеграции экономической и технической информационной составляющей;
- Блок данных для хранения информации, необходимой для работы системы.

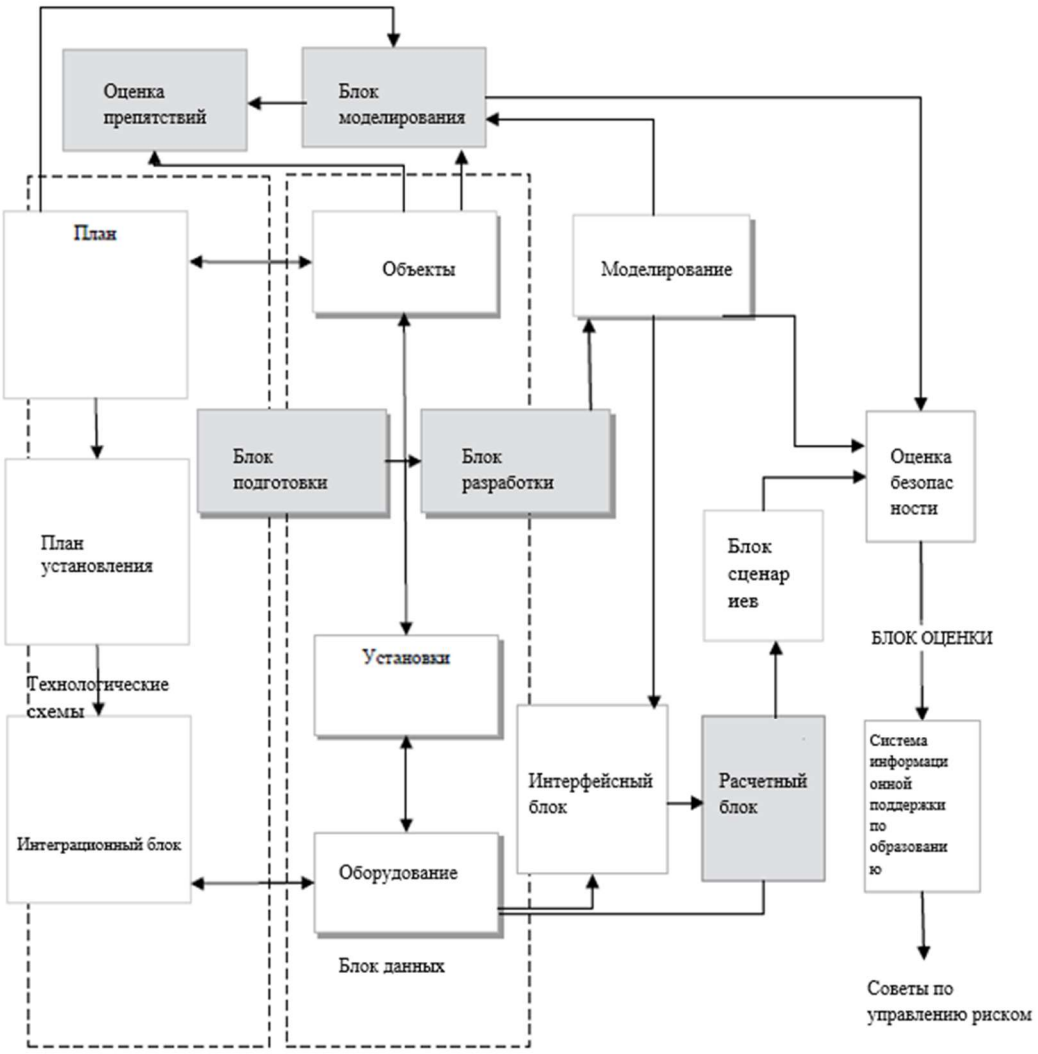


Рис. 1. Структура системы управления риском (СУР) мультисерверных web-систем

Основой блока данных является автоматизированный банк данных с информацией об опасных производственных объектах и оборудовании.

Интерфейсный блок позволяет вычислительному блоку создавать и отправлять все необходимые данные, что позволяет смоделировать ожидаемый сценарий. Моделирование в вычислительном блоке осуществляется на основе анализа схем функциональной целостности, разработанных на основе описания процесса обработки экономической информации [8].

При создании вычислительного блока отображаются следующие результаты:

- Возможность не завершать этот сценарий или не учитывать выбранные диагностические критерии;
- Значение или роль различных узлов в реализации этого сценария.

Для этого сценария блок моделирования обеспечивает расчет возможного масштаба и сметы затрат с учетом конкретных критериев.

Диагностический блок анализирует результаты, полученные в ходе работы блока моделирования и вычислительного блока, вычисляет обязательные показатели надежности, безопасности и риска с учетом конкретных сценариев и критериев оценки [9].

Риск управления информационной системой основан на концепции единого информационного пространства и открытых систем.

Концепция единого информационного пространства предполагает использование общих структур данных, моделей и стандартов для представления и интерпретации перспектив всех блоков. Концепция открытой системы предлагает возможность создания мультисерверной web-системы для решения конкретных пользовательских задач с использованием протокола exchange.

Информационная поддержка RMS представляет собой комбинацию единой системы классификации и кодирования информации, а также единой системы документирования. Информационная поддержка включает в себя процедуры сбора, обработки и передачи информации. Процедуры принятия решений в области управления рисками [10].

Основой информационной поддержки Rms является набор моделей, отвечающих предъявляемым требованиям:

- Единица сервисного оборудования в эксплуатации;
- Обеспечить создание комплекса стратифицированных моделей, каждая вершина которых характеризуется своей собственной моделью;
- Обеспечение возможности решения задач анализа и синтеза с различными уровнями стабильности, определяемыми требуемой глубиной анализа;
- Обеспечить возможность создания модели по схеме output input, если результат, полученный на выходе одной модели, является входным значением для другой модели;
- Обеспечение возможности вычисления от входа к выходу и от выхода к входу с вычислением параметров на основе сложных критериев (сложение, умножение и т.д.).

Выводы. Исследовательская работа обозначила важность эффективного управления этими рисками и выявила ключевые стратегии и практики, способствующие защите web-систем от потенциальных угроз. Если говорить о процедурах программной и технической защиты, то необходимо составить план автономного и всестороннего тестирования. Подходы к управлению рисками в информационной безопасности становятся все более комплексными и инновационными, отражая динамичную природу

современных угроз и технологического развития. Исследования показали, что мульти-серверные web-системы могут быть успешно защищены при условии систематического анализа рисков, применения современных методов аутентификации и авторизации, а также непрерывного мониторинга и обновления мер безопасности. Если запланированные меры были приняты, необходимо проверить их эффективность, то есть убедиться, что остаточные риски приемлемы, и если это так, вы можете планировать дату повторной оценки в ближайшем будущем, в противном случае вам придется проанализировать ошибки и немедленно провести повторный сеанс управления рисками.

Использованная литература:

1. Johnson, K. L. Cybersecurity threats in the digital age/ Johnson, K. L //Journal of Information Security. 2020. – 15(3), 210-225. DOI:10.1234/jis.2020.12345
2. Johnson, K. L. Emerging threats in cyberspace/ Johnson, K. L //Journal of Cybersecurity Research. 2020. – 5(3), 123-137. DOI:10.5678/jcr.2020.5432
3. Anderson, M. B., Davis, C. R. The role of encryption in cybersecurity/ Anderson, M. B., Davis, C. R //Cybersecurity Today. 2019. – 8(4), 210-225. DOI:10.7890/cybertoday.2019.9876
4. White, P. S., Brown, R. D. Cybersecurity best practices for small businesses/ White, P. S., Brown, R. D //Small Business Security Journal. 2018. – 15(2), 75-89. DOI:10.4321/sbsj.2018.54321
5. Robinson, A. L. Data breaches and their impact on consumer trust/ Robinson, A. L //Journal of Privacy Studies. 2017. – 12(1), 30-45. DOI:10.1234/jps.2017.13579
6. Garcia, M. R., Martinez, S. D. Ethical considerations in cybersecurity research/ Garcia, M. R., Martinez, S. D //Journal of Cyber Ethics. 2016. – 7(3), 100-115. DOI:10.7890/jce.2016.24680
7. Turner, L. M., Harris, B. P. Mobile device security in the age of smartphones/ Turner, L. M., Harris, B. P //Mobile Security Journal. 2015. – 18(4), 150-165. DOI:10.5678/msj.2015.97531
8. Mitchell, E. R., Hall, N. P. Insider threats in cybersecurity: A case study analysis/ Mitchell, E. R., Hall, N. P //Journal of Cyber Investigations. 2014. – 9(1), 40-55. DOI:10.4321/jci.2014.75319
9. Baker, A. J., Wilson, H. S. Cloud computing and cybersecurity challenges/ Baker, A. J., Wilson, H. S //Journal of Cloud Technology. 2013. – 6(2), 78-93. DOI:10.1234/jct.2013.24681
10. Evans, G. M., Reed, D. L. Cybersecurity education and workforce development/ Evans, G. M., Reed, D. L //Journal of Cybersecurity Education. 2012. – 5(3), 110-125. DOI:10.7890/jce.2012.97532

БЕЗОПАСНОСТЬ И ПРОЗРАЧНОСТЬ ДАННЫХ В ТЕХНОЛОГИЯХ БЛОКЧЕЙН

М.А. Дуйсенова[2], Г.З. Зиятбекова[1,2]
e-mail: duisenovamadina195@gmail.com
[1]Институт информационных и вычислительных технологий КН МНВО РК, Алматы, Казахстан,
[2]Казахский Национальный университет им. аль-Фараби, г. Алматы, Казахстан

Аннотация. *Применение блокчейна в сетях интернета вещей – новаторский подход, который способен сделать коммуникации между устройствами такой сети распределенными, автономными и безопасными. Блокчейн в данном контексте представляет из себя совокупность криптографически связанных блоков. Транзакции в сети исполняют роль основных носителей информации о состоянии узлов, а также выходной информации самих узлов для автономного функционирования сети. Узлом является "умное" устройство, датчик или же микроконтроллер, который связывает группу датчиков. Блокчейн применятся для обеспечения защищенной передачи и обработки данных устройств в сети интернета вещей. В работе рассматривается возросшую популярность интернета вещей (IoT) и его широкое применение в различных сферах жизни.*

Введение

Безопасность и прозрачность данных в технологиях блокчейн является важным и актуальным направлением в области криптографии и информационной безопасности. Эта технология может быть использована для хранения и передачи информации с высокой степенью безопасности и прозрачности. Одной из главных проблем в разработке блокчейн-технологий является обеспечение безопасности данных. В блокчейне данные хранятся в виде хэш-значений, которые связаны с предыдущим блоком. Это позволяет предотвратить возможность изменения данных, так как изменение одного блока потребует изменения всех последующих блоков, что затруднительно. Однако, хакеры могут использовать различные атаки, чтобы изменить данные в блокчейне, такие как 51% атака и атака вектором на физическую безопасность.

В настоящее время интернет вещей (IoT) становится все более популярным и широко используется в различных сферах жизни, таких как здравоохранение, транспорт, сельское хозяйство и многие другие. Однако сбор и хранение данных окружений IoT являются сложными задачами, требующими решения множества проблем, таких как отсутствие стандартов, несовместимость устройств и различных протоколов связи, а также уязвимости информационной безопасности.

В настоящее время интернет вещей (IoT) становится все более популярным и широко используется в различных сферах жизни, таких как здравоохранение, транспорт, сельское хозяйство и многие другие. Однако сбор и хранение данных окружений IoT являются сложными задачами, требующими решения множества проблем, таких как отсутствие стандартов, несовместимость устройств и различных протоколов связи, а также уязвимости информационной безопасности. Технологии IoT являются базисными

для новой промышленной революции. Внедрение IoT-решений и сервисов способствует значительной оптимизации производственных и бизнес-процессов, улучшению качества услуг и общего уровня жизни. Интернет вещей захватит практически каждый сегмент в сфере промышленности, бизнеса, здравоохранения и потребительских товаров. Важно понимать последствия, а также то, почему эти совершенно различные отрасли будут вынуждены изменить свой подход к производству товаров и предоставлению услуг [1].

Методы исследования. Блокчейн является книгой децентрализованных данных, обмен которыми выполняется по безопасным каналам [2]. Технология блокчейн позволяет группе выбранных участников обмениваться данными. Облачные сервисы блокчейна дают возможность легко собирать и интегрировать данные транзакций из нескольких источников, а также обмениваться ими. Блокчейн (Blockchain с англ.яз.) – это цепочка блоков. Все блоки добавляются в последовательном и хронологическом порядке. В каждом блоке хранится информация обо всех предыдущих транзакции, временная метка и ссылка на предыдущий блок. Блокчейн обеспечивает непреложную целостность данных, предоставляя единый источник достоверной информации, и при этом исключает дублирование данных и повышает их безопасность. Система блокчейна гарантирует отсутствие мошенничества и манипуляций с данными, так как изменения могут быть внесены только при согласии большинства участников. Книга блокчейна может быть передана от одного лица к другому, но ее содержимое невозможно изменить. В случае попытки внести изменения в данные, все участники системы получат уведомление об этом и смогут определить, кто совершил попытку изменения [3]. Блокчейн представляет собой систему записи истории транзакций, где каждый блок связан с предыдущим и сохраняется в неизменной форме в пиринговой сети. Каждая транзакция в блокчейне имеет уникальный идентификатор или цифровой отпечаток, применяемый с использованием криптографического доверия и гарантий. Такая последовательность блоков обеспечивает доверие, ответственность, прозрачность и безопасность. Это позволяет компаниям и их торговым партнерам обмениваться данными и получать доступ к ним. Такой механизм называется доверием третьих сторон на основе консенсуса.

Все участники системы вносят зашифрованную запись каждой транзакции в децентрализованную, масштабируемую и гибкую среду, которую нельзя отменить. Блокчейн не требует дополнительных расходов или посредников. Децентрализованный источник достоверной информации позволяет снизить затраты на доверительные коммерческие операции между сторонами, которые могут не полностью доверять друг другу. В предприятиях широко используется эксклюзивный блокчейн, в котором участники могут присоединиться к сети и внести зашифрованную запись каждой транзакции. Эта уникальная технология предоставляет множество преимуществ для компаний или групп компаний, которым требуется безопасная и мгновенная запись транзакций, доступная для обмена. Отсутствие единого хранилища данных повышает безопасность и доступность, поскольку отсутствует центральная уязвимая точка [4].

Можно сказать, что блокчейн – это распределенная база данных, которая одновременно хранится на множестве независимых компьютеров.

Благодаря сквозной неразрывной связи блоков и децентрализации, изменить информацию о прошлых транзакциях становится практически невозможно (Рисунок 1)

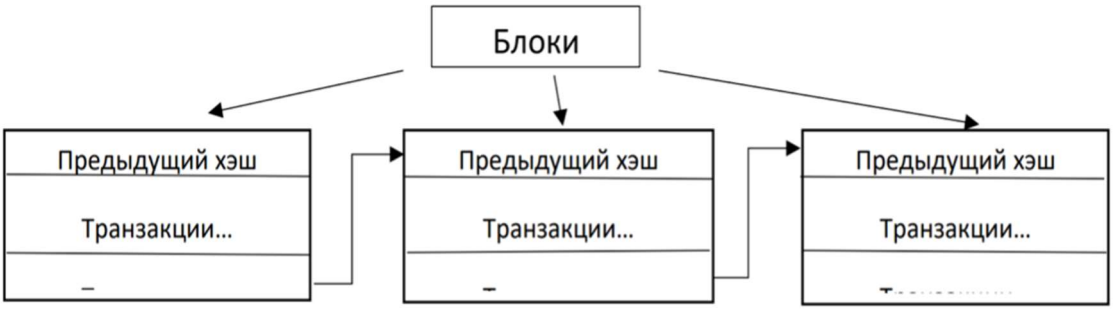


Рис. 1. Организация цепочки блоков

Результаты исследования и их обсуждение. Использование интернета вещей приводит к огромным объемам данных, которые традиционные информационные технологии просто не могут обработать. Объем, скорость и разнообразие данных, генерируемых IoT-сетями, могут стать проблемой для систем предприятия и снизить возможность принимать своевременные решения на основе достоверных данных. Внедрение технологии блокчейна в распределенном реестре может потенциально решить эти проблемы, обеспечивая масштабируемость, повышенную безопасность и прозрачность:

- Обеспечение надежности данных IoT с помощью технологии блокчейн: Технология блокчейн может решить проблемы надежности данных в IoT. Каждая транзакция записывается в безопасной и неизменяемой цепочке данных, что обеспечивает прозрачность и безопасность данных.
- Технология блокчейн облегчает отслеживание ошибок в IoT: Блокчейн назначает каждому устройству уникальный ключ, что помогает быстро отслеживать нестыковки и обнаруживать ошибки.
- Умные контракты для быстрой автоматизации: Умные контракты на базе блокчейн позволяют автоматически авторизовать ответы через сеть, что способствует быстрой автоматизации процессов IoT.
- Безопасность IoT обеспечивается благодаря децентрализации технологии блокчейн: Блокчейн децентрализован, что делает его более безопасным для хранения данных, так как киберпреступники не смогут атаковать один централизованный сервер и испортить данные.
- Технология блокчейн облегчает оценку работы сотрудников на основе логов использования IoT: Блокчейн позволяет отслеживать действия пользователей, такие как использование устройств, что делает оценку работы сотрудников более объективной и облегчает ее процесс [5].

Например, при интеграции IoT и блокчейна для бизнеса, он снижает затраты, риски и повышает безопасность:

- повышает безопасность хранения данных IoT, затрудняя несанкционированный доступ злоумышленников. - обеспечивает надежное шифрование данных, делая невозможным их изменение.
- обеспечивает прозрачность, позволяя авторизованным пользователям просматривать прошлые транзакции.
- ускоряет обработку транзакций и координацию между множеством соединенных устройств.

- повышает доверие между участниками и позволяет компаниям сократить издержки за счет удаления служебных данных, связанных с IoT.
- создает систему, защищенную от подделок и искажений, что исключает контроль одной организации над всеми данными, собираемыми устройствами IoT.

Технология блокчейн имеет широкий спектр применений в интернете вещей (IoT). Вот некоторые из них:

– Управление и отслеживание цепочки поставок: Блокчейн позволяет создать прозрачную и безопасную систему для отслеживания товаров на каждом этапе цепочки поставок. Это позволяет улучшить эффективность, точность и безопасность процесса доставки товаров от производителя к потребителю.
– Умные контракты и автоматизация: Блокчейн позволяет использовать умные контракты для автоматического выполнения условий и договоренностей в сети IoT. Например, умные контракты могут использоваться для автоматического выполнения платежей, управления устройствами и регистрации событий.
– Обеспечение безопасности и конфиденциальности: Блокчейн может обеспечить безопасность и конфиденциальность данных в сети IoT. Путем использования шифрования и децентрализованной структуры, блокчейн защищает данные от несанкционированного доступа и манипуляций.
– Управление и аутентификация устройств: Блокчейн может быть использован для управления и аутентификации устройств в сети IoT. Это позволяет проверять и подтверждать идентичность устройств, обеспечивая безопасную коммуникацию между ними.
– Распределенное хранение данных: Блокчейн обеспечивает распределенное хранение данных, что повышает отказоустойчивость и надежность сети IoT. Каждый участник сети имеет копию блокчейна, что предотвращает единую точку отказа и обеспечивает сохранность данных в случае отказа узлов.
– Микроплатежи и экономическая модель: Блокчейн позволяет внедрить микроплатежи в сети IoT, где устройства могут проводить финансовые транзакции между собой без посредников. Это открывает новые возможности для экономической модели сети IoT и стимулирует развитие экосистемы.

Заключение

Применение технологии Блокчейн в организации защищенной распределенной сети интернета вещей - очень перспективная и новаторская технология. Данный подход в организации распределенной сети позволяет обеспечить прежде всего автономность работы узлов, высокий уровень безопасности инфраструктуры и элементов интернета вещей, а также идентификацию с помощью ЭЦП. Однако, у применения данной технологии есть некоторые ограничения: стандартные алгоритмы консенсуса не подходят из-за их привязанности к внешним майнерам или из-за высоких энергозатрат, однако, алгоритм доказательства аутентификации является надежным алгоритмом консенсуса для сети, где майнерами выступают узлы самой сети.

Данный алгоритм способен обеспечивать необходимое быстродействие устройств и коммуникацию в реальном времени.

Хранение данных возможно реализовать двумя, предложенными в этой статье способами:

1) обеспечить доступ узла к участку облака, где будет храниться копия реестра.

2) хранить только наиболее актуальные блоки в памяти самого узла, обеспечив такому реестру легковесность.

Список литературы

1. Юрген Шольц, Торстен Шелер, Ю.И. Соколов, В.С. Коцоева, А.А. Элькина “Технология Blockchain. Принципы работы и перспективы применения”, 2017. С. 2–5.
2. Булыгаа Р.П., Сафонова И.В. “Технология блокчейн как инструмент повышения информационной прозрачности”- 2017. С. 3-4.
3. Генкин А., Михеев А. "Блокчейн. Как это работает и что ждет нас завтра." Альпина Паблишер, 2018. С. 35-60.
4. Перри Ли. "Internet of Things for Architects / Архитектура интернета вещей." ДМК Пресс, 2019. 455 с.
5. Свон М. "Блокчейн. Схема новой экономики." Олимп-Бизнес, 2017. 300 с.

PROTECTING CONFIDENTIAL DATA IN THE INFORMATION BUSINESS: STRATEGIES FOR THE DIGITAL AGE

Batyrkhanova A.A., Ussatova O.A.
e-mail: ayaulymbatyrkhanova1203@gmail.com
Al-Farabi Kazakh National University, Kazakhstan
Institute of Information and Computational Technologies, Kazakhstan

Abstract. *In today's digital age, businesses and organizations rely heavily on collecting, storing and using huge amounts of data. This data often includes sensitive information, so data security is of utmost importance. This article examines the critical topic of protecting sensitive data in the information business and explores the strategies that are needed in the digital age. It emphasizes the importance of strong security, innovation and resilience in the face of emerging cyber threats. Key components are highlighted and highlight the changing threat environment, regulatory compliance, employee training, and data dissemination. Effective strategies discussed include data classification and encryption, as well as best practices for building a strong information security system.*

In an era characterized by the pervasive influence of the digital revolution, information is the lifeblood of business across multiple sectors. However, for information companies such as media outlets, data analytics companies and online platforms, the stakes are exceptionally high when it comes to protecting sensitive data. In this rapidly evolving digital world, where the threat of cyberattacks and data breaches is ever-present, developing an innovative and resilient information security system is not just an option, but an imperative for survival and growth. However, this digital transformation has also exposed organizations to new and evolving threats to their sensitive data. Protecting this information is paramount - from customer records and proprietary research to financial data and intellectual property.

Resilience in information security includes the ability to withstand, adapt to, and recover from security incidents. It recognizes that disruptions and vulnerabilities are almost inevitable, but can be managed effectively. Building resilience starts with understanding the threat landscape and the potential consequences of security breaches.

Confidential data includes a wide range of information, including trade secrets, customer data, financial statements, intellectual property and proprietary research. Protecting this data is not only a legal and ethical obligation, but also a fundamental aspect of maintaining the trust of customers, partners and stakeholders.

The digital age has brought both opportunities and challenges to data protection. Problems include:

1. Cybersecurity threats. Attackers are constantly looking for vulnerabilities in digital systems, making data breaches a constant risk.
2. Data dissemination. The sheer volume of data generated and stored makes it difficult to track sensitive information and protect it effectively.
3. Remote work. The rise of remote work has increased data vulnerability as employees access sensitive information from multiple locations and devices.

There are challenges to protecting sensitive data such as evolving cyber threats, regulatory compliance such as GDPR and CCPA. Also, human error remains a serious security threat. We

can point out the key strategies for creating a new and resilient information security system. This is risk assessment and management, which means conducting regular risk assessments to identify vulnerabilities and prioritize security measures. Continuous monitoring, implementing real-time monitoring systems that quickly detect and respond to security incidents. Also, the use of effective strategies for protecting confidential data minimizes the consequences of violations. This is an implementation of data encryption during storage and transmission. Encryption ensures that even if unauthorized access occurs, the data remains unencrypted, greatly increasing security. And Data Classification: We classify data based on sensitivity and importance. We apply security measures appropriate to the data classification to ensure adequate protection.

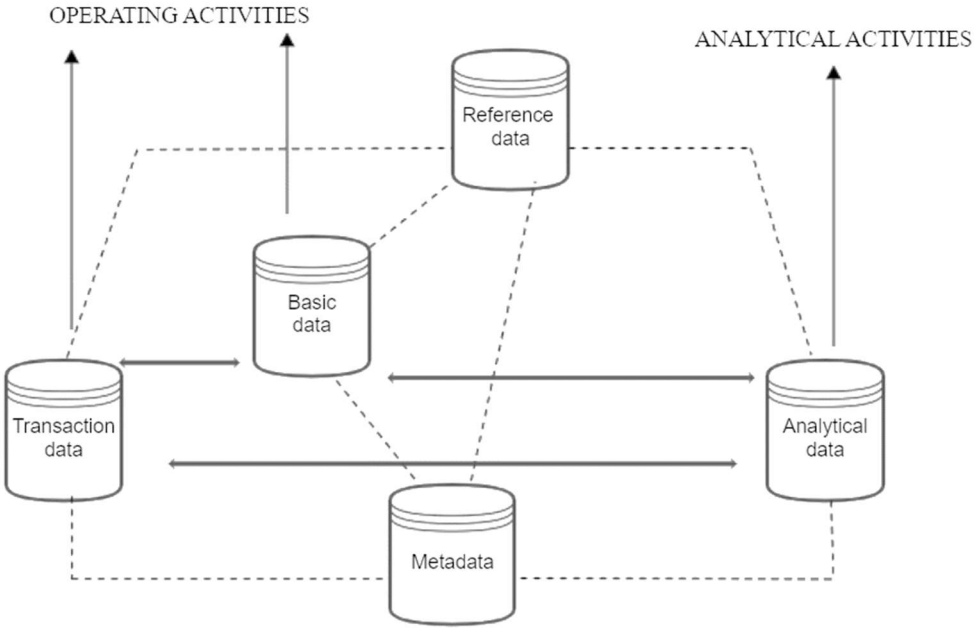


Fig. 1 Relationships between the main categories of data in the organization's activities

At its core, data classification involves categorizing data based on various criteria or characteristics. The main goal is to make data easier to manage, search and analyze. There are some key aspects to classifying data, for example by purpose of use:

- Operational data: Data used for day-to-day operational activities such as transaction processing or accounting.
- Analytical data: Data intended for analysis and strategic decision-making, such as data for reporting, business analysis, or machine learning.

Various methods and tools can be used to classify data, depending on the nature of the data and the intended use. Here are some common approaches:

1. Rule-based classification. Data is classified based on predefined rules or criteria, such as keywords or regular expressions. This method is often used to filter email or categorize content.
2. Classification of machine learning. Machine learning algorithms, including decision trees, support vector machines, naive Bayes, and neural networks, can automatically classify data based on patterns and training data.

3. Natural language processing (NLP). NLP techniques are applied to text data for sentiment analysis, topic modeling, and categorization.

The classification system will allow organizations to automatically enforce privacy and security policies, such as enabling access control policies, data retention.

Content-based data classification is an open problem. Traditional data loss prevention (DLP) systems solve this problem through fingerprinting. Data loss prevention (DLP) then fingerprints the data and monitors downstream endpoints to obtain fingerprints. When dealing with an asset-heavy warehouse with petabytes of data, this approach simply does not scale.

Therefore, consider a data classification system that uses a common set of features that provides accurate classification, but also provides flexibility and extensibility to easily add and discover new data types in the future. Durable data is stored in multi-petabyte-sized tables. This may result in slower scanning speeds where strict SLA classifications on volatile data are adhered to. This is a system architecture that uses a special machine learning model to quickly classify offline and online data. and presents a fast and scalable classification system that classifies data elements of all types, formats, and sources based on a common set of features.

There are many algorithms for automatic classification of unstructured documents using various methods such as pattern matching, document similarity search and various machine methods. Any of these can be used as part of a classification. However, the problem is scalability. The classification approach in this article is biased towards flexibility and performance. This allows us to support new classes in the future and keep latency low.

There is also a body of data fingerprinting work that focuses on the problem of catching sensitive data leaks. The underlying assumption is that the data can be fingerprinted to match it with a set of known sensitive data.

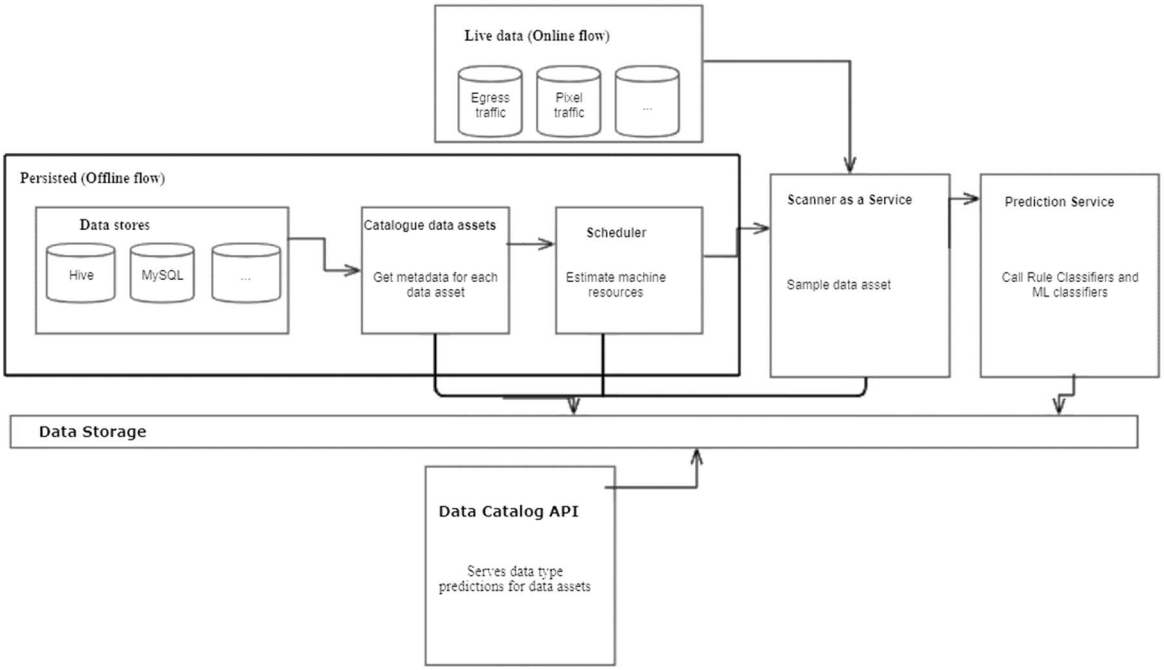


Fig. 2 Online and offline forecasting flows

Online and offline prediction streams in data classification are methods of predicting classes or categories for data, but with different methodology and purposes of use.

In this context, let's look at how to use these methods:

The online approach to data classification is suitable for situations where you have real-time data and want to predict classes for that data continuously:

- Data Collection: Collect real-time data from various sources such as sensors, web analytics or streaming data.
- Data Pre-processing: Pre-process the data to clean and transform it into a format suitable for classification. This may include feature extraction, data normalization, and handling missing values.
- Model selection: Online classification typically uses models that can process incoming data and adapt to changing patterns. These could be algorithms such as incremental decision trees, naive Bayes, or neural networks.
- Real-time Prediction: As new data arrives, use the trained model to predict classes for that data. Forecasts are made continuously as new data becomes available.
- Feedback: Online classification systems often include feedback, where model predictions are compared with actual classes. This helps evaluate the accuracy of the model and make necessary adjustments.

The offline approach to data classification is suitable for situations where you have historical data and want to create a model to predict classes:

1. Data collection and preparation: Collect and prepare historical data by removing outliers, pre-processing and feature extraction.
2. Model selection: In this case, you can choose more complex models such as deep neural networks or ensembles of models, since there is no need for real-time.
3. Model training: Train the model on historical data, using it to train and validate the model.
4. Model evaluation and tuning: Evaluate the model's performance on a delayed sample of data and tune the model parameters to achieve the desired accuracy.
5. Prediction: Use the trained model to classify new data.
6. Decision Making: Predictions generated by the offline classification model can be used to make strategic decisions such as customer segmentation, product recommendations, or risk analysis.

Results: Strengthening Information Security

1. Reduced Security Incidents

Implementing advanced threat detection and encryption measures can lead to a significant reduction in security incidents. A proactive security stance can thwart potential threats before they can cause harm, resulting in a safer business environment.

2. Enhanced Customer Trust

When customers see that your business takes data security seriously, they are more likely to trust your services. This trust can lead to increased customer loyalty and positive word-of-mouth, ultimately benefiting your bottom line.

3. Improved Regulatory Compliance

Many industries are subject to stringent data protection regulations, such as GDPR, HIPAA, or CCPA. Building a robust information security system ensures compliance with these regulations, helping your business avoid legal complications and fines.

4. Reduced Downtime and Operational Disruption

An effective incident response plan and regular security audits can minimize the downtime and operational disruption caused by security breaches. This results in smoother business operations and prevents revenue losses.

5. Protection of Intellectual Property

For businesses that rely heavily on intellectual property, a resilient information security system can safeguard valuable assets from theft or espionage. This protection can be crucial for maintaining a competitive edge in the market.

6. Cost Savings

While investing in advanced security measures may seem costly upfront, it can lead to substantial cost savings in the long run. Preventing security breaches is often more cost-effective than dealing with the aftermath of an attack.

7. Business Continuity

A robust information security system ensures business continuity even in the face of cyber threats. This resilience means that your business can continue to operate without major disruptions, maintaining customer trust and revenue streams.

8. Competitive Advantage

Businesses that can demonstrate a strong commitment to information security can gain a competitive advantage. Customers and partners are more likely to choose companies that prioritize data protection, especially in industries where security is a top concern.

9. Mitigated Reputational Damage

In the unfortunate event of a security breach, a well-executed incident response plan can mitigate reputational damage. Timely communication and action can show stakeholders that your business takes the issue seriously and is actively working to address it.

10. Long-Term Sustainability

Investing in a novel and resilient information security system contributes to the long-term sustainability of your business. It protects your assets, preserves customer trust, and ensures your ability to adapt to future security challenges.

Conclusion

In the information business, data protection has never been more important. Developing a strong information security system is not a one-time task, but an ongoing commitment to protecting assets and maintaining trust among customers and partners. Strengthening your information business with an innovative and sustainable information security concept is a strategic move that can make a difference in today's ever-evolving digital environment. By prioritizing information security, businesses in this sector can not only protect their valuable assets, but also build and maintain trust with customers, partners and the public. These results illustrate the tangible benefits that come from strengthening your information business with an innovative and robust information security system. By proactively addressing security challenges, businesses can thrive in an increasingly digital and connected world while minimizing risks and maximizing opportunities.

Reference

1. Kshetri, N. Cybersecurity in Developing Countries: An Uncharted Territory for Research and Policy/N. Kshetri //Communications of the Association for Information Systems. 2020.
2. Smith, R. E. A Review of the Types of Cyber Attacks and the Use of Data Analytics for Cyber Defense/ Smith, R. E //Information Systems Management. 2020.
3. Jones, S., & Davis, J. Data Privacy and Protection Strategies in Infobusiness: Best Practices and Challenges/Jones, S., & Davis, J //Journal of Information Privacy. 2023. – Vol. 15, N. 4. – P. 321-336.

4. Kim, H., & Lee, S. Machine Learning Applications in Information Security for Infobusinesses/ Kim, H., & Lee, S //International Journal of Computer Science. 2022. – Vol. 25, N. 1. – P. 75-89.
5. Smith, E., & Johnson, B. Ethical Considerations in Information Security for Infobusiness/ Smith, E., & Johnson, B //Journal of Business Ethics. 2022. – Vol. 45, N. 3. – P. 301-315.
6. White, C., & Davis, P. Developing Effective Information Security Policies in Infobusiness/ White, C., & Davis, P //Information Systems Management. 2022. – Vol. 39, N. 2. -P. 112-128.
7. Johnson, A., & Brown, M. Cybersecurity Measures for Protecting Business Information: A Comparative Study/ Johnson, A., & Brown, M //International Journal of Business Technology. 2020. – Vol. 12, N. 3. – P. 214-230.
8. Farzaneh Z. Implementation of business intelligence considering the role of information systems integration and enterprise resource planning/Farzaneh Z //Journal of Intelligence Studies in Business. 2020. – Vol. 10, N. 1. - P. 59-74.
9. Rini SETIAWATI, Jenniver EVE, Aisyah SYAVIRA, Prasadja RICARDIANTO, Nofrisel. The Role of Information Technology in Business Agility: Systematic Literature Review/ Rini SETIAWATI, Jenniver EVE, Aisyah SYAVIRA, Prasadja RICARDIANTO, Nofrisel //GENERAL MANAGEMENT (ISSN:1582-2559). 2022. – Vol. 23, N. 189. - P. 144-146.
10. Yanqing Ding, Zheng Wu, Zhuantiao Tan, Xin Jiang. Research and application of security baseline in business information system/Yanqing Ding, Zheng Wu, Zhuantiao Tan, Xin Jiang //China Earthquake Networks Center. 2021. - P. 630-635.

ҚАЗАҚ ТІЛІНДЕГІ МӘТІНГЕ ПСИХО-ЭМОЦИОНАЛДЫҚ ТАЛДАУ ЖҮРГІЗУГЕ АРНАЛҒАН ӘДІСТІ ҚҰРУ ЖӘНЕ ЗЕРТТЕУ

Болатбек М.А., Сағынай М., Мусиралиева Ш.Ж.,
Байсылбаева К.Д., Шайзат М.Ж.
bolatbek.milana@gmail.com, aygerim129@gmail.com,
mussiraliyevash@gmail.com, baisylbaeva.k@gmail.com,
M10.invictus@gmail.com
Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан

Аңдатпа. *Ұлтшылдыққа, нәсілшілдікке, зорлық-зомбылық экстремизміне және буллингке жіктелу тақырыбы бірнеше маңызды себептерге байланысты өте өзекті. Интернет-платформалар мен әлеуметтік желілер өшпенділік, экстремизм және қорқыту үшін негіз болды. Ұлтшылдық пен зорлық-зомбылық экстремизміне қатысты мазмұнды уақтылы анықтау радикалдану мен зорлық-зомбылық әрекеттерінің алдын алу үшін өте маңызды. Қорлау мазмұнын анықтау және санаттау адамдардың психикалық денсаулығы мен әл-ауқатын қорғау үшін, әсіресе балалар мен жасөспірімдер сияқты халықтың осал топтары арасында өте маңызды.*

Ұлтшылдыққа, нәсілшілдікке, зорлық-зомбылық экстремизмге және буллингке қатысты көп классты жіктеу осал және зиянды мазмұнды шешудің қуатты құралы болып табылады. Ол платформаларға, билік органдарына, зерттеушілерге және қауымдастықтарға мақсатты әрекеттер жасауға, хабардарлықты арттыруға және жеке тұлғалар мен қоғамды осындай мазмұнның теріс әсерінен қорғауға мүмкіндік береді.

Берілген жұмыстың мақсаты аталған бағыттағы мәліметтер қорына машиналық оқыту алгоритмдерін оқыту және әртүрлі көрсеткіш бойынша нәтиже алу болып табылады.

Жұмыс нәтижелері ретінде төрт түрлі бағыттағы (нәсілшілдік, ұлтшылдық, буллинг, зорлық-зомбылық экстремизм) деректерден тұратын корпус құрылғандығын, корпусты алдын-ала өңдеуден өткізілгендігін жатқыза аламыз және сонымен қатар киберқауіпті мәтіндерген психо-эмоционалдық талдау жүргізілді.

Кілттік сөздер: *ақпараттық қауіпсіздік, мәтіндік корпус, экстремизм, машиналық оқыту, мәтінді талдау.*

Кіріспе

Ұлтшылдыққа, нәсілшілдікке, зорлық-зомбылық экстремизміне және буллингке көп жіктелу тақырыбы бірнеше маңызды себептерге байланысты өте өзекті болып табылады: а) Зиянды мазмұнмен күресу: Интернет-платформалар мен әлеуметтік желілер өшпенділік, экстремизм және қорқыту үшін негіз болды. Көп класты жіктеу осы зиянды мазмұнды дәл анықтау және санаттау үшін өте маңызды; ә) Радикалдану мен зорлық-зомбылықтың алдын алу: Ұлтшылдық мен зорлық-зомбылық экстремизміне қатысты мазмұнды уақтылы анықтау радикалдану мен зорлық-зомбылық әрекеттерінің алдын алу үшін өте маңызды. Көп класты жіктеу радикалданудың бастапқы сатыларын анықтауға және араласуға көмектеседі; б) Инклюзивтілік пен толеранттылықты

ілгерілету: Көп класты жіктеу анағұрлым инклюзивті және толерантты онлайн ортаны дамытуға ықпал етеді. Нәсілшілдікке қатысты мазмұнды санаттау және шешу арқылы платформалар қауіпсіздік сезімін және әртүрлі пайдаланушыларға құрмет сезімін оята алады; в) психикалық денсаулық пен әл-ауқатқа қолдау көрсету: қорлау мазмұнын анықтау және санаттау адамдардың психикалық денсаулығы мен әл-ауқатын қорғау үшін, әсіресе балалар мен жасөспірімдер сияқты халықтың осал топтары арасында өте маңызды.

Көп класты жіктеу бірнеше себептер бойынша ұлтшылдық, нәсілшілдік, зорлық-зомбылық экстремизмі, буллинг сияқты осал мәселелерді анықтауда және шешуде шешуші рөл атқарады. Көп классты жіктеу мәтіннің мазмұнын неғұрлым терең және дұрыс түсінуге мүмкіндік береді. Мазмұнды жай ғана «зиянды» немесе «зиянды емес» деп белгілеудің орнына, ол мәтінді ұлтшылдық, нәсілшілдік, зорлық-зомбылық экстремизм немесе қорлау сияқты нақты санаттарға бөлуге мүмкіндік береді.

Онлайн платформалар мен әлеуметтік медиа желілері өшпенділік сөздерін, экстремизмді немесе қорлауды насихаттайтын мәтінді анықтау және реттеу үшін көп классты жіктеуді пайдалана алады. Мазмұнды әртүрлі санаттарға жіктеу арқылы платформалар ескертулер, жоюлар немесе есеп беру механизмдері сияқты тиісті модерация шараларын қолдана алады.

Зиянды мазмұны бар мәтінді ерте анықтау алдын алу және ерте араласу үшін өте маңызды. Көп класты жіктеу үлгілері білікке, тәрбиешілерге немесе ата-аналарға ықтимал мәселелер көтерілмес бұрын ескертіп, белсенді шараларды қабылдауға мүмкіндік береді.

Қорытындылай келе, ұлтшылдыққа, нәсілшілдікке, зорлық-зомбылық экстремизм-ге және буллингке көп классты жіктеу осал және зиянды мазмұнды мәтіндердің мәселесін шешудің қуатты құралы болып табылады. Ол платформаларға, билік органдарына, зерттеушілерге және қауымдастықтарға мақсатты әрекеттер жасауға, хабардарлықты арттыруға және жеке тұлғалар мен қоғамды осындай мазмұнның теріс әсерінен қорғауға мүмкіндік береді.

Ұлтшылдыққа, нәсілшілдікке, зорлық-зомбылық экстремизміне және буллингке көп классты жіктеудің мақсаты - мәтіндік деректерді осы нақты және осал санаттарға автоматты түрде жіктеп, жіктей алатын машиналық оқыту үлгілері мен алгоритмдерін әзірлеу. Бұл мақсат бірнеше мақсатқа қызмет етеді: ұлтшылдықты, нәсілшілдікті, зорлық-зомбылық экстремизмді немесе қорлауды насихаттайтын мәтінді анықтау және белгілеу арқылы онлайн платформалар мен әлеуметтік желілердегі мәтінді тиімді реттеу. Жеке адамдардың экстремистік идеологияларға радикалдануы немесе өшпенділік сөздерінің таралуы сияқты зиянды мінез-құлықтың бастапқы кезеңдерінде анықтау және араласу. Пайдаланушыларды, әсіресе балалар мен осал топтарды психологиялық, эмоционалдық немесе физикалық зиян келтіруі мүмкін зиянды мазмұнды мәтіннің әсерінен қорғау. Интернеттегі экстремизмді, нәсілшілдік пен қорлауды зерттеу үшін зерттеушілерді санатталған деректермен қамтамасыз ету, осы құбылыстарды тереңірек түсінуге көмектесу.

Берілген жұмыстың міндеттері:

- төрт түрлі бағыттағы (нәсілшілдік, ұлтшылдық, буллинг, зорлық-зомбылық экстремизм) деректерден тұратын корпус құру;
- корпусты алдын-ала өңдеуден өткізу; таңбалау және тоқтату сөздерін жою арқылы мәтіндік деректерді алдын ала өңдеу. Мәтінді көрсету үшін Word2Vec немесе GloVe сияқты сөздерді ендіруді пайдалануды қарастыру;
- психо-эмоционалдық талдауға арналған сөздік құру;

- өңделген мәтінді оқытып, қолданушының психологиялық портретін график бойынша көрсету.
- Аталған бағыттағы мәліметтер қорына машиналық оқыту алгоритмдерін үйрету және жұмыстан әртүрлі көрсеткіш бойынша нәтиже алу.

Бағалау: үлгі өнімділігін дәлдік, еске түсіру, F1 ұпайы және шатастыру матрицалары сияқты көрсеткіштер арқылы бағалау қажет болады. Оның ұлтшылдыққа, нәсілшілдікке, буллингке, зорлық-зомылық экстремизмге қатысты мазмұнды дәл ажырата алатынына көз жеткізу.

Әдебиеттерге шолу

Интернеттегі ақпарат санының өсуімен және әлеуметтік желілердің дамуымен пайдаланушылар ақпарат алмасуға мүмкіндік алды. Күнделікті әлеуметтік желі көмегімен қолданушы көптеген жазбалар мен пікірлер қалдырады. Сол пікірлердің ішінде қоғамға қауіп туғызатын жазбаларда кездеседі.

Ұлтшылдық 1930 және 1940 жылдардағы Германиядағы Адольф Гитлердің ұлтшылдық партиясымен байланысты идеология мен саясатты білдіреді. Ол шектен шыққан ұлтшылдық, нәсілшілдік, антисемитизм және авторитаризммен сипатталады. Ұлтшылдық режим Холокостқа әкелді, оның барысында миллиондаған адамдар, ең алдымен, еврейлер концлагерьлерде жүйелі түрде жойылды.

Буллинг - бұл жеке адамға немесе адамдар тобына бағытталған қайталанатын, қасақана және зиянды мінез-құлық. Ол ауызша, физикалық немесе онлайн қудалауды қоса алғанда, көптеген нысандарда болуы мүмкін. Қорқыту жәбірленушілер үшін ауыр зардаптарға әкелуі мүмкін, бұл физикалық және эмоционалдық зиян келтіреді.

Нәсілшілдік - нәсіліне немесе этникалық белгілеріне қарай жеке тұлғаларға немесе топтарға қатысты кемсітушілікпен немесе алалаушылықпен сипатталатын терең қалыптасқан сенім жүйесі және әлеуметтік мәселе. Нәсілшілдікпен күресу және нәсілдік теңдік пен әділеттілікті насихаттау бойынша күш жалғасуда. [1]

Зорлық-зомбылық экстремизм идеологиялық, діни немесе саяси мақсаттарға жету үшін күш қолдануды білдіреді. Ол терроризм немесе экстремистік зорлық-зомбылық сияқты әртүрлі нысандарда көрінуі мүмкін. Зорлық-зомбылық экстремизміне қарсы тұру радикалданудың алдын алу және экстремизмге әкеп соғуы мүмкін шағымдарды қараудың зорлықсыз балама нұсқаларын насихаттауды қамтиды.

Экстремистік мәтіндердің психологиялық аспектілерін анықтау экстремизмге қарсы тұрудың біртұтас көзқарасы үшін маңызды. Ол алдын алу, араласу және оңалту стратегиялары туралы ақпарат береді, тиімді қарсы әңгімелер әзірлеуге көмектеседі және радикалдану мен экстремистік мінез-құлыққа байланысты күрделі факторларды кеңірек түсінуге ықпал етеді [2].

[3] зерттеу жұмысы «Hate speech анықтауға арналған иерархиялық көп таңбалы жіктеу әдісі» иерархиялық көп таңбалы жіктеу тәсілін пайдалана отырып, желілік платформаларда өшпенділік сөздерін анықтау әдісін ұсынады. Авторлар бір мәтін үлгісі бір уақытта бірнеше өшпенділік санаттарына жататын бірнеше белгі классификациясын өңдеу тәсілдерін талқылай алады. Жұмыста шешім ағаштары, ансамбль әдістері немесе нейрондық желілер сияқты иерархиялық көп таңбалы жіктеу тапсырмасы үшін қолданылатын машиналық оқыту үлгілері немесе алгоритмдері сипатталуы мүмкін. Авторлар оқыту және бағалау үшін пайдаланылатын деректер жиынын, сондай-ақ деректерге қолданылатын кез келген алдын ала өңдеу қадамдарын көрсете алады.

[4] жұмыс табиғи тілді өңдеу (NLP) саласына маңызды үлес болып табылады және контексттегі нәсілдік, нәсілшілдік және нәсілшілдікке қарсы сыни мәселені қарастырады.

[5] желідегі жек көрушілікті анықтау алгоритмдерінің дәлдігін арттыру үшін динамикалық түрде жасалған деректер жиынын құруға бағытталған.

[6] зерттеу жұмысы көп класты белгілермен көп идеологиялық экстремизм туралы бастапқы деректер жинағын және теңдестірілген мәтіндік деректер жинағын әзірлеуді ұсынады. Терең оқыту арқылы табиғи тілді өңдеу соңғы уақытта экстремизмді анықтауға үлкен көңіл бөлді. зерттеу. Осылайша, бұл зерттеу экстремистік твиттерді жинауға, тазалауға және жіктеуге бағытталған. Бұл тренинг көптеген идеологиялары бар теңдестірілген деректер жиынында көп сыныпты жіктеуді енгізеді. Бұл деректер жиынтығы біріктірілген деп аталады.

MIWS деректер жинағы алдын ала дайындалған BERT және оның нұсқалары арқылы бағаланады. RoBERTa және DistilBERT сияқты үлгілер ең жоғары f1 - 0,72 жетеді. RoBERTa және DistilBERT f1- 0,68 және 0,71 пайызды көрсетеді.

Зерттеу әдістері

Әр санаттың таңбаланған мысалдары бар жоғары сапалы деректер жинағын құру көп класты жіктеу зерттеулері үшін өте маңызды. Бұл мәтіндік құжаттарды қолмен белгілеуді немесе сәйкес мазмұнды қамтитын бар деректер жиынын пайдалануды қамтуы мүмкін. Айтқанымыздай, аталған жұмысты жүзеге асыру үшін үлкен көлемді корпус қажет болады. Ол жинақталған корпус машиналық оқыту алгоритмдерін үйрету және тестілеу кезеңдерінде қолданылады. Ашық қолжетімді ресурстардан аталған төрт топ бойынша: ұлтшылдық, нәсілшілдік, буллинг және зорлық-зомбылық эсткремизмі бағытындағы мәтіндер іздестірілді. Көбіне Youtube, VK, Telegram сынды әлеуметтік желілердегі жазылған түрлі комментарийлер арасынан табылады. Бірақ бұл қолмен терілген мәтіндер саны жеткіліксіз болды. Деректерді жинау мақсатында парсинг қолдануға жүгіндік.

Жұмыс мақсаты ретінде киберқауіпсіздікті қамтамасыз ету үшін әлеуметтік желі қолданушысының психологиялық портретін құру болғандықтан, деректер негізі ретінде VK,Telegram, Youtube, парсинг негізінде жиналған деректер алынды (1-2-сурет).

1	Сіз таланттарды дамытуыңыз керек
1	Жақсы күндерді керемет етуге тырысыңыз және сол күндерден жағымды нәрсе алыңыз
1	Позитивті адам болыңыз және алға қарай жүріңіз.
1	Бұл әлемде іс жүзінде мүмкін емес ештеңе жоқ
1	Бір сәтке проблемалардан аулақ болыңыз және оң мүмкіндіктерге назар аударыңыз.
1	Іс-шара қарсаңында жеңімпаздар өздерінің оң үміттерін білдіретін әдетке ие.
1	Мен шиеленісті босатып, өмірге сенемін, Мен оны жақсы көремін;
1	Мен аяғыма мықтап және сенімді тұрамын;
1	Бейбітшілік пен адамдар маған мейірімді;
1	Мен қойылған міндеттерді оңай шешемін;
1	Әлем мол, оның көптеген мүмкіндіктері, ресурстары бар;
1	Мен өмірден бәрін батыл аламын;
1	Менің қанымда өмірлік күш көп!
1	Күн сайын мен сау және сау боламын;
1	Мен барлық Құдайға ризамын (өмір, ғалам, табиғат-кім сенеді);
1	Мен бақыттымын, сәттілік мені жақсы көреді;
1	Мен денемді жақсы көремін және әр органға оның жұмысы үшін ризамын;
1	Барлығына бақыт пен жақсылық тілеймін. Мен бай және жомарттықпын;

1-сурет. Негативті пікірлер

label	message
0	Қазақстанды бұзып жатырғандар: Ерсін Әміре, Абдуғаппар, Арман Қуанышбай, Муфтиат
0	Таліптер Алға Дінсіздерді Алланың парыз болғандарынан жүз өгіргендерді құрту керек
0	Ату керек тамырымен жойылғандай!
0	соғыс болса болсын,жемқор йттерді қуып журіп өлтіреміз,
0	Талибтердің үстінен бомба тастап жок кылып жіберетінго. Бул тек ауганстанга емес, тыныш отырган мусылман еліне тонген кауіп деп білемін.
0	Көктемде аурулар қозады ғой орыстар ауру халыққой. Бәле болды осы ауру параша;
0	Мен Оз елим,Оз жерим ушин барине дайынмын,олугеде,олтиругеде,Маскеу тубине барып жарып жиберугеде,Атым Казак
0	Орыс тілі жойылса ғой еліміз де
0	Қазақстандағы орыстарды қуу керек
0	Жабайылар жабайылыгын коррсттигой хайуан надандар биздин арамтамактардын кесириненгой...;
0	Кырткан екенсндер орыстар аирамыз кетндер шекара жаблсын Рассия нын карап отрмаспыз шошкалар
0	ҚАЗАҚТАР ОЯНЫҢДАР ЖАУ БҰЛАР АЯМАҢДАР ІШТЕГІ ТЫҢШЫЛАР БҰЛАР!
0	Америка жасасын өлсін ресей
0	Онбаган орыстар

2-сурет. Позитивті пікірлер

Нәтижелер

Мәтінге психо-эмоционалдық талдау жасаудағы ең алғашқы қадам сөздер категориясын құру болды. Екінші қадамда құрылған категориялар бойынша сөздерді тізімге енгізу жүзеге асырылды. Осы қадамнан кейін сөздерге анализ жасалып, қазақ тіліндегі уникальді сөздер қосылып, сөздер жиілігі тексерілді (3-сурет).

```
%
1  funct
2  pronoun
3  ppron
4  i
5  we
6  you
7  shehe
8  they
9  ipron
10   article
11 verb
12   auxverb
13   past
14   present
15   future
16   adverb
17   preps
18   conj
19   negate
20   quant
21   number
22   swear
121  social
122  family
137  inhib
138  incl
139  excl
140  percept
141  see
142  hear
143  feel
146  bio
147  body
148  health
149  sexual
150  ingest
250  relativ
251  motion
252  space
253  time
354  work
355  achieve
356  leisure
357  home
358  money
359  relig
360  death
462  assent
463  nonfl
464  filler
%
аяқта  354  355
шот 358
дәлдік 131  136
тұз  355
ауырсын 125  127  130  146  148
жеткіз 354  355
ауырсыну   125  127  130  146  148
қышқыл  140
мойында 131  132
безеу  146  148
таныс  121  123
иелен  355
өткір   140
бойымен 1  17   252  250
әрекеттен  250  251
әрекет251  250
белсендір   131  133
белсенді 125  126
актер  356
актриса  356
шынында   1  16
қос  131  138
нашақор 146  148
қоспа  131  138
мекенжай  357
байсалды  355
```

3-сурет. Мәтінге психо-эмоционалдық талдау сөздігі

Жалпы сөздікте 65 санат берілген және санаттар " % " таңбасының арасында орналасқан. Сөздік екі бөлімнен тұрады. Бірінші бөлім, екі " % " таңбасының арасында қай категориялардың құрылатынын анықтайды. Екінші бөлім " % "-дан кейінгі екінші жартысында қандай сөздерді санау керек және оларды қандай санаттарға қосу керектігін

анықтайды. Жоғарғы жартысында санат анықтамалары әрқашан саннан басталады, содан кейін санат атауы жазылады. Санат атауы тек бір сөзден (екі немесе одан да көп сөзден емес) тұруы мүмкін. Санат атауы санат нөміріне сәйкес келеді. Атаулар немесе сөздер бас әріппен жазылған ба, жоқ па маңызды емес. Себебі аталған жүйе регистрге сезімтал емес болып табылады. LIWC модулі орнатылғаннан кейін, функцияларды қолдану арқылы файлды талдау жүргізілді (4-сурет).

	message	stemming_message	label
0	Қазақстанды бұзып жатырғандар: Ерсін Әміре, Аб...	қазақ бұз жатыр ер әмір абдұғаппар арман қуаны...	0
1	Таліптер Алға Дінсіздерді Алланың парыз болған...	тал ал дін алла парыз бол жүз өгіргендерді құр...	0
2	Ату керек тамырымен жойылғандай!	ату керек тамыр жойылған	0
3	соғыс болса болсын,жемқор йттерді қуып журіп ө...	соғыс бол бол жемқор йттерді қу жу өлтіреміз	0
4	Талибтердің үстінен бомба тастап жок кылып жіб...	тал үс бомба таста жо кылып жіберетінго бу тек...	0
...	...	...	...

4-сурет. Тазарту және өңдеу сатысынан өткен мәліметтер қоры

.csv файлдағы деректерді жүктеу үшін pandas модулінің read_csv әдісі қолданылды.Жоғарыда аталған кітапханаларды жүктеп, plt әдісін қолдану арқылы файл мазмұны экранға шығарылды. Stemming_message файлындағы label-ы 1-ге тең деректер оқытылды. Файлды оқуға 7.54 миллисекунд жұмсалды (5-сурет)

```
[7] categories_series = liwc_df.sum().sort_values()
    categories_series = categories_series/categories_series.sum()
    categories_series

    filler     0.000000
    assent     0.000000
    death      0.000000
    ingest     0.000000
    we         0.000000
                 ...
    auxverb    0.051685
    affect     0.052809
    pronoun    0.061798
    verb       0.069663
    funct      0.137079
    Length: 64, dtype: float64
```

5-сурет. Мәтінге психо-эмоционалдық талдау жүргізу

Категориялар сериясын,категориялар суммасына бөліп жалпы сөздердің кездесу жиілігі процент бойынша көрсетіледі.Көмекші етістіктер 0.051685, әсер ету 0.052809, есімдік 0.061798, етістік 0.069663, функционалды сөздер 0.137079 процентті құрайды.

Одан бөлек етістік, есімдік, позитивті сөздер, үстеу, әлеуметтік, биологиялық, уақыт категориялары жиі қолданылған. label 1-ге тең деректерде негативті бағытты сипаттайтын бірде-бір категория кездеспеген. Осыған сәйкес бұл деректерді қалдырушылардың психологиялық портреті позитивті типке сәйкес келеді. Stemming_message файлындағы label-ы 0-ге тең деректерді оқытып шықтық.Pd.DataFrame арқылы категориялар(баған) бойынша датафрейм құрамыз.

Категориялар сериясын, категориялар сомасына бөліп, жалпы сөздердің кездесу жиілігі пайыз бойынша көрсетілген. Әсер ету категориясы 0.046278, етістік 0.0623746,

когнетивті механизм 0.080423, функционалды сөздер 0.094567 пайызды құрайды. Пайыздық көрсеткіш бойынша функционалды сөздер жиі кездесетінін байқауға болады.

Қорытынды

Әлеуметтік желі қазіргі таңда адам өмірінің ажырамас бөлігіне айналды. Әлеуметтік желіні адам өміріне әкелетін пайдасы мен зияны қатар жүреді. Әлеуметтік желі көмегімен зиянкестер, киберқылмыскерлер қоғамға, адам өміріне орасан қауіп келтіре алады. Осындай зиянкестерді анықтау құралы ретінде психологиялық портрет құру және көп классты жіктеуді келтіре аламыз. Берілген жұмыс барысында мақсатқа жету үшін келесі міндеттер орындалды: төрт түрлі бағыттағы (нәсілшілдік, ұлтшылдық, буллинг, зорлық-зомбылық экстремизм) деректерден тұратын корпус құрылды; корпус алдын-ала өңдеуден өткізілді; мәтінді психо-эмоционалдық талдауға арналған сөздік құрылды, өңделген мәтінді оқытып, қолданушының психологиялық портреті график бойынша көрсетілді, аталған бағыттағы мәліметтер қорына машиналық оқыту алгоритмдерін үйрету және жұмыстан әртүрлі көрсеткіш бойынша нәтиже алу жұмыстары жүргізілді.

Алдағы уақытта мәліметтер қорын бұдан да кеңейту және өзге де машиналық оқыту алгоритмдерін қолданып зерттеулер жүргізу жоспарлануда.

Берілген зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті қаржыландыратын “Жастар экстремизмін анықтау және заманауи ақпараттық кеңістікте жастардың қауіпсіздігін қамтамасыз етуге арналған модельдер мен әдістерді әзірлеу” жобасы аясында орындалды (грант AP19576868, жоба жетекшісі Болатбек М.А.).

Пайдаланылған әдебиеттер тізімі

1. https://ru.wikipedia.org/ - (18.09.23)
2. “Психологические аспекты экстремизма и терроризма”. Д.С.Безносов, Л.Г.Почебут- https://cyberleninka.ru/article/n/psihologicheskie-aspekty-ekstremizma-i-terrorizma/viewer (20.09.23).
3. “A Hierarchical Multi-label Classification Method for Hate Speech Detection”. Jie Zhou, Chuan Luo, Xin Wang, Tian Lan- (20.09.23).
4. “A survey of race , racism, and anti-racism in NLP”. A.Field, SL Blodgett, Z.Waseem, Y.Tsvetkov - https://arxiv.org/pdf/2106.11410.pdf (21.09.23).
5. “Learning from the worst:Dynamically generated datasets to improve online hate detection”. V.Vidgen, T.Thrush, Z.Waseem, D.Kiela - https://arxiv.org/pdf/2012.15761.pdf (22.09.23)
6. “Multi-ideology Multi-class Extremism Classification using Deep Learning Techniques”. Mayur Gaikwad, Swati Ahirrao , Ketan Kotecha , Ajith Abraham. - https://www.researchgate.net/publication/363514996_Multi-ideology_Multi-class_Extremism_Classification_using_Deep_Learning_Techniques (22.09.23)

INNOVATIVE APPROACHES TO PROTECTING DISTRIBUTED COMPUTER SYSTEMS AND INTERNET SERVERS: USING ACTIVE METHODS

[1]Daukenov N.B., [1,2]Tereikovskyi I.A.
e-mail: nb.daukenov@gmail.com
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Annotation. *The article examines innovative approaches to ensuring the security of distributed computer systems (DCS) and Internet servers. Focuses on the use of active security techniques, software server efficiency, and enterprise website security issues. The challenges and requirements in the field of cybersecurity are analyzed, and approaches to improving the protection of digital communication systems are proposed.*

Introduction

In the modern world, one of the predominant, extremely pressing problems in the field of computer security is the protection of distributed computer systems (DCS) of enterprises actively integrated into the global Internet [1,2]. Integration of such systems with the Internet is usually carried out through specialized software servers that provide remote access to certain resources. Examples of such servers include web servers that ensure continuous operation of enterprise websites, as well as e-mail servers and FTP servers that are widely used to ensure the interaction of digital communication systems with the Internet. Despite a significant number of modern solutions presented on the software market in this area, as well as extensive scientific and practical research into the security problems of Internet servers, we are increasingly faced with successful attacks on DCS, which indicates an insufficient degree of their security.

Statement of the problem and its scientific and practical significance

In the context of the modern information space, where DCSs are closely integrated with the global Internet, ensuring the security of Internet servers becomes a priority. The frequency and complexity of network attacks are constantly increasing, which requires constant improvement of protection methods.

The importance of this problem is manifested in the fact that successful attacks on Internet servers can lead to serious consequences, such as leakage of confidential information, violation of data integrity and, in extreme cases, lead to downtime of key services of enterprises and organizations.

Analysis of recent research and publications in the context of improving the security system of Internet servers in the DCS

To improve the security of an Internet server, it is necessary to analyze its functionality, potential attacks and use existing security measures, taking into account the nature of the functions performed.

Source [3] defines the structure of possible attacks on the Internet. In the context of an Internet server, the following types of attacks are of interest:

1. Denial of Service (DoS): Aimed at overloading a server, making it unavailable to legitimate users.
2. Substitution of a trusted object (subject) of the DCS: Attackers can replace a legitimate object or subject of the system, which can lead to serious consequences.
3. Injection of a false object: The purpose of such attacks is to introduce an imaginary object into the DCS in order to violate the confidentiality or integrity of information posted on the Internet server.

The mechanism of a denial of service attack is rooted in the limited capabilities of the network operating system that controls the functioning of the Internet server. This system can handle only a limited number of active virtual connections and respond to a limited volume of requests. These limits are subject to various factors, including the processing power of the system, the amount of RAM, and the data transfer speed through the communication channel. Increasing the values of these parameters helps to increase the number of processed requests per unit of time.

It is noted that this study does not consider denial of service attacks on the Internet server, which can be carried out from within a distributed computer system (DCS).

The work [4] discusses three typical attack mechanisms for denial of service on an Internet server. The first type of attack involves transmitting to the Internet server such a volume of requests from one address that it overloads its capabilities (the so-called “request storm”). The results reported in [1] examine the effectiveness of such attacks on different types of web servers and operating systems. Analysis of the results indicates the limited effectiveness of attacks with high-quality server administration and the use of powerful equipment. The second type of attack is the transmission of a specially selected request to the attacked object, which, if there are errors in the software of the Internet server or network operating system, can cause a loop in the request processing procedure and, consequently, lead to a system freeze. Findings from [1,4] indicate the low success rate of such attacks when using modern versions of Internet servers, operating systems and standard security methods. The third type of attack is based on the authorized use of server software applications, allowing anonymous use of network operating system resources and computer computing capabilities. The consequences of this attack may include exhausting the server's computing resources, which may result in limited functionality or complete shutdown. This type of attack is considered the most dangerous today. The work [4] proposes to evaluate the effectiveness of this type of attack by linking the amount of computing resources spent by the attacker with the loss of the server’s computing capabilities. In this case, the loss of computing capabilities is assessed for the attacker using the load on the central processor during the attack and the duration of the attack itself, and for the server - the duration of work with such a load at which access to the served resources is blocked.

Two metrics have been proposed to evaluate denial of service attacks: attack power (how many resources the attacker uses) and blocking effectiveness (how successful the attack is in blocking access). To protect the server, protection power (the protective resources used) and anti-blocking effectiveness (how well the protection prevents access to resources from being blocked) are allocated.

The attack power (A) is calculated as the ratio of the load on the central processor of the attacked server (B) to the total load on the attackers’ central processors during the attack (C):

$$A = B/C \tag{1}$$

Blocking efficiency (E) is defined as the ratio of the time during which the server operates at 100% load (β) to the duration of the attack (α):

$$E = L/D \tag{2}$$

The protection power (O) is determined as follows:

$$O = 1/A \tag{3}$$

The effectiveness of blocking protection (F) is defined as the ratio of the server operating time without blocking access to the duration of the attack:

$$F = 1/E \tag{4}$$

Analysis of sources [3,4] indicates the use of current methods of protecting Internet servers from denial of service attacks:

- Improvement of network operating system and Internet server software: Aimed at eliminating software errors and ensuring reliability.
- Improvement of administration technology: Includes the development of effective methods for monitoring and managing the server.
- Use of powerful hardware: Provides processing of large volume of requests and reduces vulnerability to attacks.
- Use of backup equipment with load balancing: Ensures continuous server operation through switching to backup resources and load balancing.

Reflecting massive attacks of the first and third types causes significant economic costs for excess hardware reserve capacity. At the same time, preparing and executing such attacks require significant resources, which limits their implementation to a time frame.

Attacks aimed at violating the confidentiality and integrity of information on an Internet server include:

- Hacking server software.
- Hacking passwords to access data on the server.
- Exploiting vulnerabilities in communication protocols for false authentication.
- Exploiting errors in server software and operating system.

To protect the Internet server from attacks, including hacking and violation of confidentiality, the following protection system is used:

- Firewalls: Limit access to services and manage connections to external sources.
- Proper configuration and administration: Covers setting up the Internet server and operating system, maintaining security logs.
- Attack detection systems: Used to identify and prevent possible attacks on the server.
- Vulnerability Blocking Software: Prevents exploitation of vulnerabilities in the Internet server and applications.
- Use of secure protocols.

Implementing the measures described to enhance Internet server security against attacks that violate confidentiality or data integrity may conflict with the effectiveness of protection against denial of service attacks (1–4). This is due to the consumption of significant computing resources by the protection system. For example, time-delay techniques to protect against password guessing can be abused, limiting system availability. The use of sophisticated,

computationally intensive attack detection systems can also increase the effectiveness of denial of service attacks.

Presentation of the main research material with full justification of the scientific results obtained

The study highlights that successful attacks on data privacy on an Internet server often depend on the attacker's knowledge of certain parameters. In light of this, it is proposed to improve existing protection systems by introducing an “active component”. This element involves sending false responses, requiring additional computational effort and time from the attacker.

As part of this approach, two methods for generating false responses are proposed. The first method involves providing the attacker with a fake resource that is difficult to recognize automatically. The second method involves sending an application that hinders the attacker's work or even causes potential harm to his system. These actions are aimed at exhausting the attacker's computing resources, such as loading the central processor or RAM.

This innovative approach offers a new mechanism to improve the effectiveness of Internet server security systems, actively resisting attempts to violate data confidentiality.

Additionally, even if the system is functioning normally, it seems advisable to change the standard response, for example, to “access denied.” This method also has a negative impact on the effectiveness of a potential attack. A diagram demonstrating the active component of the protection system is presented in Figure 1.

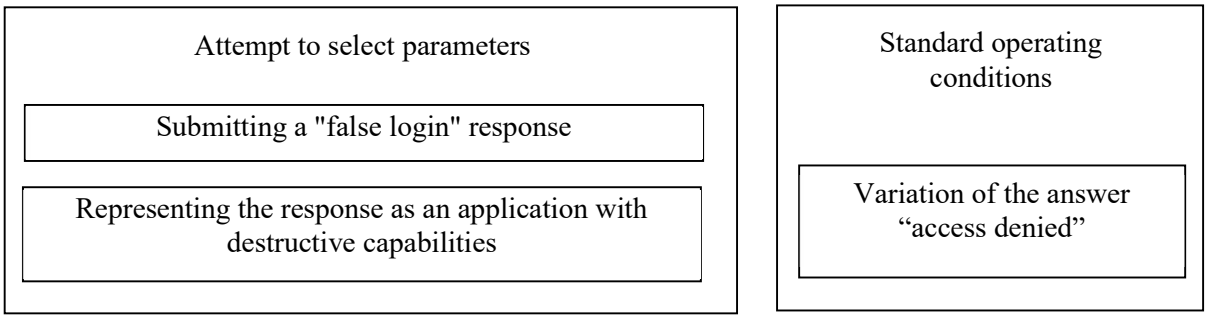


Fig. 1 Diagram of the active component of the Internet server protection system.

The presented figure shows the structure of the active component of the protection system, which includes variations in typical responses. This approach is aimed at improving the effectiveness of protection against potential attacks on the Internet server, even under normal system operation conditions.

The components of the active component, in accordance with suggestions from [4], are optimally stored in the form of small “static” files, such as TXT or XML. As noted in this study, transferring such files does not create a significant load on the server. Therefore, emulating processes such as password guessing that initiates the transfer of said files does not significantly increase the effectiveness of a denial of service attack.

The effectiveness of the active component is assessed by increasing the time required for an attacker to analyze the response from the Internet server, as well as increasing the amount of computing resources required to send requests. Similarly (1-4), the following indicators can be used:

$$\sigma = \frac{\tau_1+\tau_2}{\tau_1}, \quad (5)$$

where (σ) is a decrease in blocking efficiency, (τ_1) is the time required by an attacker to analyze the response in the absence of an active component, (τ_2) is the time required by an attacker to analyze the response when using an active component.

$$\Psi = \frac{\phi_1+\phi_2}{\phi_1}, \quad (6)$$

where (Ψ) is a decrease in the effectiveness of the attack, (ϕ_1) is the amount of computing resources required by an attacker for an attack in the absence of an active component, (ϕ_2) is the amount of computing resources required by an attacker to analyze the response when using an active component.

Conclusions from this study

Based on the research, a concept for active protection of the Internet server was developed, specially adapted to withstand attacks aimed at both denial of service and violation of confidentiality and data integrity. The main element of the proposed concept is the initiation of data transfer to the attacker's computer, the information of which creates significant difficulties for the successful implementation of an attack. Reducing the effectiveness of attacks is achieved by exploiting vulnerabilities in hacker software, generating additional load on the attacker's resources and increasing the time required to analyze the information received from the server. The work also proposes performance indicators (5,6) designed to assess the functionality of the active protection system.

Literature

1. Smith D. "Detection of attacks in computer networks" Publishing house "Technosphere", 2005. 512 p.
2. Johnson, Michael, and Richard Smith. "Classification and analysis of systems and methods for detecting attacks" Magazine "Information Security" No. 3, Ed. National Aviation University, 2006. 12–28 p.
3. Brovko I. "Mechanisms of possible attacks on the Internet " Magazine "Information Security" No. 2, National Aviation University, 2002. 45–58 p.
4. Simonov A. "Protection of Web sites of corporate information systems from interference attacks" Coll. scientific works "Information technologies in Ukraine", 2005, Issue No. 5, 220–235 p.
5. Litvinenko P. "Using the capabilities of Microsoft Word when creating web-based viruses" Magazine "Information Security" No. 9, National Aviation University, 2005. 112–125 p.

О СТАНДАРТИЗАЦИИ КРИПТОГРАФИЧЕСКИХ ПОКАЗАТЕЛЕЙ БЕЗОПАСНОСТИ СКЗИ

Оспанов Ж.Ж.[1], Горлов Л.В.[1], Ибраев Р.Б.[2]
e-mail: tbegeldin@gmail.com
[1] *КазНУ им. Аль-Фараби, Казахстан, г. Алматы*
[2] *SatbayevUniversity, Казахстан, г. Алматы*

Аннотация. *В данной работе анализируются отдельные нормы текущей редакции национального стандарта Республики Казахстан СТ РК 1073-2007 «Средства криптографической защиты информации. Общие требования», выявляются устаревшие требования, анализируются тенденции развития вычислительной техники и связанные с этим угрозы безопасности информации.*

Предлагается внесение изменений в стандарт, которые после принятия новой редакции стандарта позволят задать высокий уровень криптографической защиты сведений, защищаемых с использованием сертифицированных СКЗИ на территории Республики Казахстан.

Введение

Для оценки надёжности криптографических алгоритмов и протоколов как правило используется нижняя граница трудоёмкости наиболее успешной криптографической атаки, способной привести к частичному или полному раскрытию значения ключа шифрования или построению преобразования, эквивалентного расшифрованию, без знания ключа. В случае с СКЗИ возможно применение аналогичного подхода, при этом стойкость криптографической защиты будет определяться стойкостью наименее надёжного её узла. Следуя принципу «от обратного» задание минимально допустимой стойкости СКЗИ в виде эквивалента вычислительной сложности атаки означает, что на это СКЗИ, т.е. входящие в его состав отдельные алгоритмы и протоколы, не должно существовать эффективных или даже теоретических атак меньшей трудоёмкости.

Данный подход в настоящее время широко распространён, и именно он применяется для определения уровня безопасности СКЗИ согласно СТ РК 1073[1], на основании которого производится сертификация всех действующих в Республике Казахстан СКЗИ.

Вместе с тем, невозможно игнорировать постоянное развитие средств вычислительной техники и исследования в области криптоанализа, которые в совокупности приводят к устареванию принятых границ безопасности известных криптографических алгоритмов и протоколов.

В данной статье проанализирована актуальность и эффективность подхода, принятого в СТ РК 1073 и выдвинуты предложения для его усовершенствования.

1. Методы оценки безопасности СКЗИ

Эмпирическое наблюдение, известное как закон Мура [2], согласно которому количество размещаемых на кристалле транзисторов удваивается каждые два года, продолжает выполняться в 2023 году[3]. В совокупности с плотностью размещения

транзисторов на повышение производительности процессоров влияет оптимизация их работы и увеличение тактовой частоты. Таким образом, удвоение вычислительных мощностей фактически происходит ещё быстрее, примерно каждые 18 месяцев.

В соответствии с моделью нарушителя, вытекающей из СТ РК 1073[1], возможно произвести приблизительные расчёты для случаев, когда злоумышленник обладает ресурсами (прежде всего финансовыми) для атаки на перехваченные зашифрованные данные при полной осведомлённости об алгоритме шифрования.

Для этого нужно подсчитать, какую совокупную вычислительную мощность можно получить за количество ресурсов, которое целесообразно потратить на атаку. Естественно, что в реальности с одной стороны построение системы для атаки криптографической защиты требует не только финансов, но и соответствующей квалификации, а также влечёт множество побочных расходов помимо непосредственно вычислительной техники, а с другой стороны закупка больших партий оборудования и вообще налаженная промышленная система организации атак позволяют значительно оптимизировать расходы. Не имея возможности учитывать эти и другие факторы, примем в расчётах известные из открытых источников цены на аренду и закуп только основных компьютерных комплектующих и исключим из расчётов побочные расходы. Это позволит получить верхнюю оценку эффективности атаки с применением известных на данный момент средств.

Также для упрощения расчётов предположим, что атаки носят детерминированный характер, т.е. гарантированно приводят к успеху точно за определённое время.

2. Анализ развития вычислительной техники и возможностей нарушителей

Пусть требуется вскрыть ключ шифрования информации, защищённой с применением СКЗИ 4-го уровня безопасности в соответствии с СТ РК 1073. Это требует проведения как минимум 2^{160} вычислительных операций. Такая информация может оцениваться в 100 000 000 (сто миллионов) МРП или порядка одного миллиарда долларов США (оценка в долларах наиболее релевантна в реалиях рынка вычислительной техники).

При условии, что защищаемые сведения будут оставаться актуальными в течение 40 лет, а за это время вычислительные мощности вырастут приблизительно в миллиард раз, реалистичная угроза атаки возникает при наличии на данный момент вычислительных мощностей около 2^{133} степени операций в год.

Существует несколько способов получить большую совокупную вычислительную мощность, рассмотрим наиболее перспективные из них.

2.1. Суперкомпьютеры

Мощнейший на данный момент (по открытым данным) суперкомпьютер Frontier находится в США, в национальной лаборатории Oak Ridge[4]. Он был построен в 2021 году и содержит 8,730,112 процессорных ядер. Его мощность в пике составляет 1,685.65 PFlop/s, что позволяет выполнять до 2^{60} операций в секунду или 2^{85} операций в год. Этого недостаточно для атаки полным перебором на СКЗИ 4-го уровня безопасности, поскольку атака заняла бы 2^{100} секунд, что составляет 2^{75} лет.

На втором месте с пиковой мощностью 537.21 PFlop/s находится суперкомпьютер Fugaku, работающий в центре компьютерных наук Японии[5]. Порог в 200 PFlop/s на данный момент преодолевают 5 суперкомпьютеров, в том числе Summit, находящийся в той же лаборатории Oak Ridge[6].

Аренда суперкомпьютеров является весьма специфической услугой с непрозрачной тарификацией. Постройка суперкомпьютеров производится в индивидуальном порядке и может сильно отличаться по стоимости.

Находящийся на втором месте по мощности в мире суперкомпьютер Fugaku (537.21 PFlop/s) стоит около 1 миллиарда долларов[7], занимающий третье место LUMI с производительностью 428.70 PFlop/s был построен всего за 150 миллионов долларов[8]. В 2020 году за 34 миллиона долларов был построен компьютер мощностью 50 PFlop/s в Австралии[9], из чего можно предположить приблизительную цену в 1,5 миллиона долларов за 1 PFlop/s мощности.

Из вышеизложенного следует, что злоумышленник, располагающий финансированием в 1 миллиард долларов, может обладать суперкомпьютером, по мощности сравнимым с Fugaku (без учёта расходов на эксплуатацию, ремонт и электропитание). Он выполняет до $2^{83,8}$ операций в год и не представляет опасности для СКЗИ 4-го уровня.

2.2. Распределённые и облачные вычисления

Другим вариантом является использование дешёвых компьютеров, но в огромном количестве.

Мощнейшими на данный момент процессорами для персональных компьютеров являются Intel Core i9-13900 (мощностью 1 485 GFlop/s)[10] и AMD Ryzen 9 7950X (2 120 GFlop/s)[11]. На сумму в миллиард долларов можно приобрести до полутора миллионов таких процессоров, с помощью которых можно выполнять до 2^{60} операций в секунду или 2^{86} операций в год, что недостаточно для атаки на СКЗИ 4-го уровня безопасности.

Мощнейшими графическими вычислительными устройствами являются NVIDIA RTX A6000 с производительностью 75.6 TFlop/s[12]. В розницу она продаётся за 4700 долларов. Без учёта остальной инфраструктуры, электропитания и аппаратных составляющих за миллиард долларов можно получить систему из двухсот тысяч таких графических вычислительных модулей, производящую около 2^{62} операций в секунду или 2^{88} операций в год, что всё ещё далеко от производительности, необходимой для атаки на СКЗИ 4-го уровня, так как атака будет длиться 2^{71} лет.

Ещё одна опция для проведения высокопроизводительных вычислений – аренда мощности у облачных провайдеров. Наиболее крупные из них – Microsoft Azure, Amazon web service (AWS), Google Cloud.

Одна миллисекунда вычислений на AWS стоит от 0.0000000017 до 0,0000001667 USD в зависимости от требований к оперативной памяти приложения. Таким образом, за миллиард долларов можно приобрести 588 235 294 117 секунд вычислений на одном ядре Graviton 2 типа ARM производительностью около 1000 GFlop/s[13], что позволит осуществить около 2^{79} операций при условии, что Amazon сможет предоставить такие мощности в распоряжение одного заказчика.

2.3. Специализированные устройства

Распространение криптовалют придало развитие технологиям узконаправленных аппаратных вычислительных модулей. Наиболее мощными из них являются Antminer S19 Pro мощностью до 110 Th/s (триллионов хешей в секунду)[14] при стоимости устройства 2860 долларов и WhatsMiner M30S++ мощностью 112TH/s при стоимости 4000 долларов.

Если допустить, что создано аналогичное устройство для взлома алгоритмов шифрования, то злоумышленник, располагающий миллиардом долларов, может построить вычислительный кластер из 350 000 таких узлов и достичь производительности в 2^{64} операций в секунду или 2^{90} операций в год. Принципиальным отличием от предыдущих расчётов является то, что в данном случае одной операцией является полная процедура расшифрования данных, которая может состоять из нескольких тысяч элементарных операций.

Тем не менее, такой производительности также очевидно недостаточно для атаки на СКЗИ 4-го уровня безопасности.

2.4. Квантовые вычислительные мощности

Дискуссия о применимости квантовых компьютеров в задачах криптографического анализа продолжается не один десяток лет, так как началась задолго до практического запуска первого квантового компьютера. Теоретические модели и компьютерные симуляции разной степени соответствия реальности предсказывают существование вероятности эффективной на практике атаки как на асимметричные[15], так и на симметричные[16] криптосистемы.

Мощнейшим на данный момент квантовым компьютером является компьютер Osprey фирмы IBM[17], запущенный в 2022 году. Он содержит 433 кубита, а к 2025 году компания надеется построить компьютер, содержащий более 4000 кубитов. Помимо количества кубитов критичным является показатель количества операций поворота кубитов в секунду, Circuit Layer Operations Per Second (CLOPS). Этот показатель планируют увеличить с текущих 1400 до 15000. Вместе с тем идёт постоянная борьба за уменьшение коэффициента ошибок при квантовых вычислениях.

Параллельно происходит наращивание научного и практического потенциала в сфере создания алгоритмов для квантовых компьютеров и написания программного обеспечения для них.

Однако, несмотря на видимое приращение разнообразных количественных характеристик, всё ещё не являются, очевидны возможности квантовых компьютеров в области криптографического анализа, отсутствуют доказанные случаи успешного взлома тех или иных алгоритмов шифрования помимо тренировочных (созданных специально для тестирования квантовых алгоритмов). Самым близким к практике применением квантового компьютера в целях криптоанализа является взлом упрощённого DES-подобного алгоритма группой китайских учёных [16], которым при помощи вероятностного алгоритма удалось достичь производительности алгоритма Гровера[18], а в некоторых случаях и превзойти её.

2.5. Хранилища данных

По прогнозам экспертов компании Seagate[19] размер данных, хранящихся в сети Интернет, к 2025 году составит около 175 зеттабайт, то есть 2^{28} байт.

Для проведения дифференциального и линейного криптографического анализа современных алгоритмов шифрования требуется хранить огромное количество данных, сравнимое с количеством всевозможных открытых и шифрованных текстов. Т.е. при размере блока алгоритма 128 битов требуется хранить до 2^{128} блоков, а это 2^{132} байтов. Для атаки на алгоритм ГОСТ 28147-89 с размером блока 64 бита потребуется хранить около 2^{67} байтов данных, что далеко за пределами возможностей всего человечества.

2.6. Обобщение полученных данных

По аналогии можно рассчитать возможности нарушителей с бюджетом, соответствующим остальным уровням безопасности. В следующей таблице приведены данные о предположительной длительности атаки в годах.

Таблица 1.

Количество лет, необходимых для атаки СКЗИ по уровням безопасности

Уровень	Суперкомпьютеры	Распределённые вычисления	Специализированные устройства
1	2^{74}	2^{71}	2^{70}
2	2^{41}	2^{38}	2^{36}
3	2^{7}	2^{4}	2^{3}
4	2^{-15}	2^{-18}	2^{-20}

Методом аппроксимации установлено, что злоумышленник, вознамерившийся атаковать СКЗИ 1-го уровня безопасности и готовый затратить на это 100 МРП (345 000 тенге или около 750 долларов США на момент написания статьи), может приобрести или арендовать оборудование, которое позволит осуществить такой взлом за несколько десятков минут.

Действительно, за эту сумму можно приобрести процессор Intel Core i9-13900 производительностью 1485 GFlop/s, т.е. около 2^{40}операций в секунду, что всего в 1024 раза меньше вычислительной сложности атаки на СКЗИ 1-го уровня безопасности. Эта оценка согласуется с ранее приведённой в работе [20].

2.7. Предположение о сохранении стойкости СКЗИ

СКЗИ более высоких уровней безопасности на текущий момент кажутся достаточно надёжными, однако при оценке безопасности информации следует учитывать, в том числе тенденции роста вычислительных мощностей и время актуальности защищаемой информации. Например, криптографическая защита информации, которую требуется сохранить в тайне в течение 50 лет, будет потенциально атакована средствами, превосходящими ныне существующие в 2^{33} раз при сохранении тенденции удвоения каждые полтора года, т.е. после 33 удвоений, к тому же при этом неизбежно повысится эффективность методов атаки вследствие естественного развития теоретической и прикладной науки, причём этот фактор не поддаётся даже приблизительной оценке. Если допустить, что по аналогии с ростом вычислительных мощностей наука позволит спроектировать механизмы атаки, оптимизированные относительно ныне существующих в 2^{33} раз, то СКЗИ 4-го уровня будут взломаны за $2^{10} = 1024$ года, что кажется весьма солидным показателем их стойкости, а вот взлом остальных СКЗИ уже займёт часы и минуты.

Естественно, что информация, защищаемая менее серьёзными средствами, обладает не только меньшей ценностью, но и, как правило, меньшим временем актуальности. Это должно напрямую влиять на требования, предъявляемые к СКЗИ, однако в текущей версии стандарта время актуальности информации никак не учитывается.

В соответствии со статьёй 20 закона РК «О государственных секретах» [21] срок хранения информации, составляющей государственные секреты, ограничен 30 годами и в отдельных случаях может быть увеличен. Тогда примем за максимальный срок

хранения информации, защищённой с использованием СКЗИ 4-го уровня безопасности, период в 40 лет. За это время удвоение вычислительных мощностей произойдёт примерно 27 раз.

СКЗИ 1-го уровня предназначены для защиты информации, время актуальности которой составляет несколько дней, для определённости примем значение в два месяца, т.е. в 243 раза меньше. Поскольку $\sqrt[3]{243} \approx 6.24$, то примем этот значение в качестве отношения времени актуальности информации, защищаемой с применением СКЗИ соседних уровней защищённости.

Тогда можно рассчитать, с каким запасом СКЗИ соответствующего уровня безопасности будут выдерживать атаки с применением средств и методов, которые предположительно будут доступны на момент потери актуальности защищаемых сведений.

Таблица 2.

Стойкость СКЗИ к прогнозируемым атакам

Уровень	Актуальность (годы)	Количество удвоений	Существующие атаки (время в годах)	Будущие атаки (время в годах)
1	40	26.67	2^{70}	$2^{16.7}$
2	6.41	4.27	2^{36}	$2^{28.1}$
3	1.03	0.68	2^{3}	$2^{1.9}$
4	0.16	0.11	2^{-20}	$2^{-20.2}$

Таким образом, только СКЗИ 1-го уровня безопасности остаются явно уязвимыми к атакам на всё время актуальности защищаемой информации, однако и СКЗИ 2-го уровня уже находятся опасно близко к границам возможностей атакующих как сейчас, так и в будущем, что также согласуется с выводами работы [22].

3. Обновление стандарта

На основании приведённых расчётов, а также соображений, изложенных в работе[23] предлагается внести ряд изменений в текущую редакцию стандарта, а именно:

1. Привести для оценки надёжности не только вычислительную сложность лучшей атаки, но и пространственную сложность, т.е. требования к памяти и/или накопленным данным.
2. Пересмотреть суммы, определяющие зависимость требуемого уровня безопасности СКЗИ от потенциального ущерба при компрометации защищённой информации.
3. Конкретизировать список методов криптоанализа, анализ стойкости к которым должен быть проведён в обязательном порядке.
4. Внести нормы, касающиеся отдельных этапов жизненного цикла СКЗИ (разработка, изготовление, реализация, эксплуатация, утилизация).
5. Требования к наличию в эксплуатационной документации на СКЗИ регламента его эксплуатации, порядок работы с ключами и т.п.
6. Добавить требования к регистрации в СКЗИ основных событий.
7. Внести пункты, касающиеся технических каналов утечки информации.

СТ РК 1073 в новой редакции, включающей в себя изменения и дополнения, соответствующие результатам проведённых научно-исследовательских работ [22] [23] [24], будет разработана в соответствии с законодательством Республики Казахстан в сфере технического регулирования.

4. Заключение

Текущая редакция СТ РК 1073 является устаревшей в техническом отношении и в целом не охватывает некоторые критически важные аспекты безопасности, связанные с разработкой, реализацией, эксплуатацией и утилизацией СКЗИ, их учётом и контролем, техническими каналами утечки информации, особенностями лицензирования и другими.

Несколько групп отечественных специалистов в последние годы указывали на отдельные несовершенства этого стандарта, однако новая версия до сих пор не была предложена на рассмотрение широкой общественности.

Мы предлагаем внести ряд важных поправок в нормы стандарта, а также внести несколько принципиально новых требований, не противоречащих остальным частям документа и позволяющих повысить как защищённость информации, так и в целом научный и технический уровень организаций, работающих с СКЗИ на всех этапах их жизненного цикла.

Данные изменения позволят задать высокий уровень криптографической защиты сведений, защищаемых с использованием сертифицированных СКЗИ на территории Республики Казахстан.

Список литературы

1. Комитет по техническому регулированию и метрологии Министерства индустрии и торговли Республики Казахстан (Госстандарт), *Государственный стандарт Республики Казахстан "Средства криптографической защиты информации. Общие требования",* Астана, 2007.
2. Wikipedia, «Moore's law,» Wikipedia, 14 08 2023. Available: https://en.wikipedia.org/wiki/Moore%27s_law. (Дата обращения: 23 08 2023).
3. D. A. Kelleher, «Moore’s Law – Now and in the Future,» 16 02 2022. Available: https://www.intel.com/content/www/us/en/newsroom/opinion/moore-law-now-and-in-the-future.html#gs.slud9x. (Дата обращения: 23 08 2023).
4. OP500.org, « TOP500 June 2023,» TOP500.org, 07 2023. Available: https://top500.org/lists/top500/2023/06/. (Дата обращения: 23 08 2023).
5. TOP500.org, «Supercomputer Fugaku,» TOP500.org, Available: https://top500.org/system/179807/. (Дата обращения: 23 08 2023).
6. TOP400.org, «Summit,» TOP400.org, Available: https://top500.org/system/179397/. (Дата обращения: 23 08 2023).
7. Wikipedia, «Fugaku (supercomputer),» Wikipedia, 08 07 2023. Available: https://en.wikipedia.org/wiki/Fugaku_(supercomputer). (Дата обращения: 23 08 2023).
8. Wikipedia, «LUMI,» Wikipedia, 04 08 2023. Available: https://en.wikipedia.org/wiki/LUMI. (Дата обращения: 23 08 2023).
9. servernews.ru, «Суперкомпьютер Pawsey,» 19 10 2020. Available: https://servernews.ru/1023311. (Дата обращения: 23 08 2023).
10. askgeek.io, «Обзор процессора Intel Core i9-13900,» Available: https://askgeek.io/ru/cpus/Intel/Core-i9-13900. (Дата обращения: 23 08 2023).
11. gadgetversus.com, «AMD Ryzen 9 7950X Processor Benchmarks and Specs,» Available: https://gadgetversus.com/processor/amd-ryzen-9-7950x-specs/. (Дата обращения: 23 08 2023).
12. NVIDIA, «NVIDIA RTX A6000,» Available: https://www.nvidia.com/content/dam/en-zz/Solutions/design-visualization/quadro-product-literature/proviz-print-nvidia-rtx-a6000-datasheet-us-nvidia-1454980-r9-web%20(1).pdf. (Дата обращения: 23 08 2023).
13. arm.com, «Arm Performance Libraries benchmark data,» Available: https://developer.arm.com/tools-and-software/server-and-hpc/compile/arm-compiler-for-linux/resources/tutorials/benchmarks. (Дата обращения: 23 08 2023).

14. softwaretestinghelp.com, «Top 10 Bitcoin Mining Hardware,» Available: https://www.softwaretestinghelp.com/bitcoin-mining-hardware/. (Дата обращения: 23 08 2023).
15. Wikipedia, «Shor's algorithm,» Wikipedia, 18 08 2023. Available: https://en.wikipedia.org/wiki/Shor%27s_algorithm. (Дата обращения: 23 08 2023).
16. ZeGuo Wang, ShiJie Wei, Gui-Lu Long, Lajos Hanzo, «A Variational Quantum Attack for AES-like Symmetric Cryptography,» *https://arxiv.org/pdf/2205.03529.pdf,* 2022.
17. IBM, «IBM Quantum's mission is to bring useful quantum computing to the world.,» 2022. Available: https://research.ibm.com/blog/next-wave-quantum-centric-supercomputing. (Дата обращения: 23 08 2023).
18. Wikipedia, «Grover's algorithm,» 07 08 2023. Available: https://en.wikipedia.org/wiki/Grover%27s_algorithm. (Дата обращения: 23 08 2023)
19. David Reinsel, John Gantz, John Rydning, «The Digitization of the World. From Edge to Core,» 2018. Available: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf. (Датаобращения: 23 08 2023).
20. А. Абдрахманов, «Модели нарушителей криптографической защиты и стандарт СТ РК 1073-2007,» *Доклады национальной академии наук Республики Казахстан,* т. 6, pp. 62-71, 2017.
21. Р. Казахстан, «Закон Республики Казахстан "О государственных секретах",» 07 03 2023. Available: https://online.zakon.kz/Document/?doc_id=1012633. (Дата обращения: 24 08 2023).
22. А. Абдрахманов, «Модели нарушителей криптографической защиты и стандарт СТ РК 1073-2007,» *Доклады Национальной академии наук Республики Казахстан,* т. 6, pp. 62-71, 2017.
23. A.E. Abdrakhmanov, G.T. Turdiyeva, «Safe thresholds for the parameters of cryptographic algorithms and standard ST RK 1073-2007,» *News of the National Academy of Sciences of the Republic of Kazakhstan Physico-Mathematical series,* т. 5, № 333, p. 76 – 85, 2020.
24. Р.Б. Ибраев, Л.В. Горлов, О.Э. Возный, А.М. Кунгожин, «Анализ полноты требований государственного стандарта Республики Казахстан СТ 1073-2007 "Средства криптографической защиты информации. Общие технические требования",» *Сатпаевские чтения,* т. 2, pp. 374-378, 2020.

О ПОДХОДАХ К УСТАНОВЛЕНИЮ НОРМАТИВНЫХ ТРЕБОВАНИЙ К ПРОЦЕССАМ ЖИЗНЕННОГО ЦИКЛА СКЗИ

Ибраев Р.Б.[1], Горлов Л.В.[2]
e-mail: r.ibrayev@satbayev.university
[1] *Satbayev University, Казахстан, г. Алматы,*
[2] *КазНУ им. Аль-Фараби, Казахстан, г. Алматы*

Аннотация. *В данной работе рассматриваются роль и место Государственного стандарта Республики Казахстан СТ РК 1073-2007 «Средства криптографической защиты информации. Общие технические требования» при обеспечении безопасности средств криптографической защиты информации. Приведены результаты анализа угроз, возникающих из-за отсутствия требований к процессам жизненного цикла средств криптографической защиты информации. Выработаны предложения и подходы к установлению этих требований.*

Введение

В соответствии со статьей 4 Закона Республики Казахстан «О национальной безопасности Республики Казахстан»[1] информационная безопасность является одним из видов национальной безопасности Республики Казахстан.

Существование такой нормы в законодательстве Республики Казахстан определяет важность поддержания в актуальном состоянии национальных и государственных стандартов Республики Казахстан, регламентирующих требования к средствам защиты информации, так как они напрямую влияют на ход разработки, реализации и эксплуатации этих средств.

В соответствии с подпунктом 18) статьи 1 Закона Республики Казахстан от 30 декабря
2020 года № 396-VI ЗРК «О техническом регулировании» под жизненным циклом средств криптографической защиты информации (далее – СКЗИ) следует понимать процессы их разработки (проектирования, включая изыскания – научная составляющая), производства, реализации (обращение на рынке), эксплуатации и утилизации.

Нормы Закона Республики Казахстан от 16 мая 2014 года № 202-V ЗРК «О разрешениях и уведомлениях»[2] разработка и реализация СКЗИ, в том числе иная их передача, относятся к лицензируемому виду деятельности, для которого требуются специальные лицензии и разрешения уполномоченного государственного органа Республики Казахстан – Комитета национальной безопасности Республики Казахстан.

Вместе с тем, требования к остальным процессам жизненного цикла не описаны ни в этом, ни в каких-либо других нормативно-правовых актах, включая технические регламенты и национальные стандарты, в том числе Государственный стандарт Республики Казахстан СТ РК 1073-2007 «Средства криптографической защиты информации. Общие технические требования»[3] (далее – СТ РК 1073-2007). Данное обстоятельство может являться источником рисков безопасности информации, защищаемой с применением СКЗИ.

5. Роль и место СТ РК 1073-2007

В Республике Казахстан нормы национальных и государственных стандартов носят добровольный характер[4], если иное не установлено законодательством Республики Казахстан.

При этом ряд нормативных правовых актов Республики Казахстан переводят нормы СТ РК 1073-2007 из добровольного характера в обязательные.

Так, обязательность применения норм СТ РК 1073-2007 государственными органами, местными исполнительными органами, государственными юридическими лицами, субъектами квазигосударственного сектора, собственниками и владельцами негосударственных информационных систем, интегрируемых с информационными системами государственных органов или предназначенных для формирования государственных электронных информационных ресурсов, а также собственниками и владельцами критически важных объектов информационно-коммуникационной инфраструктуры при обеспечении конфиденциальности с помощью СКЗИ нормативно закреплена постановлением Правительства Республики Казахстан от 20 декабря 2016 года № 832 «Об утверждении единых требований в области информационно-коммуникационных технологий и обеспечения информационной безопасности»[5].

Создание и использование электронных документов, удостоверенных посредством электронных цифровых подписей (далее – ЭЦП), предусматривающих установление, изменение или прекращение правоотношений, а также прав и обязанностей участников правоотношений, возникающих в сфере обращения электронных документов, включая совершение гражданско-правовых сделок, нормативно закреплены Законом Республики Казахстан «Об электронном документе и электронной цифровой подписи»[6].

Для удостоверения соответствия открытого ключа ЭЦП закрытому ключу ЭЦП, а также подтверждения достоверности регистрационного свидетельства в Республике Казахстан действуют удостоверяющие центры (далее – УЦ).

Легитимность деятельности УЦ обеспечивается проведением специальной комиссией процедуры аккредитации УЦ в соответствии с Правилами выдачи и отзыва свидетельства об аккредитации удостоверяющих центров [7].

В соответствии с подпунктом 9) пункта 25 указанных правил СКЗИ, используемые в УЦ и его пользователями, должны иметь сертификат соответствия требованиям СТ РК 1073-2007.

Кроме того, подпунктом 3) пункта 3 приказа Министра по инвестициям и развития Республики Казахстан от 28 декабря 2015 года № 1261 «Об утверждении Правил регистрации и прекращения взаимодействия удостоверяющих центров, доверенных третьих сторон иностранных государств с доверенной третьей стороной Республики Казахстан»[8] предписана обязательность применения в УЦ СКЗИ, соответствующих требованиям СТ РК 1073-2007.

Помимо этого, требования Правил отнесения сведений к служебной информации ограниченного распространения и работы с ней, утвержденных постановлением Правительства Республики Казахстан от 24 июня 2022 года № 429[9], предусматривают применение отечественных СКЗИ[1] для защиты сведений с ограничительной пометкой

[1] 12. База данных облачного хранилища должна быть защищена с использованием отечественных средств криптографической защиты информации не ниже третьего уровня безопасности по государственному стандарту СТРК 1073-2007 "Средства криптографической защиты информации. Общие технические требования.".
14. Подключение к облачному хранилищу, а также передача служебной информации ограниченного распространения в электронном виде осуществляются по каналам связи с использованием отечественных

«Для служебного пользования». При этом, в нормативных правовых и иных документов не приведены методы и условия отнесения СКЗИ к отечественным, что создает правовые коллизии.

Таким образом, роль и место СТ РК 1073-2007 в обеспечении информационной безопасности критической информации критически важна, следовательно, установление актуальных требований информационной безопасности и процессам жизненного цикла СКЗИ является важнейшим условием обеспечения самых разных интересов Республики Казахстан.

6. Жизненный цикл СКЗИ

Для защиты информации криптографическими методами требуются соответствующие технические средства, как правило программные и аппаратные, соответствующие условия работы (вспомогательные технические средства, защищённые помещения, системы электропитания и связи, технический персонал и пользовательская документация), а также криптографические ключи. Для обеспечения действительно надёжной защиты эти средства необходимо обслуживать, проводить регламентный контроль исправности, не допускать компрометации средств и ключей, обеспечивать необходимый набор организационных мер, направленных на предотвращение утечки информации тем или иным способом, а также периодически проводить оценку, обновление и совершенствование принятых мер защиты информации.

6.1. Действующие НПА

В ходе изучения законодательства Республики Казахстан не обнаружены нормативные правовые акты, напрямую управляющие жизненным циклом СКЗИ. Косвенно на него влияет СТ РК 1073-2007, но только в части обеспечения некоторых показателей надёжности защиты, предоставляемой СКЗИ. При этом требования СТ РК 1073-2007 не являются полными, так как не учитывается происхождение СКЗИ, механизмы контроля компрометаций и регламента устранения их последствий, возможности наличия незадекларированных возможностей, процедуры управления СКЗИ и ключевой информации, надёжность источников случайности, наличие технических каналов утечки данных и множество иных факторов, способных негативно повлиять на безопасность защищаемых сведений.

6.2. Объективная реальность

В результате отсутствия единого подхода к управлению полным жизненным циклом СКЗИ сложившаяся в результате неконтролируемого эволюционного развития система практик не покрывает все этапы, через которые должны проходить СКЗИ для того, чтобы обеспечить требуемый уровень безопасности информации, что порождает серьёзные риски безопасности в случае, если организация, эксплуатирующая СКЗИ, не

средств криптографической защиты информации не ниже третьего уровня безопасности по государственному стандарту СТРК 1073-2007 "Средства криптографической защиты информации. Общие технические требования.".

15. Руководители организации принимают все необходимые организационные, технические и правовые меры по обеспечению защиты служебной информации ограниченного распространения вне зависимости от вида, формы, содержания и предоставления электронного носителя, в том числе с использованием соответствующих средств криптографической защиты информации и (или) информационных систем в защищенном исполнении, отнесенных к государственным секретам.

занимается целенаправленно построением собственной цельной системы информационной безопасности, включая анализ рисков, построение моделей угроз, разработку политик безопасности, организацию доверенных служб, аудит и учёт, собственные узлы депонирования ключей и так далее. Очевидно, что по тем или иным причинам не все организации обладают достаточным ресурсом для проведения таких работ, причём более серьёзной преградой может являться даже не сложность и дороговизна достижения достаточно высокого научного и технического потенциала, а в целом отсутствие понимания угроз, которые влечёт за собой бессистемное применение СКЗИ.

6.3. Наиболее серьёзные угрозы для СКЗИ

Из угроз, порождаемых отсутствием контроля за жизненным циклом СКЗИ, наиболее серьёзными представляются следующие:

1. Наличие незадекларированных возможностей, заложенных разработчиком и/или производителем СКЗИ.
2. Наличие ошибок в реализации криптографических и иных механизмов, влияющих на работу СКЗИ.
3. Внесение в СКЗИ изменений, нарушающих его работу и понижающих защищённость обрабатываемых сведений, уже на этапе его работы.
4. Случайное или умышленное использование в СКЗИ слабых или устаревших криптографических алгоритмов и протоколов.
5. Депонирование заведомо слабых криптографических ключей.
6. Извлечение ключевых данных из СКЗИ, изъятых с участка работы для ремонта, хранения или утилизации.
7. Появление новых и совершенствование существующихатак на находящиеся в эксплуатации СКЗИ, реализованные в них алгоритмов и протоколов.
8. Отсутствие действенных механизмов реагирования на компрометацию.

Отсутствие механизмов уничтожения ключевой информации.

6.4. Краткий обзор мировых практик

В мировой практике также не встречается описание единого подхода ко всему жизненному циклу СКЗИ. Такая ситуация порождает разобщённость акторов по месту, времени и зонам ответственности: производитель, ритейлер, управляющая и эксплуатирующая организации руководствуются собственными различными положениями и регламентами, и эти сотни и тысячи акторов не имеют возможности выработать единый стандарт, и не заинтересованы в этом, так как стандарт наложит дополнительные требования на существующие бизнес-процессы и усложнит их.

Вместе с тем, в некоторых странах СНГ приняты соответствующие нормативные правовые акты, регламентирующие требования к жизненному циклу СКЗИ, утвержденные приказами уполномоченных государственных органов.

Так, в Украине в соответствии со статьей 3 Закона Украины «О Государственной службе специальной связи и защиты информации Украины» № 3475-IV[10] формирование и реализация государственной политики в сфере криптографической защиты информации возложена на Государственную службу специальной связи и защиты информации Украины.

Основным нормативным правовым документом, регулирующим деятельность в области криптографической защиты информации в Украине, является Указ Президента Украины «Об утверждении Положения о порядке осуществления криптографической защиты информации в Украине» от 22 мая 1998 года № 505[11]. Данное Положение

определяет порядок осуществления криптографической защиты информации с ограниченным доступом, разглашение которой наносит (может нанести) ущерб государству, обществу или лицу.

При этом требования к порядку разработки, производства и эксплуатации средств криптографической защиты государственных информационных ресурсов или информации, требование относительно защиты которой установлено законом, средств, специально предназначенных для разработки, исследования, производства и испытания средств криптографической защиты информации, нормативно закреплены в Положении о порядке разработки, производства и эксплуатации средств криптографической защиты информации, утвержденном приказом Администрации Государственной службы специальной связи и защиты информации Украины от 20 июля 2007 года № 141[12].

В Украине в процессе разработки, производства и эксплуатации средств криптографической защиты участвуют и взаимодействуют между собой:

- заказчики;
- разработчики;
- производители;
- организации, эксплуатирующие средства криптографической защиты;
- организации, производящие сертификационные испытания (экспертные работы);
- поставщики ключевых документов (организация или подразделение, которые определяются Администрацией Государственной службы специальной связи и защиты информации Украины для осуществления поставки ключевых документов к средствам криптографической защиты информации;
- Администрация Государственной службы специальной связи и защиты информации Украины.

Согласно пункту 25 Положения о Федеральной службе безопасности Российской Федерации, утвержденного Указом Президента Российской Федерации от 11 августа 2003 года № 960[13], Федеральная служба безопасности осуществляет регулирование в области разработки, производства, реализации, эксплуатации, ввоза в Российскую Федерацию и вывоза из Российской Федерации шифровальных (криптографических) средств и защищенных с использованием шифровальных средств систем и комплексов телекоммуникаций, а также в области предоставления на территории Российской Федерации услуг по шифрованию информации.

Пунктом 47 того же Положения на Федеральную службу безопасности Российской Федерации возложена функция по организации и проведению исследований в области защиты информации, экспертных криптографических, инженерно-криптографических и специальных исследований шифровальных средств, специальных и закрытых информационно-телекоммуникационных систем.

Кроме того, в соответствии с пунктами 21 и 22 вышеуказанного Положения именно Федеральная служба безопасности Российской Федерации занимается сертификацией средств криптографической защиты информации и лицензированием деятельности в области разработки, производства и распространения средств криптографической защиты информации.

Общие положения отрасли защиты информации, в частности, криптографической защиты, регулируются Федеральным законом от 27 июля 2006 года № 149-ФЗ «Об информации, информационных технологиях и о защите информации»[14].

Основным нормативным документом в сфере криптографической защиты информации является Положение о разработке, производстве, реализации и эксплуатации шифровальных (криптографических) средств защиты информации (Положение ПКЗ-

2005), утвержденное приказом Федеральной службы безопасности Российской Федерации от 9 февраля 2005 года № 66[15].

Следует отметить, что несмотря на наличие в этом приказе раздела, посвященного эксплуатации средств криптографической защиты информации, данный приказ в большей степени ориентирован на разработчиков средств криптографической защиты информации, а не на конечных пользователей.

В Республике Беларусь вопросы криптографической защиты информации, не содержащей сведений, отнесенных к государственным секретам, регулируются:

- Законом Республики Беларусь от 10 ноября 2008 года № 455-З «Об информации, информатизации и защите информации»[16];
- Законом Республики Беларусь от 28 декабря 2009 года «Об электронном документе и электронной цифровой подписи»[17];
- Указом Президента Республики Беларусь от 16 апреля 2013 года № 196 «О некоторых мерах по совершенствованию защиты информации»[18], которым утверждено Положение о технической и криптографической защите информации в Республике Беларусь;
- приказом Оперативно-аналитического центра при Президенте Республики Беларусь от 30 августа 2013 года № 62 «О некоторых вопросах технической и криптографической защиты информации»[19];
- приказом Оперативно-аналитического центра при Президенте Республики Беларусь от 03 марта 2011 года № 18 «Об утверждении Положения о порядке применения средств криптографической защиты информации в системах защиты информации»[20].

Требования к безопасности средств криптографической защиты информации нормативно закреплены в Государственном стандарте Республики Беларусь СТБ 34.101.27-2022 «Информационные технологии и безопасность. Средства криптографической защиты информации. Требования безопасности»[21].

В указанном стандарте помимо прочих требований в разделе «Гарантированные требования» приводятся требования к проектированию и разработке, поддержке жизненного цикла, программе испытаний, анализу программ, а в приложениях к нему приводятся содержание функциональной спецификации, принципы построения генераторов случайных чисел и механизмы тестирования их качества, а также памятка по анализу исходных текстов программ.

Указанный стандарт не устанавливает требования процедурного характера, а также вообще не рассматривает требования к производству и утилизации (уничтожению) средств криптографической защиты информации.

Следует отметить, что разработка, производство (либо выборка из указанного перечня работ) технических средств обработки информации в защищенном исполнении, программных средств обработки информации в защищенном исполнении, технических, программных, программно-аппаратных средств защиты информации и контроля ее защищенности, средств криптографической защиты информации (либо выборка из указанного перечня средств) относятся к лицензируемому виду деятельности, лицензия выдается Оперативно-аналитическим центром при Президенте Республики Беларусь (Положение о лицензировании отдельных видов деятельности, утвержденное Указом Президента Республики Беларусь от 1 сентября 2010 года № 450 «О лицензировании отдельных видов деятельности»[22]).

7. Предложения к установлению требований к жизненному циклу СКЗИ в республике Казахстан

Простейшим способом установить требования к вышеперечисленным акторам и строгость их соблюдения является совершенствование механизма лицензирования. Необходимо внести в них несколько изменений, которые в комплексе серьёзно поменяют как технические, так и бизнес-процессы:

1. Лицензия не должна выдаваться единожды без дальнейшей периодической аккредитации. Это заставит акторов поддерживать свой научно-технический потенциал на должном уровне.

2. Различные виды деятельности должны требовать различных лицензий. Например, очевидно, что большинство акторов задействованы только в процессах реализации, но не разработки СКЗИ, тогда и требования к ним должны предъявляться на уровне опытных эксплуататоров, а не криптографов-теоретиков. Отдельно должны быть лицензированы такие виды деятельности, как разработка алгоритмов и криптографических схем и протоколов, разработка аппаратных модулей для СКЗИ, сертификация СКЗИ, реализация СКЗИ, эксплуатация СКЗИ, утилизация СКЗИ.

3. При выдаче лицензии проводить более глубокий анализ научно-технического уровня актора. Так, для проведения теоретических исследований он должен обладать реальным научно-практическим опытом, что подтверждается наличием соответствующего образования, публикациями в международных рецензируемых журналах, участием в авторитетных международных выставках в роли докладчика – набор требований, хорошо известный в научной среде. Такие акторы, а также претендующие на лицензию изготовителя СКЗИ, должны также обеспечить необходимые организационные и технические меры защиты своих разработок и производства, и так далее.

4. В свете общемировых тенденций на быстрое устаревание научных и технических решений требуется постоянное совершенствование средств криптографической защиты информации. В этой связи предлагается постоянно совершенствовать требования, выдвигаемые к акторам в области разработки и изготовления СКЗИ. Для этого требуется создание профильной комиссии (например, на базе профильных вузов или Академии наук), обладающей достаточным научным потенциалом для отслеживания новейших достижений в области информационной безопасности и криптографии, которая будет с заданным периодом вносить изменения в список требований к акторам и экзаменационные вопросы, применяемые для оценки их научно-технического потенциала.

5. Крайне желательно ввести механизмы периодического контроля компетенции органов подтверждения соответствия, которая будет включать в себя выборочную проверку выданных ими заключений, а также оценку результатов их деятельности (ретроспектива инцидентов информационной безопасности). По результатам этого контроля в случае выявления нарушений, низкой компетенции исполнителей и руководителей организации, а также в случае, когда лицензируемая деятельность фактически не осуществляется, возможен отзыв лицензии.

По остальным вопросам достаточно простого упорядочивания регулирования отношений между различными акторами и ведения систематизированного учёта. В этой связи предлагается:

1. Обязать всю цепочку акторов вести поэкземплярный учёт СКЗИ и журнал с историей всех манипуляций над ними начиная с изготовления и заканчивая утилизацией. Этот учёт должен осуществляться централизованно с использованием государственной

базы данных. Сертификационные органы должны быть обязаны вносить соответствующую информацию о сертифицируемых СКЗИ в эту базу.

2. Обязать эксплуатирующие подразделения проводить периодический контроль технического состояния СКЗИ с внесением соответствующих записей в журнал.

Сделать обязательным использование в СКЗИ отечественных криптографических алгоритмов или гармонизированных стандартов. Это исключит возможность применения преднамеренно ослабленных и устаревших СКЗИ, а также создаст условия для развития отечественных разработчиков.

8. Ожидаемый результат

После нормативного закрепления требований к жизненному циклу СКЗИ:

1) повысится защищённость информационных систем и защищаемой информации;

2) усилится контроль за организационными и техническими мерами, принимаемыми для защиты критической информации;

3) появится и разовьётся независимый отечественный рынок СКЗИ. Средства, разработанные и произведённые в РК, должны получить преимущества при сертификации и подтверждении качества за счёт наличия у локального производителя исчерпывающей конструкторской документации и документов, подтверждающих историю СКЗИ;

4) усложнится внедрение в эксплуатацию низкокачественного оборудования в сфере защиты информации, а также оборудования, имеющего незадекларированные возможности.

Список литературы

1. Закон Республики Казахстан № 527-IV, «О национальной безопасности Республики Казахстан,» 6 января 2012 года.
2. Закон Республики Казахстан № 202-V ЗРК, «О разрешениях и уведомлениях,» 16 мая 2014.
3. Комитет по техническому регулированию и метрологии Министерства индустрии и торговли Республики Казахстан (Госстандарт), «Государственный стандарт Республики Казахстан "Средства криптографической защиты информации. Общие требования",» 2007.
4. Закон Республики Казахстан № 183-VI, «О стандартизации,» 5 октября 2018 года.
5. Постановление Правительства Республики Казахстан № 832, «Об утверждении единых требований в области информационно-коммуникационных технологий и обеспечения информационной безопасности,» 20 декабря 2016 года.
6. Закон Республики Казахстан N 370, «Об электронном документе и электронной цифровой подписи,» 7 января 2003 года .
7. Приказ Министра цифрового развития, инноваций и аэрокосмической промышленности Республики Казахстан № 224/НҚ, «Правила выдачи и отзыва свидетельства об аккредитации удостоверяющих центров,» 1 июня 2020 года.
8. Приказ Министра по инвестициям и развитию Республики Казахстан от 28 декабря 2015 года № 1261. Зарегистрирован в Министерстве юстиции Республики Казахстан № 13098, «Об утверждении Правил регистрации и прекращения взаимодействия удостоверяющих центров, доверенных третьих сторон иностранных государств с доверенной третьей стороной Республики Казахстан,» 17 февраля 2016 года.
9. Постановление Правительства Республики Казахстан № 429, «Об утверждении Правил отнесения сведений к служебной информации ограниченного распространения и работы с ней,» 24 июня 2022 года .

10. Закон Украины № 3475-IV, «О Государственной службе специальной связи и защиты информации Украины,» 23 февраля 2006 года.
11. Указ президента Украины №505/98, «О Положении о порядке осуществления криптографической защиты информации в Украине,» 22 мая 1998 года.
12. Приказ Администрации Государственной службы специальной связи и защиты информации Украины № 141, «Об утверждении Положения о порядке разработки, производства и эксплуатации средств криптографической защиты информации,» 20 июля 2007 года.
13. Указ Президента РФ N 960, «Вопросы Федеральной службы безопасности Российской Федерации,» 11 августа 2003 г..
14. Федеральный закон N 149-ФЗ, «Об информации, информационных технологиях и о защите информации,» 27.07.2006.
15. Приказ ФСБ РФ № 66, «Об утверждении Положения о разработке, производстве, реализации и эксплуатации шифровальных (криптографических) средств защиты информации (Положение ПКЗ-2005),» 9 февраля 2005 года.
16. Закон Республики Беларусь № 455-З, «Об информации, информатизации и защите информации,» 10 ноября 2008 г..
17. Закон Республики Беларусь № 113-З, «Об электронном документе и электронной цифровой подписи,» 28 декабря 2009 г..
18. Указ президента Республики Беларусь №196, «О некоторых мерах по совершенствованию защиты информации,» 16 апреля 2013 года.
19. Приказ ОАЦ при Президенте РБ 62, «О некоторых вопросах технической и криптографической защиты информации,» 30.08.2013.
20. Приказ Оперативно-аналитического центра при Президенте Республики Беларусь № 18, «Об утверждении Положения о порядке применения средств криптографической защиты информации в системах защиты информации,» 3 марта 2011 года.
21. СТБ 34.101.27-2022, «Информационные технологии и безопасность. Средства криптографической защиты информации. Требования безопасности,» 2022.
22. Указ Президента Республики Беларусь № 450, «О лицензировании отдельных видов деятельности,» 1 сентября 2010 г..

CURRENT AREAS OF SECURITY ENSURING IN WIRELESS SENSOR NETWORKS

Zhukabayeva T.[1,2], Adamova A.[1,3], **Khu Ven-Tsen B.**[1,4], Zholshiyeva L.[1]
e-mail: aigul.adamova@astanait.edu.kz
[1]*International Science Complex «ASTANA», Kazakhstan*
[2]*L. Gumilyov Eurasian National University, Kazakhstan*
[3]*Astana IT University, Kazakhstan*
[4]*M. Auezov South Kazakhstan University, Kazakhstan*

Annotation. *Wireless sensor networks (WSNs) represent one of the most significant achievements of scientific and technological progress in the field of modern information technologies. They are widely and universally used since they allow one to successfully solve various problems of a production, military, and domestic nature. As a result, various aspects of ensuring the reliable and efficient operation of WSNs, including security, stability, and confidentiality, are becoming increasingly important.*

The purpose of this study is to systematize and analyze approaches to solving this problem known in world practice and identify emerging trends and the most relevant areas of research in this area. The results of this study can be used in planning research and identifying research problems, as well as in the practice of creating and operating cyber-physical systems based on WSNs.

Introduction

A wireless sensor network WSN is a distributed self-organizing network in which various sensors and automatic actuators are located in its nodes and it is informationally interconnected via a radio channel. An important advantage of such networks is their availability for implementation in rugged and inaccessible terrain. They can cover large areas due to the possibility of relaying messages from one network node to another. WSNs form the telecommunications basis for the rapidly developing Internet of Things (IoT) technology.

A WSN is subject to various vulnerabilities, which pose certain risks to its reliable and efficient operation in terms of security, stability, and confidentiality like any telecommunications network. For this reason, scientific and technical developments in the field of countering these threats are becoming increasingly important.

To carry out such developments, it is necessary to have reliable information on the current status in this area, as well as observed trends in the approaches used in practice.

This paper presents the results of a systematic analysis of approaches to ensuring security in WSN networks known in world practice, as well as emerging trends and the most current areas of research in this area. The research was based on materials and data from specialized literature and other open sources. The results obtained are oriented towards use by specialized specialists in planning research and identifying research problems, as well as in the practice of creating and operating cyber-physical systems based on WSNs.

The article discusses the priority aspects of ensuring the safe and reliable functioning of the WSN.

1. Device compatibility and energy consumption in the WSN

Currently, WSNs have become an integral element of IoT structures, Smart systems, and other cyber-physical systems. Such systems are created and function in various areas of human life. Such as healthcare, urban planning and organization of the living environment, industrial automation, military, and many other areas (Fig. 1).

Manufacturers of various elements of technical, energy, and communication equipment for WSNs are appearing. In connection with this, the problem of ensuring the compatibility of these elements has become urgent, which is reflected in the creation of appropriate protocols.

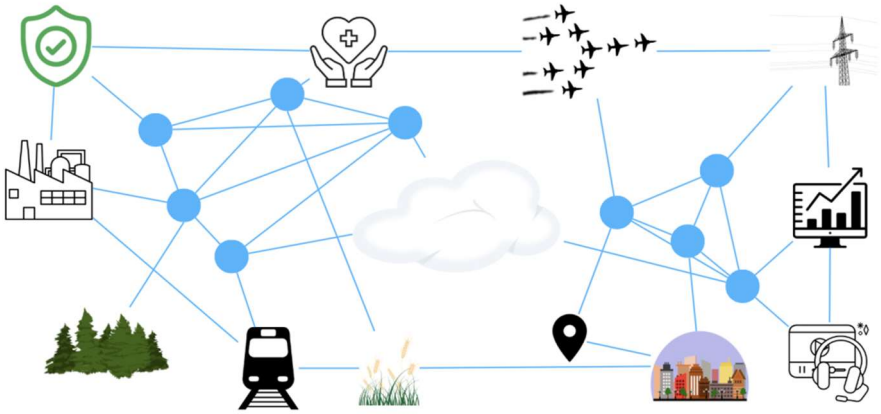


Fig. 1 WSN Applications

There are more than 50 companies specializing in the design and production of products for the standardized Thread And Matter protocol at the 6LoWPAN level, which is effective in Smart objects in the categories “smart house”, “smart building”, etc. (https://www.thread-group.org /).

Current research areas include measures to extend the lifespan of the WSNs network, particularly the sensor nodes, as they become unavailable once the network is deployed. In this regard, minimizing energy consumption is a priority criterion in the design process of the sensor network.

The protocol stack for WSN consists of five layers (Fig. 2). Approaches to reducing energy consumption can be considered at each level of the architecture.

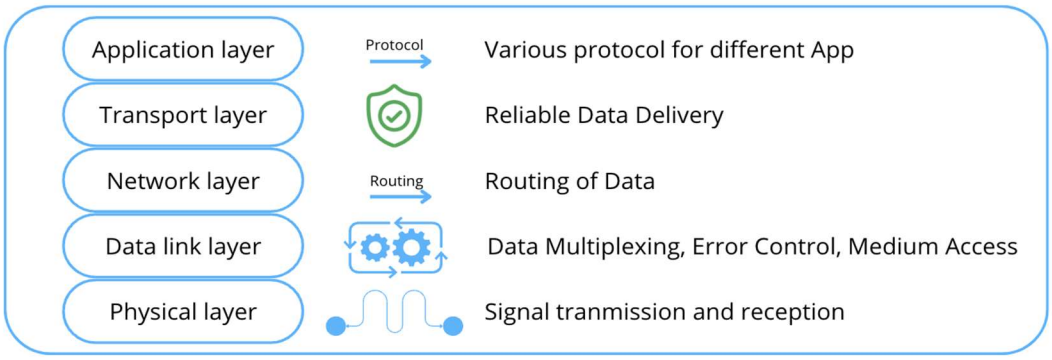


Fig. 2 WSN Protocol Stack

The presented protocol architecture layers ensure the following functional and operational requirements are met.

The physical layer regulates the processes of receiving and transmitting data in the physical equipment used in terms of matching the characteristics of the receiving and transmitting devices. The layer determines carrier frequency generation and selection, signal detection, signal modulation, and encryption, and signal reception.

Due to the use of a radio channel to transmit and receive data, energy consumption is significant. Based on this, three different operating modes are defined for it: standby mode, active, and sleep modes. Energy consumption is minimized by enabling sleep mode [1].

Data link layer. This layer ensures that adjacent signals do not interfere with each other in a noisy environment. This layer must have appropriate access, error control, multiplexing, and error detection and correction.

One of the means of effective energy management in this case is to reduce the time between frame transmission over the network channel and listening in standby mode [2].

Network layer. At this level, topology management and routing schemes are regulated to increase the service life of network equipment.

It should be noted that choosing the optimal topology that can provide a well-connected network is often a difficult task. Routing also plays an important role in increasing the lifespan of network equipment. Since it must ensure the selection of the most energy-efficient path from sensor nodes to the base station. Well-known routing algorithms with hierarchical clustering are effective in determining the optimal route [3].

Transport layer. Designed to provide effective regulation of network traffic flow to a remote end node. At the same time, additional traffic is provided with reliability measures. For this purpose, it is divided into successive segments when data is sent by the upper-level application, which are then reassembled into data packets [4].

Application layer. Serves as a connection point between users and network services: email, file transfer, virtual terminals, and file servers.

Solving the problems of reducing the energy dependence of WSN equipment is facilitated by extracting energy from the environment using solar, thermal, and vibration technologies, combined with the use of recharging equipment after its energy depletion [4].

2. Countering network attacks

This section describes WSN security issues and analyses possible solutions proposed by researchers.

Figure 3 presents a three-level classification of attacks by type. The practical measures to eliminate the types of attacks considered are presented.

These attacks are divided into two parts - active and passive.

Active attacks are those that involve or make changes to the data transmitted from the sending node to the receiving node. In this case, attackers are directly involved in the process of receiving and transmitting data.

In a passive attack scenario, nodes can be tracked over a channel of two communicating sensor nodes. The attacker collects and determines important information without interfering with network information exchange, simply observing the situation. Passive attackers are more difficult to detect because they do not change the content of information transfer from one node to another. The intention of a passive attacker is usually to obtain sensitive information [5].

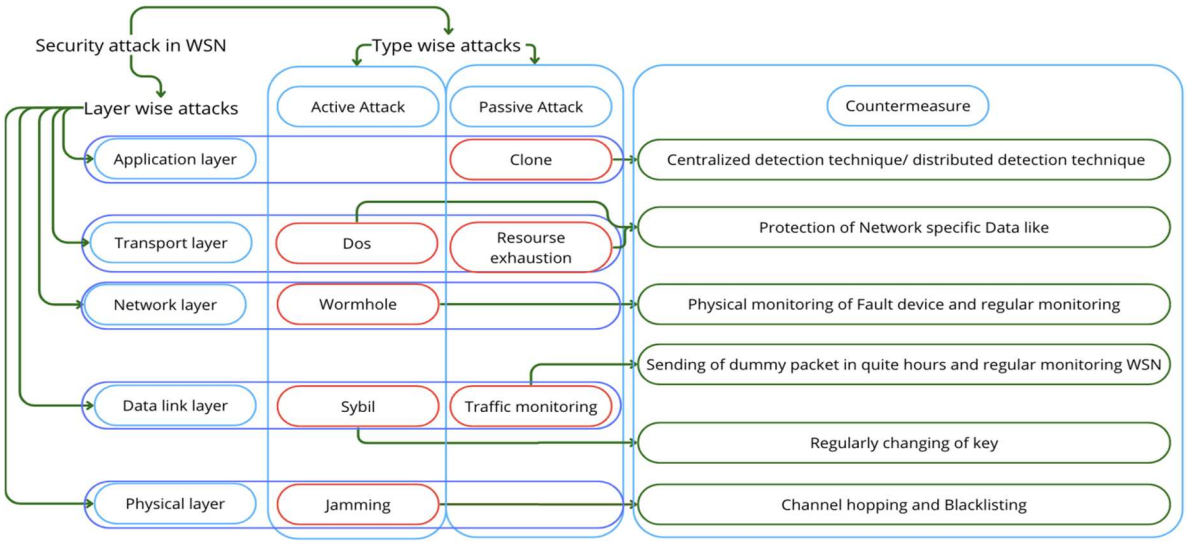


Fig. 3 Classification of network attacks by type and levels of architecture

3. Machine learning algorithms used for WSN security

This section discusses the known types of machine learning algorithms that are used to provide WSN security. They are divided into levels of WSN architecture and are classified including controlled, uncontrolled, reinforcement learning, and deep learning.

The classification of algorithms is presented in Figure 4.

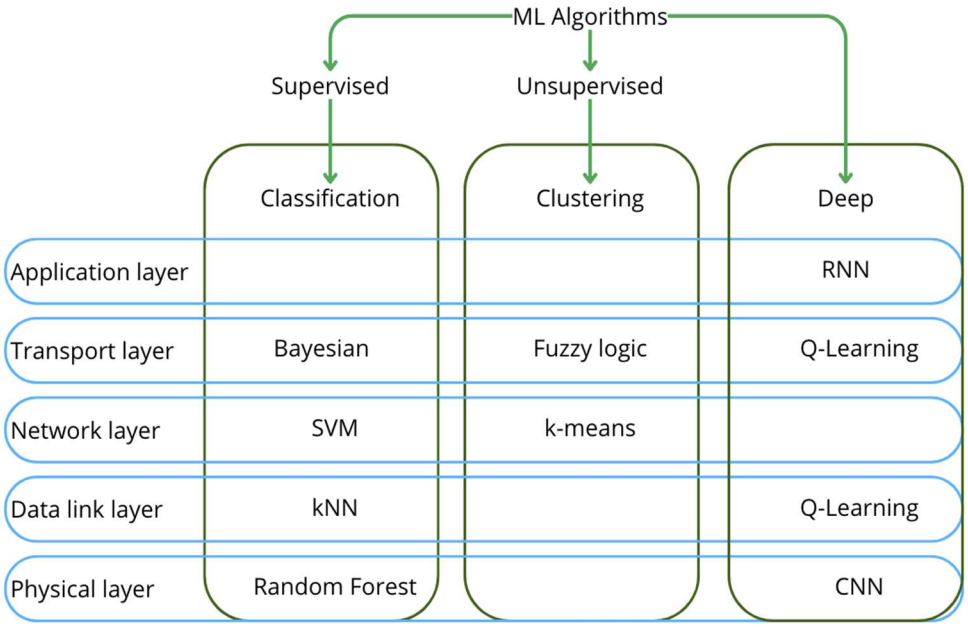


Fig. 4 Classification of used machine learning algorithms

Supervised category algorithms allow to prediction of future samples with high performance and accuracy, but the learning process requires considerable time and hardware

resources. These limitations make it difficult to apply the learning process to applications that require real-time outputs. Bayesian, SVM, kNN, Random Forest, etc.

Unsupervised class algorithms are widely used for machine learning in the neural network environment. Training is based on the use of various data patterns without knowledge of the output characteristics. The main area of application of Unsupervised algorithms is the assessment of data density, to find signs of commonality between elements and order them statistically.

The difference between Unsupervised and Supervised algorithms is that training using unsupervised algorithms is less labor-intensive and faster to implement, but the accuracy of predicting output data is lower. For example, these are the Fussy Logic and K-means algorithms.

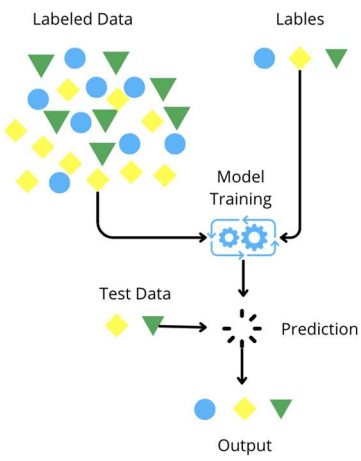
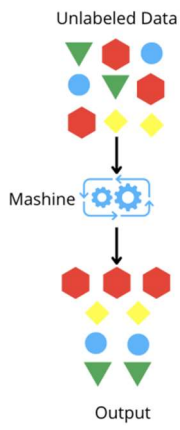
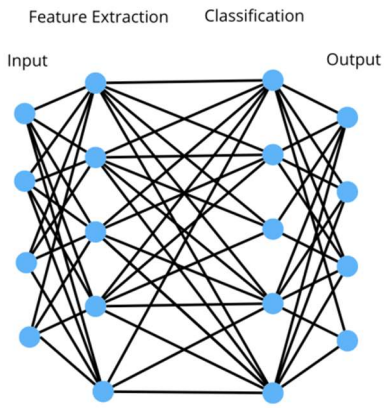
Deep learning algorithms based on classification technology that uses a multi-layered way of representing data known as deep learning methods (between an input layer and an output layer). The deep learning technique for deep learning as well as supervised and unsupervised is presented in Table 1.

The basis of the Deep Learning Technique is the use of basic nonlinear modules that transform the description from a lower level to a higher level to achieve an optimal result. Deep learning has several advantages, including the ability to extract high-level features of data, the ability to work with or without labels, and the ability to learn to achieve different goals.

Algorithms in the deep learning category are RNN, Q-Learning, CNN, etc.

Table 1.

Examples of Machine Learning Algorithms

Supervised learning	Unsupervised learning	Deep learning
		

3. Review and analysis of research publications

Table 2 shows the systematic basic points of the literature examined on the subject. The review was conducted for specialized publications in the public domain over three years from 2020 to 2023.

Figure 5 presents the number of research publications.

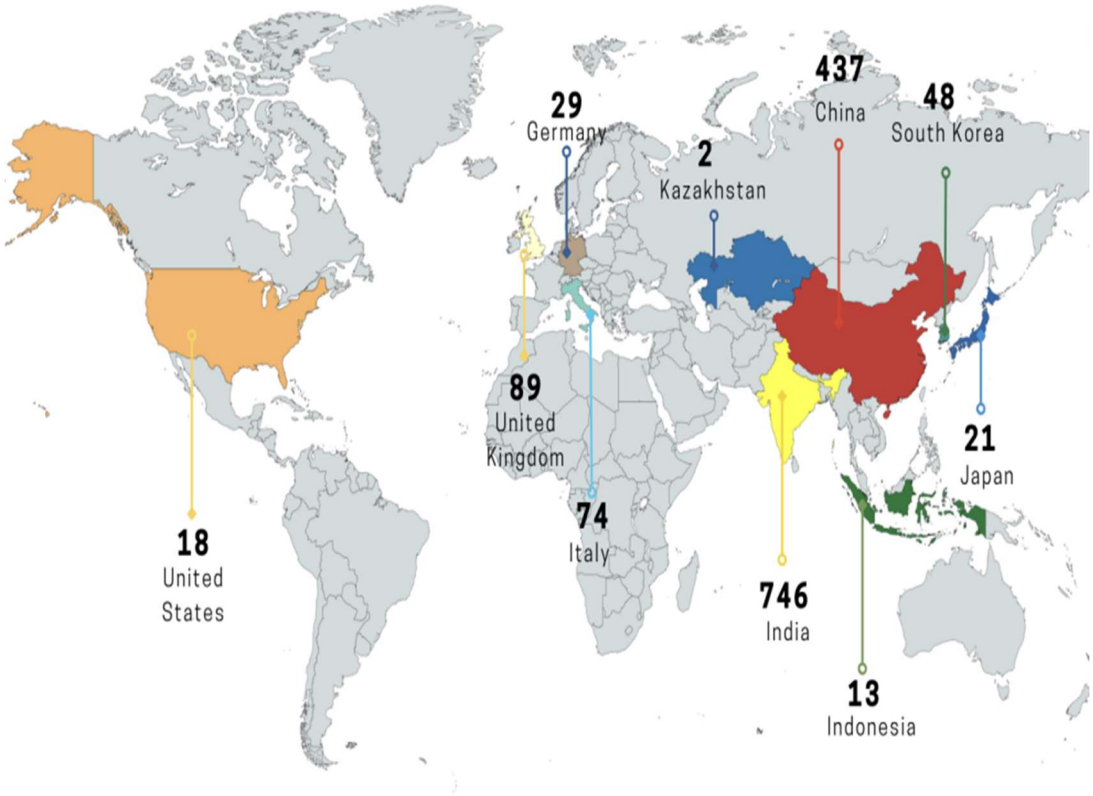


Fig. 5. Number of scientific publications by country.

The literary sources were selected according to the frequency of citations and references in scientific articles and publications. The review period is linked to the timing of the first WSN publications.

Table 2.

Systematization of literature analysis.

Research	Authors	Basic ideas
1	2	3
[7]	K. Shafique et al.	The paper presents a detailed and extensive review of the 5G-IoT scenario.
[8]	I. Butun et al.	The paper presents a detailed overview of attacks on WSN security systems and the Internet of Things, as well as methods of prevention, detection, and mitigation of these attacks.
[9]	A. Ahmed et al.	The article offers an energy-efficient mechanism of data aggregation, protected by a blockchain, which uses a mechanism of data aggregation at the cluster level for energy saving.
[10]	A. Halbouni et al.	The article offers a hybrid model of intrusion detection systems.
[11]	M. N. M. Bhutta et al.	The article offers a monitoring and reporting system based on the Internet of Things.

1	2	3
[12]	S. Pundir et al.	The article offers detailed information about the threat model applicable for communication security based on WSN and IoT. Along with this, the security requirements and various attacks possible in communication environments based on WSN and IoT were discussed.
[13]	Z. Cui et al.	The paper proposes an authentication scheme with multiple-based blockchain for the Internet of Things.
[14]	M. A. Al-Naeem	The work focuses on distributed denial of service in WSN.
[15]	O. R. Ahutu et al.	The article presents a simplified routing protocol for WSN, the purpose of which is to minimize energy consumption, as well as to detect attacks through wormhole attacks.

The conducted analysis shows that the scope of WSN network proliferation is rapidly expanding, resulting in an update of all aspects of ensuring their effective functioning, including security, sustainability, and confidentiality. At the same time, ensuring the energy independence of the technical equipment of the WSN is an important component.

Development of different approaches to security management based on protocol compatibility of WSN architecture elements, and the use of proven algorithms for machine learning of intelligent network elements are topical ways of achieving progress in this field.

Conclusion

Wireless sensor networks represent a significant and promising element of modern IT infrastructure for various purposes, and the scope of their distribution is rapidly expanding. The problem of ensuring the security of the WSN is important and urgent, as the efficiency and reliability of the network operations depend critically on it.

Current areas for ensuring WSN security are protocol compatibility of network elements, non-volatility of technical equipment, and countering possible network attacks based on machine learning of intelligent network subsystems.

The conducted studies systematized various aspects of this problem. The results obtained can be applied in a practical way to the informed planning of research work, selection of topics, and directions of research.

ACKNOWLEDGMENT

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP19680345).

References

1. 1. O. JOUINI and K. SETHOM, "Physical Layer Security Proposal for Wireless Body Area Networks," 2020 IEEE 5th Middle East and Africa Conference on Biomedical Engineering (MECBME), Amman, Jordan, 2020, pp. 1-5, doi: 10.1109/MECBME47393.2020.9265157.
2. 2. M. Meribout and A. Al Naamany, "A collision-free data link layer protocol for wireless sensor networks and its application in intelligent transportation systems," 2009 Wireless Telecommunications Symposium, Prague, Czech Republic, 2009, pp. 1-6, doi: 10.1109/WTS.2009.5068957.
3. U. Ghugar, G. Sivanarayan, G. Rajareddy, K. S. Varma and G. Lakshmeeswari, "TBSMFA: Trust-Based Security Mechanism for SYN Flooding Attack at Transport Layer using DRICC technique in WSN," 2022 OPJU International Technology Conference on Emerging

Technologies for Sustainable Development (OTCON), Raigarh, Chhattisgarh, India, 2023, pp. 1-6, doi: 10.1109/OTCON56053.2023.10113952.

4. S. P. Win, W. L. Lelt Phyu and T. Z. Thaw, "Shedding Light on Technological Treasures within IoT Application Layer Protocol," 2023 IEEE Conference on Computer Applications (ICCA), Yangon, Myanmar, 2023, pp. 164-169, doi: 10.1109/ICCA51723.2023.10182211.
5. H. Banerjee and S. Yadav, "Energy-efficient Security Technique Implementation for Selective Forwarding Attack in WSN," 2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), Jaipur, India, 2023, pp. 1-10, doi: 10.1109/IEMECON56962.2023.10092329.
6. J. l. A and J. P. P M, "A Survey on Security and Network management of SDWSN with ML Technique," 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India, 2023, pp. 241-246, doi: 10.1109/ICISCoIS56541.2023.10100545.
7. K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi and M. Mustaqim, "Internet of Things (IoT) for Next-Generation Smart Systems: A Review of Current Challenges, Future Trends and Prospects for Emerging 5G-IoT Scenarios," in IEEE Access, vol. 8, pp. 23022-23040, 2020, doi: 10.1109/ACCESS.2020.2970118.
8. I. Butun, P. Österberg, and H. Song, "Security of the Internet of Things: Vulnerabilities, Attacks, and Countermeasures," in IEEE Communications Surveys & Tutorials, vol. 22, no. 1, pp. 616-644, First quarter 2020, doi: 10.1109/COMST.2019.2953364.
9. A. Ahmed, S. Abdullah, M. Bukhsh, I. Ahmad and Z. Mushtaq, "An Energy-Efficient Data Aggregation Mechanism for IoT Secured by Blockchain," in IEEE Access, vol. 10, pp. 11404-11419, 2022, doi: 10.1109/ACCESS.2022.3146295.
10. A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "CNN-LSTM: Hybrid Deep learning Neural Network for Network Intrusion Detection System," in IEEE Access, vol. 10, pp. 99837-99849, 2022, doi: 10.1109/ACCESS.2022.3206425.
11. M. N. M. Bhutta and M. Ahmad, "Secure Identification, Traceability and Real-Time Tracking of Agricultural Food Supply During Transportation Using Internet of Things," in IEEE Access, vol. 9, pp. 65660-65675, 2021, doi: 10.1109/ACCESS.2021.3076373.
12. S. Pundir, M. Wazid, D. P. Singh, A. K. Das, J. J. P. C. Rodrigues, and Y. Park, "Intrusion Detection Protocols in Wireless Sensor Networks Integrated to Internet of Things Deployment: Survey and Future Challenges," in IEEE Access, vol. 8, pp. 3343-3363, 2020, doi: 10.1109/ACCESS.2019.2962829.
13. Z. Cui et al., "A Hybrid BlockChain-Based Identity Authentication Scheme for Multi-WSN," in IEEE Transactions on Services Computing, vol. 13, no. 2, pp. 241-251, 1 March-April 2020, doi: 10.1109/TSC.2020.2964537.
14. M. A. Al-Naeem, "Prediction of Re-Occurrences of Spoofed ACK Packets Sent to Deflate a Target Wireless Sensor Network Node by DDOS," in IEEE Access, vol. 9, pp. 87070-87078, 2021, doi: 10.1109/ACCESS.2021.3089683.
15. O. R. Ahutu and H. El-Ocla, "Centralized Routing Protocol for Detecting Wormhole Attacks in Wireless Sensor Networks," in IEEE Access, vol. 8, pp. 63270-63282, 2020, doi: 10.1109/ACCESS.2020.2983438

АШЫҚ КІЛТ ИНФРАҚҰРЫЛЫМЫНДАҒЫ БИОМЕТРИЯҒА НЕГІЗДЕЛГЕН АУТЕНТИФИКАЦИЯ ӘДІСІ

Шаметова Г.Қ.
email: gauharshametova@gmail.com
Әл-Фараби атындағы Қазақ ұлттық университеті, Қазақстан

Андатпа. *Бұл мақалада ашық кілттер инфрақұрылымы үшін биометрияға негізделген аутентификация әдісін ұсынады. Аутентификация әдісі Шамир құпияларды ілестіру және соқыр қол қою әдістерін пайдаланады. Биометриялық деректер бір үлес ретінде сақталады. Соңғы тәсіл биометриялық үлгілердің қауіпсіздігін жақсартуға көмектеседі және оларды ашық кілттер инфрақұрылымы шеңберінде пайдалануға мүмкіндік береді.*

Кілт сөздер: ашық кілттер инфрақұрылымы, аутентификация, биометрия, Шамир сұлбасы

Кіріспе

Заманауи биометрияға негізделген аутентификация жүйелері [9] ең жоғары дәлдік деңгейі мен жоғары өнімділікті қамтамасыз ететіндей қылып жасалынған. Бұл тәсіл жабық ортада жұмыс істейтін жүйелерде жақсы қызмет атқарады. Алайда пайдаланушылар талаптарына сәйкес Интернет сияқты ашық ортада жұмыс істеуге көшу кезінде жағдай қатты өзгереді. Пайдаланушы талаптары өте маңызды болғанмен, қауіпсіздік аспектісі одан да маңызды мәселе. Қазіргі уақытта стандартталған жаһандық биометриялық жүйе жоқ, дегенмен кейбір зерттеу топтары биометрия әдістері мен ашық кілттер инфрақұрылымын (PKI) бірге пайдалану мәселесін сипаттайтын бірнеше маңызды мақалалар жариялады [5,7,10].

Ең алғашқы жұмыстардың бірі Мексикалық зерттеушілер жасаған шағын көлемде биометриялық аутентификацияны қолданатын жүйе жобасы [8]. Олардың шешімі сандық сертификаттарға (X.509) түрлендірулер енгізіп, бірегей мобильді сертификаттар орталығын (CO) құруға негізделген. Биометрияны тексеру және тану модулі ашық бастапқы жобадан алынған [3]. Бұл жүйенің ең үлкен кемшілігі сертификат өлшемінің таңдалған криптографиялық әдіске тәуелді болуы.

Келесі жұмыс АҚШ зерттеушілерінің PKI-мен биометрияны қалай пайдалану керектігін сипаттайтын құнды мақаласы [2]. Олар биометриялық жүйелерді қабылдаудың маңызды кезеңі болған қайтарылатын биотокендерді ұсынды. Бұл мәселеге жаңа көзқарас сеанс биотокендерін құруға ықпал етті, олардың бұзылуы түпнұсқалық биометриялық ақпаратты алу мүмкіндігіне әсер етпейді. Сонымен қатар биожүйелер төзімді болатын екі шабуылды ұсынды.

Ақырында, оңтүстік кореялық зерттеушілер озық биометриялық үлгілерді өңдеуге негізделген өз шешімін ұсынды [10]. Олар биометриялық мүмкіндіктерді клиенттер сертификатының ашық кілтімен байланыстырды, бұл оларға цифрлық сертификат пен өңделген мәндерді бір уақытта қайтарып алуға мүмкіндік берді. Оның үстіне ұсынылған жүйе сертификатты өзгертуді де, кеңейтуді де талап етпейді.

Жүйеге шолу

Біздің жүйе тәжірибе мен тиісті жұмыстардың нәтижелері негізінде құрылған [2, 8, 10]. Біз бірнеше болжамдарды анықтадық, оларға біздің жүйе сәйкес келеді. Бұл бізге қауіпсіздіктің күтілетін деңгейін қамтамасыз етуге мүмкіндік берді. Олар төменде көрсетілген.

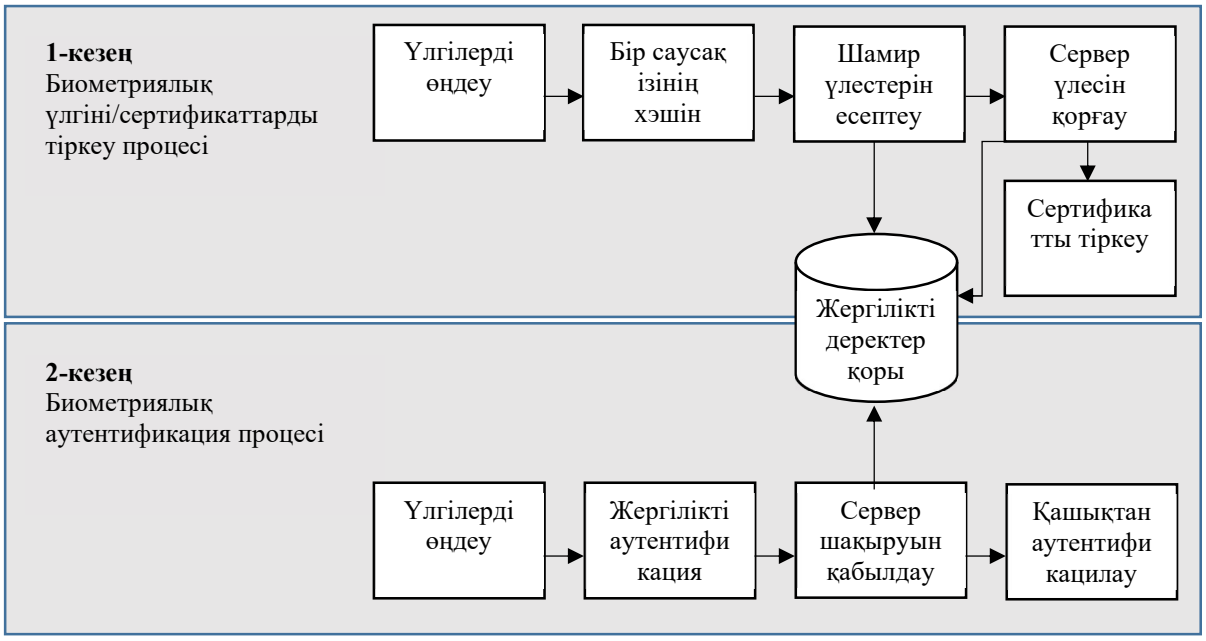
1. Биометриялық аутентификация процесі клиенттің сандық сертификатының ашық кілтімен тікелей байланысты болуы керек.

2. Клиент сервер сұрауларына жауап беру үшін қажет деректердің ең аз мөлшерін ғана сақтауы керек.

3. Серверде сақталған өңделген биометриялық деректерді бұзу бастапқы биометриялық үлгіні қалпына келтіруге мүмкіндік бермеуі керек.

4. Өңделмеген биометриялық үлгі емес, тек хэш мәндері ғана кері қайтарылуы керек.

Жоғарыда сипатталған критерийлерге қол жеткізу жүйенің биометриялық деректерді нақты үлгісіз тексеру мүмкіндігін береді.



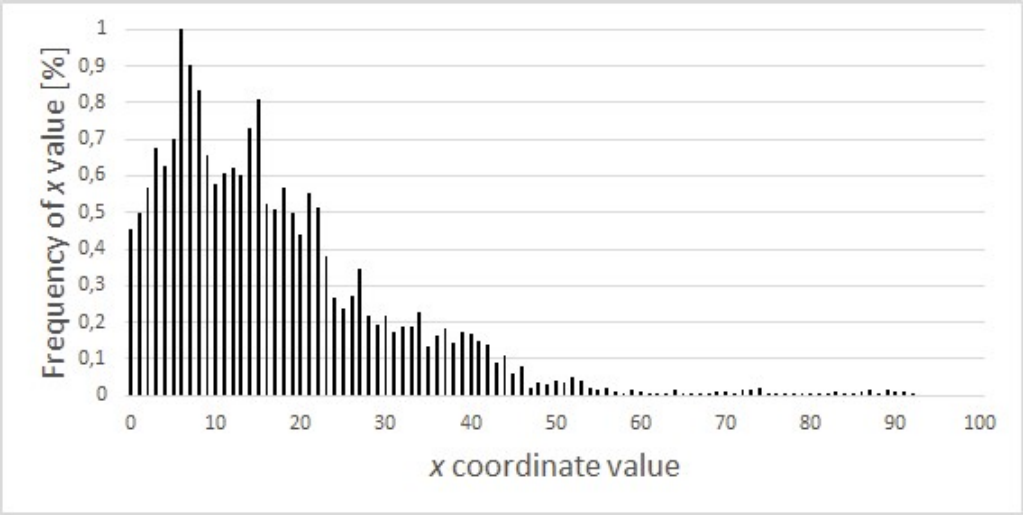
1-суретте жүйемізге шолу жасалған. Шешімде деректерді өңдеудің екі кезеңі көрсетілген. 1-кезең жүйедегі кіріс деректерді оқудан бастап, ұсақ-түйек құрылымдарды өңдеуге, Шамир үлестерін құру және сандық куәлікті тіркеуге дейінгі өңдеу қадамдарын көрсететін блоктарды қамтиды. 2-кезең аутентификация процесі қалай жұмыс істейтінін көрсетеді. Жүйе сондай-ақ ортақ жергілікті дерекқорды қамтиды, ортақ дерекқор мазмұны екі кезеңде де пайдаланылатынын білдіреді. Жалпылай айтқанда, бірінші кезең биометриялық үлгіні немесе сертификатты тіркеу процесін қамтиды, екіншісі тек биометриялық аутентификация процесін қамтиды.

Үлгілерді өңдеу

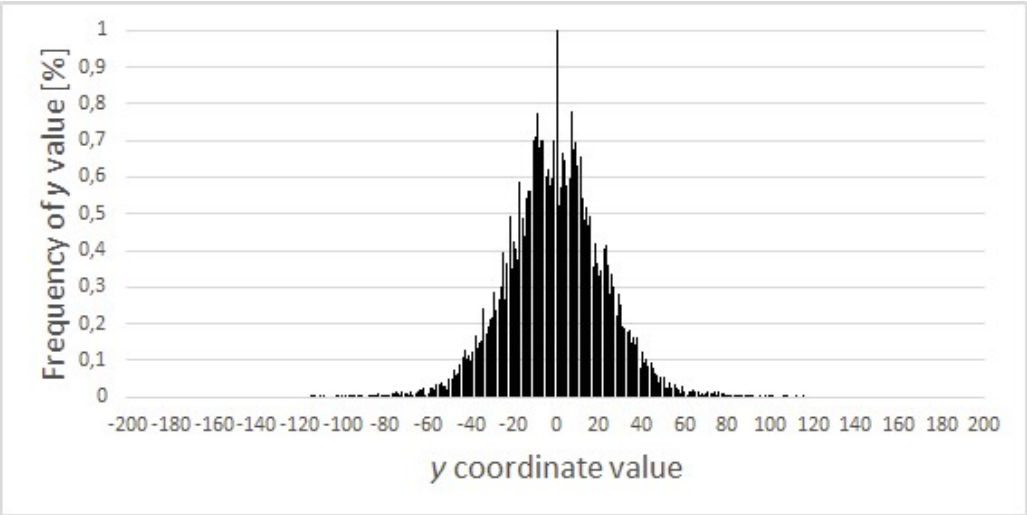
Үлгілерді өңдеу «Жергілікті құрылымдардың ішкі жиындарын пайдалана отырып, саусақ ізі хэштерін қорғау» [9] мақаласында егжей-тегжейлі сипатталған. Бүкіл процесс ұсақ-түйектерді анықтаудан, үшбұрыштарды анықтаудан, үшбұрыштарды аударудан,

үшбұрыш комбинацияларының үштіктерінен, синтетикалық саусақ іздерін генерациялаудан, кванттаудан және соңғы хэш мәндерін генерациялаудан тұрады. Жалпы алғанда, жоғарыда көрсетілген қадамдар жақсы сипатталған, бірақ кванттау белгілі бір мағынада қалдырылды және сегмент диапазондары туралы маңызды ақпарат айтылмады. Біздің жұмыста қосымша зерттеулер жүргізілді және шығыс мәндерін теңестіруге мүмкіндік беретін қолайлы диапазондарды табылды.

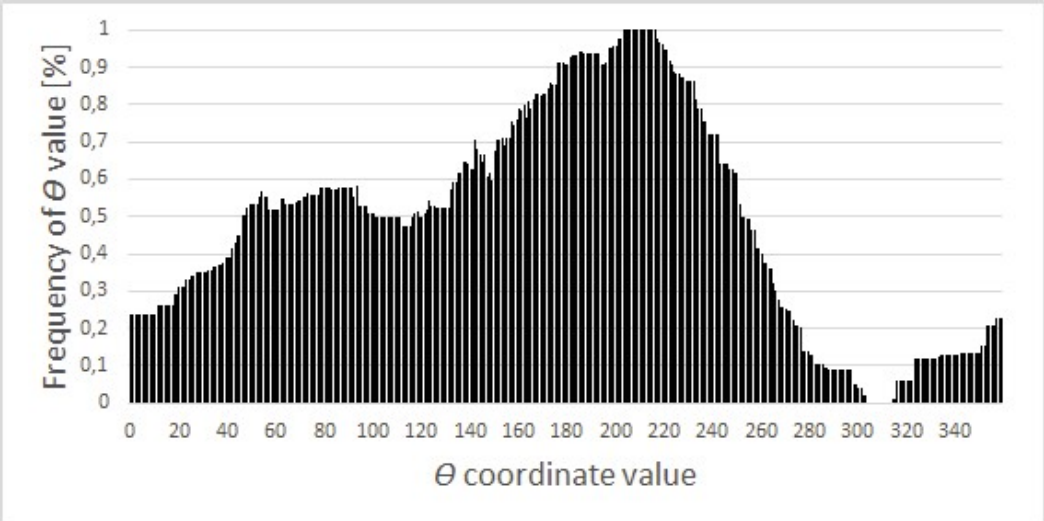
Біріншіден, біз саусақ іздерінің суреттерін жинап, уақытша мәліметтер базасын жасадық. Содан кейін біз кездейсоқ 100 элементті таңдадық. Осыдан кейін ұсақ нәрселерді анықтадық. Содан кейін біз үшбұрыштарды тұрғызып, аталған құжатта сипатталғандай координаттар жүйесіне айналдырдық. Біздің пайдаланушы қолданбамыз әр x, y, θ үшін қайталану санын санайды және барлық мәндерді <0;1> диапазонына дейін қалыпқа келтіреді. Бұл нормаланған деректер бір орташа нәтижеге біріктірілген, ол 2, 3 және 4-суреттерде әрбір координат үшін жеке көрсетілген.



2-сурет. x координатының мәндерінің орташа таралуы



3-сурет. y координатының мәндерінің орташа таралуы



4-сурет. θ координатының мәндерінің орташа таралуы

Көріп отырғанымыздай, x және y координаталары үшін таралу заңдылығы бар. Мәндерді негізінен диаграмманың бірінші жартысында (<0; 50>) x координатасы үшін және y координатасы үшін <25;25> диапазонында орналастыру керек, себебі ол сәйкес мәндердің ең көп мөлшерін қамтиды. Ұсақ бөліктердің айналу бағытын білдіретін θ жағдайында нәтиже мүлдем басқаша болады. Бұл мәндер тұрақты емес болғандықтан ауқымды шығыс толығымен теңестірілетін дұрыс бөліктерге бөлу мүмкін емес. Біз саусақ іздерінің шағын тобын таңдадық және диапазондарды мүмкіндігінше эксперименттік жолмен анықтадық.

Бір саусақ ізі үшін хэшті анықтау

Алдыңғы бөлімдегі өңдеу үлгісі мыңдаған хэш мәндерін жасайды. Үлкен топтан кездейсоқ түрде бір өкілді мәнді таңдау іс жүзінде мүмкін емес. Біздің шешіміміз жиындардың қиылысуларына негізделген.

Біз бірнеше саусақ ізі суретін түсіреміз және бір мәнді немесе қайталанатын мәндер тобын алған кезде қиылысатын жерлерді іздейміз. Біздің сынақтар бұл процестің салыстырмалы түрде жылдам екенін көрсетті. Алгоритм төменде сипатталған.

1. Саусақ іздерінің хэштер жинағын жүктеу және оны *B* негізгі жинағы ету.
2. Басқа саусақ іздерінің хэштерін F жинағын жүктеу.
3. Екі жиынға жататын хэштерді қамтитын *E* жиынының ортақ элементтерін құрастыру (∀e ∈ E, e ∈ B ∧ e ∈ F).
4. *E* жинағы бос болса (саусақ ізі хэштері жинағы қате пішімделген болуы мүмкін) 2-қадамға өту. Басқа жағдайда 5-қадамға өту.
5. *E* жиынын жаңа *B* негізгі жиынды жасау.
6. Егер негізгі жиынтықтың негізгілігі 1-ге тең болса, қиылысу процесін аяқтап, 7-қадамға өтіу. Басқа жағдайда 2-қадамға өту.
7. Кездейсоқ S_H санын анықтау.
8. Негізгі хэш мәнін анықтау (1-формула).

$$h(H) = h(h(H)||S_H) \quad (1)$$

мұндағы H – таңдалған хэш мәні, h – SHA-256 (қауіпсіз хэш алгоритмі) хэш функциясы.

Шамир үлестерін есептеу

Аутентификация үлгісінде жалғыз хэш мәнін пайдалану мүмкін емес. Оны зиян келтіруші оңай бұзып алып, қолдана алады. Біздің жүйеде деректерді хэшті қажет ететін және сәйкес құпияны білетін адам ғана сервер тапсырмасын сәтті орындай алатындай етіп түрлендіреді. Біз Шамирдің құпия бөлісуін [12,14,15] орынды шешім деп шештік.

Біріншіден, егер процесс қайталанатын болса, қажетті үш коэффициентті құру қажет.

$$\begin{aligned} a_0 &= secret \\ a_1 &= H\,M\,AC_S\,(kbio) \\ a_2 &= H\,M\,AC_S(a_1) \end{aligned} \tag{2}$$

мұндағы құпия(лар) пайдаланушы кілті, kbio - жалғыз хэш мәні. Құпия бүкіл жүйенің маңызды бөлігі болып табылады — біз кванттау процесімен байланысты [9] ықтимал осалдықты таптық.

Біз SHA-512 хэш функциясы бар HMAC (кілттелген-Хэш хабарламасының аутентификация коды) алгоритмін қабылдадық.

$$S(s, x, a, p) = s + \sum_{j=1}^{2} a_j\, x^j \; mod\; p \tag{3}$$

мұндағы S – Шамирдің үлесі, x – үлестің идентификаторы (0,1 немесе 2), a – коэффициенттер жиыуы және p – үлкен жай сан. 3-формула S_0, S_1 және S_2 үлестерін есептеу үшін қолданылады.

Сервер үлестерін қорғау

Сервердегі кейбір мәндерді бірнеше қызмет жеткізушілері пайдаланатындықтан сервер үлестерін (S_2) өңдеу қауіпті. Себебі қорғалуы тиіс деректерге қауіп төнуі мүмкін. Сонымен қатар, өңделмеген мәнді Z_N тобында бір мәнді түрлендірулер үшін пайдалану мүмкін емес — бұл аутентификация мәселелерін тудырады. Біз сертификаттың ашық кілтін сервер үлесімен байланыстыратын қауіпсіз сервер үлесін жасаймыз. S_2 үшін кездейсоқ индексті (RI) есептейміз

$$GCD(RI * S_2, n) = 1. \tag{4}$$

мұндағы GCD – ең үлкен ортақ бөлгіш, n – клиенттің ашық кілт модулі.

Соңында біз төменде көрсетілгендей қауіпсіз сервер үлесін (S_{SS}) жасаймыз.

$$S_{SS} = (RI * S_2)^{E_k} \tag{5}$$

мұндағы E_k - клиенттің ашық кілті.

Сипатталған әдіс соқыр цифрлық қолтаңба әдісіне негізделген [1]. Ұсынылған тәуелділік кілттер жұбымен бір уақытта сервер үлесін қайтарып алуға мүмкіндік береді. Ақырында, нәтиже клиенттің сандық сертификатталған ашық кілтімен шифрланады.

Сертификатты тіркеу

Дайындықтың бірінші бөлігіндегі соңғы қадам – екінші Шамир үлесі (S_1) бар, маңызды кеңейтімі анықталған клиент үшін X.509 сертификатын тіркеу. Бұл өте маңызды, себебі S_1 жергілікті аутентификация процесі кезінде қажет болады.

Жергілікті деректер қорын сақтау

Бұрын айтылғандай, жүйе мүмкіндігінше аз деректерді сақтауы керек. Біз тек P үлкен жай санның мәнін, RI кездейсоқ индекс мәнін және таңдалған хэштің мәнің сақтаймыз. Қажет болса, жергілікті дерекқорда клиент сертификатының саусақ ізі немесе сертификаттың ашық кілті болуы мүмкін.

Жергілікті аутентификация

Жергілікті аутентификация серверге қашықтан жасалатын аутентификациядан бұрын болатын процесс. Бұл биометриялық деректерді өңдеу классикалық шақыруға жауап беру әдісі сияқты жылдам емес болғандықтан енгізілді. Сонымен қатар, уақыт белгілерін пайдалану мүмкін емес (олардың мерзімі аяқталады). Басында біз бір саусақ ізі хэшін анықтау бөлімінде көрсетілгендей жалғыз хэш мәнін аламыз. Жүйе сонымен қатар S_1 үлесін, RI кездейсоқ индексін, үлкен P саны мен клиент құпиясын S қажет етеді. Құпия мәнді клиент өзі енгізуі керек. Осыдан кейін біз Шамирдің құпия бөлісу коэффициенттерін есептеп, S'_0, S'_2, S'_3 үлестерін аламыз. Егер анықталған S'_1 сандық сертификат кеңейтімінде сақталған S_1 мәніне тең болса, біз клиенттің өзі мәлімдеген адам екеніне сенімді бола аламыз.

Сервер шақыруын қабылдау

Таңдалған жалғыз хэш мәні дұрыс екеніне сенімді болсақ, екі жақты аутентификация процесін бастауға болады. Біз бұл бөлікті мұнда сипаттамадық, себебі оның орнына SSL/TLS4 протоколын пайдалануға болады. Қауіпсіз байланыс арнасында біз биометриялық аутентификация процесін бастаймыз. Толық аутентификация протоколы төменде сипатталған.

1. Сервер кездейсоқ R_V сынақ мәнін жасайды.
2. Сервер төменде сипатталғандай R^*_V мәнін анықтайды

$$R_V^* = R_v * S_{SS} \, mod \, n \tag{6}$$

мұндағы S_{SS} – қауіпсіз сервер үлесі, n – клиенттің ашық кілт модулі.

3. Сервер m хабарламасын жібереді

$$\mathrm{m} = (R_V^* \,||\, T_B)^{D_S} \tag{7}$$

мұнда T_B сервердің уақыт белгісі, D_S сервердің жабық кілті болып табылады.

4. Клиент m хабарламасын алады және уақыт белгісінің жарамдылығын тексереді.
5. Клиент шығару процедурасын бастайды.

Клиент R_V мәнін шығаруға және оны серверге қайтаруға міндетті. Бұл бөлік қашықтан аутентификациялау деп аталады.

Қашықтан аутентификациялау

Клиент m хабарламасы қабылданғаннан кейін және R^*_V мәні алынғаннан кейін шығару процедурасын бастайды. Бастапқы мән қалпына келтірілген S_2 үлесі негізінде

есептеледі. Бүкіл процесс ZN тобындағы қауіпсіз сервер үлесіне және бірнеше RSA шифрлауларына кері мәнді анықтауды талап етеді.

1. Клиент шығару процедурасын бастайды:

- R'_V анықтау

$$R'_V = (R^*_V)^{D_K} \, mod \, n \tag{8}$$

- Үлгілерді өңдеу және бір саусақ ізі хэшін анықтау бөлімдерінде сипатталғандай жалғыз хэш мәнін алу;
- Шамир үлесін есептеу;
- Z_N тобындағы қауіпсіз үлестің кері мәнін анықтау

$$S_s = (S_2 * RI)^{-1} \tag{9}$$

мұндағы S_2 – Шамир үлесінің бастапқы мәні, RI – жергілікті дерекқорда сақталған кездейсоқ индекс мәні.

- R'_V соқырлығын жою

$$R^{**}_V = R'_V * S_S \, mod \, n \tag{10}$$

- өзіңіздің ашық кілтіңізді пайдаланып R^{**}_V шифрлаg және R_{VK} мәнін алу. R_{VK} - шығарылған сынақ мәні.

2. Клиент m хабарламасын осылайша жібереді

$$m = (R_{VK} \, || \, T_A)^{D_K} \tag{11}$$

3. Сервер m хабарламасын алады, уақыт белгісінің мәнін тексереді және R_{VK} мен R_V салыстырады. R_{VK} мен R_V тең болса, аутентификация процесі сәтті аяқталды, клиентте R_V шығаруға мүмкіндік беретін кейбір деректер алынады.

Жоғарыда келтірілген процесті 12-формуламен көрсетуге болады.

$$R_{VK} = [[(RI * S_2)^{E_K} * R_V \, mod \, n]^{D_k} * (RI * S_2)^{-1} \, mod \, n]^{E_K} \, mod \, n \tag{12}$$

Қорытынды

Бұл мақала қорғалған үлгілерде жұмыс істейтін ашық кілт инфрақұрылымы үшін биометрияға негізделген аутентификация жүйесін ұсынады. Біздің шешіміміз аз ғана шешімдердің бірі ретінде биометриялық мүмкіндіктер деңгейінде қауіпсіздік механизмін қамтамасыз етеді. Мақалады мысал келтірілген басқа тәсілдерге қарағанда, біз криптографиялық мағынада жалғыз қауіпсіз хэш мәнін және соқыр цифрлық қолтаңбалармен Шамир құпиясын бөлісетін үлгіні өңдеуді қолдандық. Сондай-ақ, жүйе құрылымы хаттаманың кейбір мәліметтерімен ұсынылды.

Болашақ жұмыстар үлгіні өңдеу өнімділігін жақсартуға және табылған ұсақ бөліктер санын көбейте алатын бірнеше алгоритмдерді қосуға арналады.

Пайдаланған әдебиеттер

1. Бишоп В., Боулт Т.Э., Шайрер В.Дж. За пределами PKI: инфраструктура биокрипто-графических ключей, 2010. – http://www.wjscheirer.com/papers/wjs_spb2011_bki.pdf. (08.05.2023)
2. Болле Р.М., Сениор А.В., Ратха Н.К., Панканти С. Мелочи отпечатков пальцев : конструктивное определение, 2002. – https://www.research.ibm.com/people/a/aws/documents/papers/BolleMinutiae.pdf.(10.06.2023)
3. Круз-Кортес Н., Эртаул Л., Мартинес-Сильва Г., Родригес-Энрикес Ф. О создании сертификатов X.509v3 с биометрической информацией, 2007 - http://delta.cs.cinvestav.mx/~francisco/certFRH_SAM07.pdf. (01.07.2023)
4. Эффланд Т., Шнеггенбургер М., Шулер Дж. Защита хэшей отпечатков пальцев с использованием подмножеств локальных структур, 2014 - http://www.cse.buffalo.edu/~hartloff/papers/SPIE2014.pdf. (09.08.2023)
5. Cappelli R., Jain A.K., Maio D., Maltoni D., Wayman J.L.: FVC2004: Third Fingerprint Verification Competition, 2004. – http://www.cse.msu.edu/ biometrics/Publications/Fingerprint/FVC/ Maioetal_FVC2004_ICBA04.pdf.(18.06.2023)
6. Han-Ul J., Heung-Kyu L.Biometric-PKI Authentication System Using Fingerprint Minutiae, 2013. – http:// dx.doi.org/10.4236/jcc.2014.24004. (20.08.2023)
7. Network Working Group, Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile, 2008. – https://www.ietf.org/rfc/ rfc5280.txt. (03.09.2023)
8. R. Vidya, N. Nagaveni. Evaluation of a Perturbation-Based Technique for Privacy Preservation in Multi-Party Clustering Scenario /Information Sciences – Elsevier 2012. Vol. 232 N.20 P. 437-448
9. Shamir A. How to share a secret /A. Shamir// Communications of the ACM – N.Y., 1979. – Vol. 22 N.11 – P. 612-613, – http://cs.jhu.edu/~sdoshi/crypto/papers/shamirturing.pdf. (01.06.2023)
10. M. Mignotte, How to share a secret, / T. Beth // Cryptography-Proceedings of the Workshop on Cryptography. – Feuerstein., 1982. Vol. 149. – P. 371–375.

СТРУКТУРА АЛГОРИТМА ХЕШИРОВАНИЯ HAS03

Алгазы К.Т, Дюсенбаев Д.С.
e-mail: kunbolat@mail.ru, dimash_dds@mail.ru
Институт информационных и вычислительных технологий
КН МОН РК, Казахстан

Аннотация. *В статье описана структура разработанного алгоритма хеширования HAS03. Также обсуждается актуальность хеш-алгоритмов и общие требования к ним. Большинство современных алгоритмов хеширования основаны на блочных алгоритмах и структуры губки. Рассматриваемый алгоритм относится ко второй группе, в которых структура состоит из внутреннего состояния фиксированной длины и этапов впитывания и выжимания. Размер внутреннего состояния алгоритма составляет 1024 бита, этапы впитывания и выжимания выполняются с помощью функции f. В работе описаны преобразований, используемые в функции f и их результаты действия поэтапно отражены на реальном примере. Одним из обязательных свойств криптографических алгоритмов хеширования является наличии лавинного эффекта. Результаты исследования показали, что данный алгоритм хеширования полностью соответствует критериям лавинного эффекта. Алгоритм является криптостойким, если вероятность нахождения коллизии и прообраза очень мала. Поэтому, исследования в этом направлении активно продолжаются.*

Введение

Основное назначение алгоритмов хеширования - преобразовать исходный текст произвольной длины в текст определенной фиксированной длины. Обычно длина хеш-значения намного короче исходного текста и составляет от 128 до 512 битов. Алгоритмы хеширования разрабатываются как блочные шифры, которые удовлетворяют условию лавинного эффекта, что означает, что при изменении любого бита или байта исходной информации результирующее значение хеш-функции должно быть полностью изменено. Если это не учитывать, злоумышленникам будет легче осуществить успешную атаку. Хеширование – это односторонний процесс, поэтому вы никогда не сможете вернуться к исходной информации. Качественный алгоритм хеширования должен быть настолько сложным, чтобы он не мог генерировать одно и то же хеш-значение для двух разных наборов данных. Обнаружение такой пары называется коллизией. Хеш-функция называется стойкой к коллизиям, если вероятность нахождения коллизии очень (ничтожно) мала или для самых современных компьютеров для ее вычисления требуется очень большое время. Также не должно быть возможности определить прообраз результирующего значения хеш-функции. Как правило, для хорошей хеш-функции должны выполняться следующие условия [1,2]:

- одно и то же значение хеш-функции получено из одного и того же сообщения, т.е. хеш-функция является детерминированной;
- высокая скорость вычисления хеш-значения любого сообщения;
- невозможно найти исходное сообщение, зная его хеш-значение;
- невозможно найти два разных сообщения с одинаковым хеш-значением;

– небольшое изменение исходного сообщения должно полностью изменить хеш-значение, а новое и исходное хеш-значения не должны быть связаны между собой.

На сегодняшний день криптографические хеш-функции имеют очень широкое применение. Это незаменимый инструмент, используемый для выполнения различных задач, таких как аутентификация, проверка целостности данных, защита файлов, обнаружение вредоносных программ и поиск по базам данных. Хеширование является одним из наиболее распространенных методов хранения паролей. Если пользователь забудет пароль, вам придется воспользоваться функцией восстановления. Однако в этом случае пользователь не получает старый пароль, потому что онлайн-сервис не сохраняет пароли пользователя в первоначальном виде. Вместо этого он хранит их хеш-значения. Серверу не известно значение пароля. Хеширование также используется для защиты целостности сообщения, отправляемого по сети. Для этого вместе с текстом отправляется его хеш-значение. Получатель вычисляет значение хеша и сравнивает с полученным хеш-значением. В случае несоответствия целостность считается нарушенной. Использование хеш-функций широко распространено при организации обмена документами с электронной подписью. В этом случае файл хешируется и значение хеш-функции подписывается электронной цифровой подписью [3-6].

Разработка алгоритма хеширования

Алгоритм хеширования выполняется в рабочем диапазоне длиной 1024 бита (то есть размер внутреннего состояния равен 1024 бита). Исходный открытый текст делится на блоки длиной 512 бит. Если общая длина открытого текста не кратна 512, т.е. длина последнего блока меньше 512 бит, то производится его дополнение. В качестве дополнения используется битовая последовательность, на первой и последней позиции которой находятся единицы, а середина заполнена нулями. В основе алгоритма лежит конструкция губки, общая схема которой представлена на рисунке 1. Внутреннее состояние разделено на две части. Разработанный алгоритм использует внутреннее состояние, состоящее из двух 512-битных частей A и V.

На этапе впитывания очередной блок открытого текста A_i с помощью операции xor объединяется с обоими частями внутреннего состояния алгоритма и записывается вместо первой части, а вторая часть записывается без изменений: $A_i' = A_i \oplus A_{i-1} \oplus V_i$, $V_i' = V_{i-1}$, $i = \overline{0, n-1}$. В качестве начального внутреннего состояния можно выбрать последовательность состоящую только из нулей (в общем случае, можно выбрать любую фиксированную последовательность) длиной 1024 бита. С помощью функции f преобразуем данные, полученную в результате этих операций. После того, как все блоки открытого текста обработаны, переходим на этап выжимания. На этом этапе входными данными функции f является внутреннее состояние $A_{n-1} \parallel V_{n-1}$, а первые 64 бита выходных данных выбираются в качестве хеш-значения. Это повторяется до тех пор, пока не будет получено хеш-значение требуемой длины. То есть, чтобы получить 256- или 512-битное хеш-значение, нужно повторить эту процедуру 4 или 8 раз соответственно.

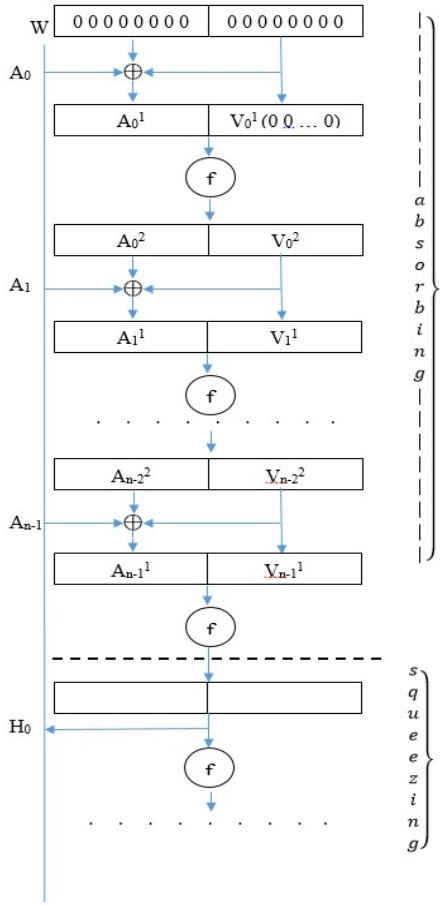


Рис. 1 Структура губки

Функция f

Структура внутреннего состояния, используемого функцией f, представляет собой квадратную матрицу размером 4×4, каждый из элементов которой состоит из 64-битных слов (Рисунок 2). W_i, C_i $i = \overline{0,7}$ – 64-битные слова, W_i – данные из открытого текста (в общей схеме алгоритма - A_j), C_i – вторая часть, участвующая в преобразовании (в общей схеме V_j).

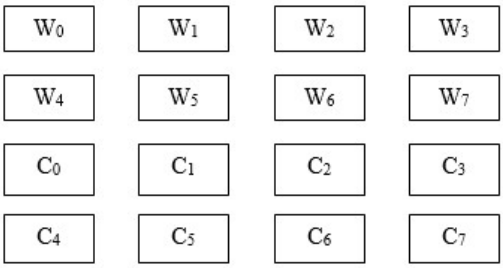


Рис. 2 Структура внутреннего состояния, используемого функцией f

Функция f состоит из преобразований X, S, R и P, причём X выполняется дважды. Опишем каждое из этих преобразований.

X-преобразование. Выполняет сложение соответствующих элементов первого и второго столбца матрицы с помощью операции xor и помещает результат в первый столбец, т.е. $W_0' = W_0 \oplus W_1$, $W_4' = W_4 \oplus W_5$, $C_0' = C_0 \oplus C_1$, $C_4' = \mathrm{C}_4 \oplus \mathrm{C}_5$. Это же преобразование используется для получения элементов второго и третьего столбца: $W_1' = W_1 \oplus W_2$, $W_5' = W_5 \oplus W_6$, $C_1' = C_1 \oplus C_2$, $C_5' = \mathrm{C}_5 \oplus \mathrm{C}_6$, $W_2' = W_2 \oplus W_3$, $W_6' = W_6 \oplus W_7$, $C_2' = C_2 \oplus C_3$, $C_6' = \mathrm{C}_6 \oplus \mathrm{C}_7$. Четвертый столбец оставим без изменении.

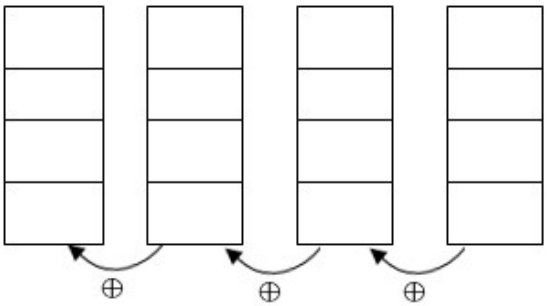


Рис. 3 Схема X-преобразования

S-преобразование. Каждое из значений W_i, C_i $i = \overline{0,7}$ состоит из 8 байтов, а их общие количество равно 16. Тогда общее количество байтов, обрабатываемых функцией f, равно 128. В результате преобразования S эти 128 байтов будут заменены на другие байты с использованием S-блока, то есть $S: \{0,1\}^{1024} \to \{0,1\}^{1024}$.

После S-преобразования выполняется X-преобразование для первого и четвертого столбцов, и результат записывается как новые элементы четвертого столбца: $W_3'' = S(W_0') \oplus S(W_3')$, $W_7'' = S(W_4') \oplus S(W_7')$, $\mathrm{C}_3'' = S(\mathrm{C}_0') \oplus S(\mathrm{C}_3')$, $\mathrm{C}_7'' = S(\mathrm{C}_4') \oplus S(\mathrm{C}_7')$.

R-преобразование. Обозначим каждое из 16 слов внутреннего состояния как X_n, $n = \overline{0,15}$. Тогда при $n = 0,1,2,3$ выполняется операция $SHRL(X_n)$, , а в остальных случаях выполняется операция $ROTRL(X_n)$. Здесь $SHRL(X_n)$ – операция логического сдвига 64-битного аргумента на $n + 3$ бита вправо, если n – чётное, и влево, если n – нечётное. Аналогично, $ROTRL(X_n)$ – операция циклического сдвига 64-битного аргумента на $n + 3$ бита вправо, если n – чётное, и влево, если n – нечётное.

Таблица 1

Таблица перестановки для P-преобразования

P-преобразование															
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
0	16	32	48	64	80	96	112	1	17	33	49	65	81	97	113
16	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
2	18	34	50	66	82	98	114	3	19	35	51	67	83	99	115
32	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**
4	20	36	52	68	84	100	116	5	21	37	53	69	85	101	117
48	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**	**61**	**62**	**63**
6	22	38	54	70	86	102	118	7	23	39	55	71	87	103	119
64	**65**	**66**	**67**	**68**	**69**	**70**	**71**	**72**	**73**	**74**	**75**	**76**	**77**	**78**	**79**
8	24	40	56	72	88	104	120	9	25	41	57	73	89	105	121

80	**81**	**82**	**83**	**84**	**85**	**86**	**87**	**88**	**89**	**90**	**91**	**92**	**93**	**94**	**95**
10	26	42	58	74	90	106	122	11	27	43	59	75	91	107	123
96	**97**	**98**	**99**	**100**	**101**	**102**	**103**	**104**	**105**	**106**	**107**	**108**	**109**	**110**	**111**
12	28	44	60	76	92	108	124	13	29	45	61	77	93	109	125
112	**113**	**114**	**115**	**116**	**117**	**118**	**119**	**120**	**121**	**122**	**123**	**124**	**125**	**126**	**127**
14	30	46	62	78	94	110	126	15	31	47	63	79	95	111	127

P-преобразование. Основная цель преобразования – перемешивание входных элементов. Каждый байт входной последовательности перемещается на другое место в определенном порядке (Таблица 1).

Функция f состоит из 14-кратного повторения вышеуказанных преобразований, за исключением того, что R-преобразование не выполняется в последнем раунде.

Покажем результаты преобразований, используемых в функции f, на простом примере. Пусть M – двоичная последовательность длиной 512 бит, состоящая только из нулей. Обработка начинается с этапа впитывания.

Результаты исследование

Лавинный эффект – одно из обязательных свойств криптографических алгоритмов. Небольшое изменение входных данных блочных шифров и криптографических хеш-функций должно значительно изменить выходное значение (например, изменить половину выходных битов) [7-9].

Чтобы проверить лавинный эффект построенного алгоритма хеширования, рассмотрим приведенный выше пример, т.е. в качестве исходного текста выберем двоичную последовательность, состоящих только из нулей. Инвертируя один из нулевых битов в каждой позиции этой последовательности, будем получать новый 512-битовый текст. Вычисляем вероятность изменения k_i, $i = \overline{0,511}$, где k_i – это вероятность изменения i-го бита в выходном значении в случае инверсии i-го элемента входного значения по сравнению с первоначальным выходным значением. Чем ближе значение каждого k_i к 0,5, тем лучше лавинный эффект. В шестом раунде алгоритм показывает хороший результат (Рисунок 4).

Идеально или грамотно построенные хеш-функции обладают следующими свойствами: результаты преобразования одних и тех же сообщений всегда приводят к одному и тому же хеш-значению, то есть являются детерминированными; высокая скорость вычисления; исходное сообщение не может быть найдено по его хеш-значению; небольшое изменение сообщения полностью меняет значение хеш-функции; невозможно найти два сообщения с одинаковым хеш-значением.

Одной из универсальных атак, проводимых на все криптографические алгоритмы является полный перебор или атака методом грубой силы. Если хеш-значение сообщения M_1 равно $H(M_1)$, атакующему необходимо найти такое сообщение M_2, для которого выполняется $H(M_1) = H(M_2)$. Если длина полученного хеша равна n, то сложность этого метода составляет $O(2^n)$. Обычно при разработке алгоритмов этому условию уделяется особое внимание и алгоритм строится таким образом, чтобы обеспечивался требуемый уровень безопасности. В алгоритме HAS03 длина полученного хеш-значения составляет 256 и 512 бит, поэтому использование метода полного перебора неэффективно.

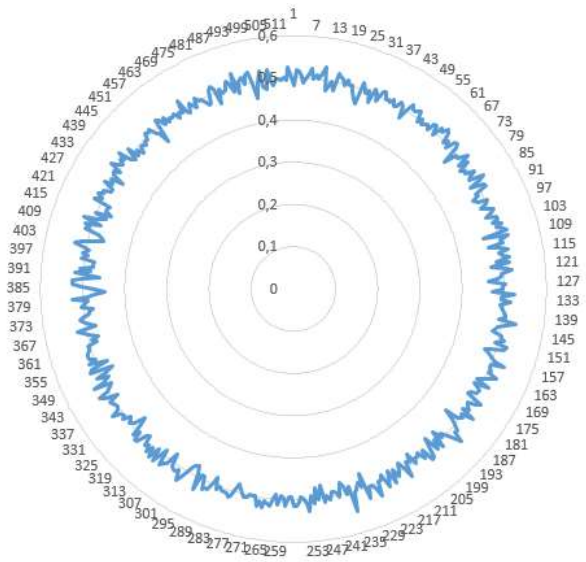


Рис. 4 Лавинный эффект 6-го раунда

Еще один универсальный метод атаки, используемый для хеш-функций, это «парадокс дней рождений». Этот метод криптографического анализа оценивает, сколько сообщений необходимо рассмотреть, чтобы обнаружить коллизию с вероятностью более 0,5. Чтобы найти коллизию, нужно сгенерировать два набора, каждый из которых содержит $2^{n/2}$ сообщений и вычислить их хеш-значения. Согласно парадоксу дней рождения, среди них есть сообщения с одинаковыми хеш значениями с вероятностью более 0,5. Однако для этого метода требуется большой объем памяти. Даже если длина хеша равна 256, современные компьютеры не смогут реализовать этот метод. Чтобы найти коллизию второго рода, необходимо произвести расчеты в размере не менее 2^{128}.

Благодарность

Исследовательская работа выполнена в рамках проекта грантового финансирования AP 14870719 – «Разработка и исследование алгоритмов пост-квантовой криптографии, основанных на хеш-функциях» Министерства науки и высшего образования Республики Казахстан.

Литература

1. B. Denton and R. Adhami // Modern Hash Function Construction 1 Dept. of Electrical & Computer Engineering, The University of Alabama in Huntsville, Huntsville, AL, USA, 2012.
2. Saif Al-Kuwari, James H. Davenport, Russell J. Bradford // Cryptographic Hash Functions: Recent Design Trends and Security Notions, URL: https://eprint.iacr.org/2011/565.pdf
3. Анин Б.Ю. Защита компьютерной информации / Б.Ю. Анин. СПб.: БХВ-Петербург, 2000. - 384 с.
4. Jonathan Katz, Jonathan Katz Introduction to Modern Cryptography: Principles and Protocols / Chapman & Hall/CRC Cryptography and Network Security Series, 2007. – 552 p.
5. Zegzhda P.D. Secure information systems based on a secure OS // Problems of information security. Computer systems. 1999. -No 1. P.96-99.

6. Зима В.М. Безопасность глобальных сетевых технологий /В,М. Зима, А.А. Молдовян, Н.А. Молдовян. Спб.: БХВ-Петербург, 2000. - 320 с.
7. Соколов А.В. Защита от компьютерного терроризма: Справочное пособие /А.В. Соколов, О.М. Степанюк. БВХ-Петербург: Арлит, 2002. -496 с.
8. Vergili I., Yücel M. D. Avalanche and Bit Independence Properties for the Ensembles of Randomly Cho-sen S-Boxes // Turk J Elec Engin.. - № 2. С. 137–145. -2001
9. Б. Шнайер, Прикладная криптография, 2-е изд: Пер. с англ. –Триумф, 2002. – 816 с.

ИНТЕГРАЦИЯ КИБЕРБЕЗОПАСНОСТИ В BIM: СОВРЕМЕННЫЕ ТРЕНДЫ И ПЕРСПЕКТИВЫ

Усатова Ольга Александровна[1,2], Пейішбек Ағахан Нұрлыбекұлы[2]
email: 222134@astanait.edu.kz
[1]*Институт информационных и вычислительных технологий КН МНВО РК, Алматы, Республика Казахстан*
[2]*Университет «Astana IT University», Астана, Республика Казахстан*

Аннотация. *Статья рассматривает актуальную тему интеграции кибербезопасности в область информационного моделирования зданий (BIM) в современной строительной и проектировочной индустрии. BIM, как интегрированная информационная система, предоставляет обширные данные о проектах, их структуре, стоимости и характеристиках, что делает безопасность данных критически важной. В статье анализируются основные проблемы, связанные с кибербезопасностью в BIM. Также рассматриваются перспективы интеграции кибербезопасности в BIM и ключевые тренды в области кибербезопасности, которые будут важными для отрасли в ближайшие годы.*

Введение

Ядром цифровой трансформации в строительной отрасли стали технологии информационного моделирования, известные как BIM (Building Information Model). Внедрение BIM-технологий началось в 2000-х годах и привнесло революцию в способ проектирования и строительства. BIM - это не просто трехмерная компьютерная модель здания, которая заменила старые двухмерные чертежи на бумаге.

BIM идет далеко за пределы простой геометрии здания. Эта технология интегрирует огромное количество информации о каждом элементе объекта, включая сведения о материалах, спецификации, стоимости, график строительных работ, функциональные и эксплуатационные характеристики, а даже условия окружающей среды [1]. Более того, изменение любого параметра здания автоматически влияет на все связанные с ним параметры и элементы.

В современном мире строительной индустрии BIM становится все более популярной и важной технологией. Она предоставляет возможность создания трехмерных моделей зданий и инфраструктуры, которые значительно упрощают процесс проектирования, строительства и управления объектами.

С развитием технологий информационного моделирования (BIM) в строительной индустрии стало очевидным, что обеспечение кибербезопасности становится приоритетной задачей. BIM не просто предоставляет трехмерные модели для улучшенного проектирования и строительства, но и интегрирует множество слоев информации, включая материалы, спецификации, стоимость, график работ, функциональные характеристики и даже окружающую среду. Это делает BIM зоной повышенного интереса для кибератак и подчеркивает необходимость интеграции кибербезопасности.

В этой статье мы рассмотрим современные тренды и перспективы интеграции кибербезопасности в BIM.

Основная часть.

В развитых странах более половины строительных организаций успешно применяют BIM-технологии. Этот внушительный уровень внедрения BIM в строительную индустрию, зачастую, объясняется внедрением опережающих требований со стороны государственных органов. Например, в Великобритании, в 2019 году, доля организаций, использующих BIM, составила 70%, что внушительно в сравнении с 10% в 2011 году [2]. Этот значительный рост связан с решением правительства Великобритании внедрить обязательное требование использовать «fully collaborative 3D BIM» в государственных проектах.

Подобное обязательство в отношении применения инструментов BIM было введено в Сингапуре с 2015 года, начиная с проектов, занимающих площадь от 5 тысяч квадратных метров. Согласно информации от Министерства Строительства России, в 2017 году только 5-7% российских компаний использовали BIM, и преимущественно это касалось крупных городов и мегапроектов. Однако, с течением времени, количество проектов, в которых применяют BIM, значительно увеличилось. Этот рост связан с развитием нормативной и законодательной базы, ростом спроса со стороны заказчиков, а также с уменьшением затрат на внедрение этой технологии [3].

Ранее использование BIM в проектировании обходилось в 2-3 раза дороже по сравнению с традиционными методами, но сегодня разница в стоимости сократилась до приблизительно 30% [4]. Это изменение цен обусловлено в значительной степени развитием стандартизированных элементов информационных моделей и плагинов.

В 2019 году доля организаций в России, работающих с BIM в инвестиционно-строительной сфере, составляла в среднем 22% [5]. Основной акцент на использование BIM сосредотачивается на стадии проектирования, где его применение составляет 80%, в то время как на стадии строительства этот показатель снижается до 15%, а на стадии эксплуатации всего лишь 5% [6].

Казахстан, учитывая опыт других стран, таких как Россия и Великобритания, недавно официально принял решение о внедрении поэтапной интеграции BIM в процесс проектирования и строительства. Согласно этому плану, правительство внимательно рассмотрит возможность обязательного использования BIM с 1 января 2021 года [6].

Одним из основных трендов в интеграции кибербезопасности в BIM является рост ценности данных, хранимых в моделях BIM. Эти данные содержат информацию о проекте, его конструкции, материалах, бюджете и сроках выполнения. Такие данные становятся желанными целями для кибератак, и их утечка может привести к серьезным последствиям.

Чтобы обеспечить безопасность данных BIM, требуется внимательное управление доступом и шифрование информации. Кроме того, необходимо разрабатывать стратегии реагирования на инциденты, чтобы быстро реагировать на возможные угрозы и минимизировать ущерб.

Эксперты прогнозируют восемь ключевых трендов в области кибербезопасности на ближайшие годы, которые помогут компаниям разработать эффективные стратегии защиты информационных инфраструктур.

Ориентированные на сотрудников подходы к киберзащите: К 2027 году 50% директоров по информационной безопасности будут внедрять стратегии, акцентирующие внимание на действиях сотрудников для минимизации операционных рисков и повышения эффективности контроля.

Регулирование конфиденциальности данных: В 2024 году современное регулирование конфиденциальности данных будет охватывать большую часть клиентских

данных. Однако менее 10% компаний смогут использовать это как конкурентное преимущество, рекомендуется соответствовать Общему регламенту по защите данных (GDPR).

Инфраструктура "нулевого доверия": К 2026 году 10% крупных предприятий будут иметь устойчивую инфраструктуру "нулевого доверия", что является значительным ростом по сравнению с менее чем 1% в начале 2023 года [7].

Внешние технологии и сотрудники: В 2027 году около 75% сотрудников компаний будут внедрять, модифицировать или создавать технологии без вмешательства ИТ-отдела. Это требует пересмотра операционных моделей кибербезопасности для более тесного взаимодействия с персоналом.

Количественная оценка киберрисков: К 2025 году 50% руководителей в области кибербезопасности будут безуспешно пытаться использовать количественную оценку киберрисков для корпоративных решений. Рекомендуется уделять внимание запросам на количественные оценки от лиц, принимающих решения.

Постоянная смена мест работы руководителей кибербезопасности: К 2025 году почти половина руководителей в области кибербезопасности сменит место работы, причем 25% из-за стресса, связанного с профессиональной деятельностью.

Участие в советах директоров: К 2026 году 70% компаний включат в состав совета директоров представителя с опытом в области кибербезопасности.

Использование данных управления рисками: К 2026 году более 60% средств обнаружения, расследования и реагирования на инциденты будут использовать данные управления рисками для приоритизации обнаруженных угроз.

Эти тренды показывают, что кибербезопасность будет оставаться важной и динамично развивающейся областью, требующей постоянной адаптации и инноваций со стороны компаний [8].

Еще одним трендом является рост интереса к защите от манипуляции моделями BIM. Киберпреступники могут внести изменения в модели, что может привести к ошибкам и недоразумениям на стройплощадке. Это может повлечь за собой не только задержки и перерасход бюджета, но и повысить риски безопасности.

Таким образом, для борьбы с этими угрозами разрабатываются механизмы проверки целостности моделей BIM. Эти механизмы могут обнаруживать несанкционированные изменения и предупреждать об их возможных последствиях. Это помогает поддерживать надежность данных в BIM и уменьшать риски.

Один из ключевых аспектов интеграции кибербезопасности в BIM - это обучение персонала. Сотрудники, работающие с данными BIM, должны быть обучены соблюдать меры безопасности, включая правила использования паролей, соблюдение политик доступа и уведомление об инцидентах.

Регулярное обучение и тестирование персонала на знание правил кибербезопасности становятся стандартной практикой в организациях, использующих BIM. Это помогает усилить защиту от внутренних источников угроз [9].

Компании могут настроить систему BIM так, чтобы ограничить доступ к данным только для авторизованных сотрудников и управлять уровнями доступа. Например, инженеры могут иметь доступ только к разделам модели, необходимым для их работы, и не иметь доступа к финансовым данным или конфиденциальной информации. Данные BIM могут быть зашифрованы, чтобы защитить их от несанкционированного доступа. Это может предотвратить утечку конфиденциальной информации, даже если кто-то получит доступ к файлам. Компании могут использовать системы мониторинга активности для отслеживания необычной активности в BIM. Это может включать в себя обнаружение попыток несанкционированного доступа или изменения данных.

Регулярные аудиты безопасности могут помочь выявить уязвимости и потенциальные проблемы в системе BIM. Это позволяет компаниям быстро реагировать на угрозы и вносить необходимые изменения.

Компании могут уделять особое внимание соблюдению нормативов и стандартов в области кибербезопасности, таких как GDPR или HIPAA, в зависимости от характера данных, хранящихся в BIM.

Применение этих методов и практик позволяет компаниям укрепить кибербезопасность в BIM, обеспечивая защиту данных и предотвращая возможные киберугрозы. Это особенно важно в сфере строительства и проектирования, где конфиденциальность и целостность данных играют критическую роль в успешном выполнении проектов.

Интеграция кибербезопасности в BIM будет продолжать развиваться в ближайшие годы. С развитием технологий и ростом угроз кибербезопасности будут разрабатываться новые методы и решения для обеспечения безопасности данных и моделей BIM. Возможно, появятся интегрированные решения, которые позволят управлять безопасностью BIM на всех этапах проекта, от проектирования до эксплуатации.

Будущее BIM будет включать в себя более интегрированные системы мониторинга и обнаружения угроз. Это позволит оперативно выявлять и реагировать на потенциальные атаки и изменения в BIM-моделях. Использование искусственного интеллекта и машинного обучения в кибербезопасности BIM может помочь в автоматизации процессов обнаружения и реагирования на угрозы. Алгоритмы могут выявлять аномальное поведение и предупреждать о возможных атаках.

С ростом объемов данных BIM, инфраструктура обработки данных будет включать более мощные серверы и облачные вычисления. Это позволит обеспечивать высокую доступность и защиту данных.

Гарантировать целостность данных в BIM-моделях будет одной из важных задач. В будущем предполагается развитие механизмов контроля целостности и возможность быстрого восстановления данных после атак или сбоев [10].

Интеграция кибербезопасности в BIM будет продолжать развиваться, чтобы соответствовать растущим угрозам и требованиям безопасности данных. Это позволит сохранять целостность и конфиденциальность информации в BIM-проектах, а также обеспечить успешное и безопасное выполнение строительных проектов.

Интеграция кибербезопасности в BIM становится все более актуальной в современной строительной индустрии. С ростом ценности данных BIM и увеличением угроз кибербезопасности необходимо принимать меры по защите информации и обучению персонала. Правильно реализованная стратегия кибербезопасности в BIM помогает минимизировать риски и обеспечивать успешное выполнение проектов.

Заключение

В заключение, интеграция кибербезопасности в сферу Building Information Modeling (BIM) представляет собой неотъемлемую часть современной строительной и проектировочной индустрии. Внедрение этой интеграции становится все более актуальным, учитывая рост киберугроз и увеличение объемов данных, хранящихся в BIM-моделях.

Мы рассмотрели несколько ключевых аспектов данной темы:

Сущность BIM: BIM представляет собой не просто трехмерную модель здания, но и интегрированную информационную систему, включающую данные о материалах, стоимости, графиках, характеристиках и других аспектах проекта.

Необходимость кибербезопасности в BIM: Внедрение кибербезопасности в BIM обусловлено не только ростом угроз, но и значительным увеличением количества информации, которую необходимо защищать.

Тренды в сфере кибербезопасности: Мы рассмотрели восемь ключевых трендов в области кибербезопасности, ориентированных на будущее, включая фокус на сотрудниках, регулирование конфиденциальности, инфраструктуру "нулевого доверия" и другие.

Перспективы интеграции кибербезопасности в BIM: В будущем, интеграция кибербезопасности в BIM будет направлена на обеспечение защиты данных, мониторинг угроз, использование искусственного интеллекта и машинного обучения, а также соблюдение стандартов и обучение персонала.

Все эти аспекты подчеркивают важность безопасности данных в BIM и необходимость постоянного развития методов и технологий для ее обеспечения. Интеграция кибербезопасности в BIM позволяет минимизировать риски и обеспечивает надежность и конфиденциальность данных, что становится фундаментом для успешного развития строительных и проектировочных проектов в будущем.

Список использованных источников

1. Цифровая трансформация отраслей: стартовые условия и приоритеты: докл. к XXII Апр. междунар. науч. конф. по проблемам развития экономики и общества, Москва, 13–30 апр. 2021 г. / Г. И. Абдрахманова, К. Б. Быховский, Н. Н. Веселитская, К. О. Вишневский, Л. М. Гохберг и др.
2. 10th annual BIM report 2020 - [Электронный ресурс]. URL: https://www.thenbs.com/PublicationIndex/documents/details?Pub=RIBAE&DocID=329337. (дата обращения: 27.09.2023).
3. Как интеграторы развивают BIM: состояние рынка, опыт отечественных компаний, комментарии - [Электронный ресурс]. URL: https://spbit.ru/analytics/kak-integratory-razvivayut-bim-sostoyanie-rynka-opyt-otechestvennyh-kompaniy-kommentarii-39705. (дата обращения: 28.09.2023).
4. Может ли российская ТИМ заменить иностранную BIM - [Электронный ресурс]. URL: https://www.cnews.ru/articles/2023-06-21_mozhet_li_rossijskaya_tim_zamenit?erid=%D0%BC%D0%BF10007%D0%90. (дата обращения: 28.09.2023).
5. Отчет по исследованию «Уровень применения BIM в России 2019» - [Электронный ресурс]. URL: http://concurator.ru/information/bim_report_2019/. (дата обращения: 28.09.2023).
6. BIM технологии становятся неотъемлемой частью процесса проектирования и строительства - [Электронный ресурс]. URL: https://www.pwc.com/kz/ru/blogs/lu-march-2017.html. (дата обращения: 28.09.2023).
7. Gartner Unveils Top Eight Cybersecurity Predictions for 2023-2024.
8. Составлен список 8 главных прогнозов по кибербезопасности на ближайшие годы - [Электронный ресурс]. URL: https://www.tadviser.ru/index.php/. (дата обращения: 28.09.2023).
9. Гулик, В. Ю. Перспективы внедрения BIM-технологий / В. Ю. Гулик. – Текст : непосредственный // Архитектура, строительство, транспорт. – 2021. – № 2. – С. 58–63.
10. Кибербезопасность в 2022–2023. Тренды и прогнозы [Электронный ресурс]. URL: https://www.ptsecurity.com/ru-ru/research/analytics/ogo-kakaya-ib/. (дата обращения: 28.09.2023).

СОВРЕМЕННЫЕ МЕТОДЫ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В BIM: АНАЛИЗ И СРАВНЕНИЕ

Усатова Ольга Александровна[1,2], Пейішбек Ағахан Нұрлыбекұлы[2]
email: 222134@astanait.edu.kz
[1]*Институт информационных и вычислительных технологий КН МНВО РК, Алматы, Республика Казахстан*
[2]*Университет «Astana IT University», Астана, Республика Казахстан*

Аннотация. *Статья представляет исследование актуальных методов и стратегий обеспечения информационной безопасности в контексте информационного моделирования зданий (BIM). Автор проводит сравнительный анализ различных подходов к обеспечению безопасности данных в BIM-процессах, выявляя их преимущества и ограничения. Статья также обсуждает вызовы и перспективы применения современных технологий, таких как блокчейн, в области BIM. Исследование призвано предоставить читателям более глубокое понимание важности и эффективности обеспечения информационной безопасности в BIM и помочь им выбрать наилучшие методы для своих проектов.*

Введение

Информационное моделирование зданий (Building Information Modeling, BIM) становится неотъемлемой частью строительной и проектировочной отрасли. Однако с ростом популярности BIM возникают новые вызовы в области информационной безопасности. В этой статье мы проанализируем современные методы обеспечения информационной безопасности в BIM и сравним их, чтобы помочь компаниям и организациям принимать информированные решения о защите своих данных в этой среде.

Building Information Modeling (BIM) - это инновационная методология, которая революционизировала строительную и архитектурную отрасли, предоставляя комплексный подход к проектированию, строительству и управлению зданиями и инфраструктурой.

BIM позволяет создавать трехмерные цифровые модели зданий и инфраструктуры, интегрировать данные о конструкции, материалах, графиках и расписаниях, что упрощает совместное взаимодействие разных участников проекта [1].

BIM позволяет улучшить эффективность проектирования и строительства, повысить качество проектов, сократить бюджеты и сроки строительства. BIM используется в различных сферах, включая архитектуру, инженерное проектирование и строительство [2].

BIM - это эффективный инструмент, который преобразовывает способ проектирования, строительства и управления зданиями. Ссылки на официальные ресурсы и образовательные материалы помогут вам глубже понять и освоить эту важную технологию.

В свою очередь, информационная безопасность охватывает следующие аспекты:

1. **Защита информационного пространства**: Гарантирует создание и развитие информационной среды, которая служит интересам граждан, организаций и государства.
2. **Безопасное использование информации:** Обеспечивает, что информация используется только согласно своему назначению и не наносит вред системе при ее использовании.
3. **Защита свойств информации:** Гарантирует сохранность свойств информации, таких как конфиденциальность (секретность), целостность (непротиворечивость и защита от несанкционированных изменений) и доступность.
4. **Экономические аспекты:** Охватывают управление информацией в экономической сфере, включая системы сбора, хранения и обработки информации для управления производственными структурами, анализа экономического развития, управления в промышленности, транспорте, энергосистемах и других областях.
5. **Финансовые аспекты:** Включают информационные сети и базы данных банков и финансовых организаций, системы финансового обмена и расчетов [3].

Обеспечение информационной безопасности начинается с определения заинтересованных сторон, связанных с использованием информационных систем. Их интересы включают доступность (возможность получения информации в приемлемые сроки), целостность (сохранность и актуальность информации, защита от вмешательства и изменений), конфиденциальность (секретность информации).

Рис. 1 Структура понятия «Информационная безопасность»

На самом высоком уровне абстракции, информационная безопасность может быть определена как способность сохранить неприкосновенность свойств объекта безопасности, которые связаны с информацией и информационной инфраструктурой (см. Рисунок 1) [3].

На сегодняшний день есть разные методы обеспечения информационной безопасности в BIM, которые показаны в таблице 1 [5].

Таблица 1.

Преимущества и недостатки методы обеспечения информационной безопасности в BIM

№	Методы обеспечения информационной безопасности в BIM		Преимущества	Недостатки
1	Шифрование данных		Шифрование обеспечивает конфиденциальность данных, делая их непригодными для несанкционированного доступа	Эффективное шифрование требует значительных вычислительных ресурсов, что может замедлить рабочие процессы BIM
2	Аутентификация и авторизация		Аутентификация и авторизация гарантируют, что только уполномоченные пользователи имеют доступ к данным BIM	Управление пользователями и правами доступа может потребовать значительных усилий и ресурсов.
3	Мониторинг и обнаружение угроз:		Системы мониторинга и обнаружения угроз способствуют раннему обнаружению аномального поведения и потенциальных нарушителей	Для эффективной работы эти системы требуют постоянного мониторинга и настройки
4	Физическая безопасность		Защита физического доступа к инфраструктуре BIM, такой как серверы и хранилища данных	Дорогие меры, которые могут быть недоступны для малых компаний
5	Обучение пользователей		Обучение пользователей о правилах безопасности и социальной инженерии может снизить риск человеческих ошибок	Не всегда эффективно, и требует времени и ресурсов
6	Обновления и патчи		Регулярные обновления и установка патчей помогают устранять известные уязвимости	Требует времени и усилий для обслуживания системы BIM.

Вместе с тем, на сегодняшний день поднимается вопрос о применения технологии блокчейн в BIM для усовершенствования процессов передачи, хранения и обеспечения

безопасности данных. Интеграция BIM и блокчейн является относительно новой областью, и на данный момент доступные материалы по этому вопросу ограничены. В глобальной практике реального опыта интеграции BIM и блокчейн пока недостаточно [6].

На данный момент большинство людей ассоциируют блокчейн в основном с криптовалютами, такими как биткоин. Однако блокчейн гораздо более широкая технология, чем просто функционирование новых цифровых валют. Его суть заключается в создании непрерывной и устойчивой последовательной цепочки блоков, содержащих зашифрованные данные. Блокчейн - это программный продукт и протокол обмена данными, представляющий собой распределенную базу данных, в которой устройства хранения данных не зависят от центрального сервера. Каждый блок в цепочке содержит метку времени и ссылку на предыдущий блок [7].

Но в то же время, один из важных вопросов заключается в создании ясной и конкретной спецификации BIM для интеграции с технологией блокчейн. Это поможет определить, какая информация необходима для того, чтобы BIM-моделирование соответствовало требованиям строительной отрасли и инструментам блокчейн [8].

Ограниченная совместимость между технологиями BIM и блокчейн представляет собой значительную проблему. Это означает, что стандарты данных и оперативная совместимость остаются критически важными факторами для успешной интеграции блокчейн в BIM-технологии.

Блокчейн - это мощный инструмент для обеспечения надежности и прозрачности данных. Однако, если его использовать как обычную базу данных, производительность и скорость могут стать ограничивающими факторами. Следует учитывать, что технически блокчейн можно настроить на запись данных без их перезаписи, сохраняя при этом историю.

Эти аспекты подчеркивают важность детального исследования и разработки практических решений для интеграции блокчейн и BIM-технологий.

Между тем, шифрование и аутентификация/авторизация обеспечивают высокий уровень безопасности, но могут быть затратными и требовать значительных ресурсов. Мониторинг и обнаружение угроз предоставляют баланс между безопасностью и ресурсами. Физическая безопасность и обучение пользователей важны, но не являются единственными методами защиты [9].

Также, обновления и патчи легко внедряются, но не могут обеспечить полную безопасность. Шифрование и аутентификация/авторизация требуют более сложных систем. Мониторинг и обнаружение угроз требуют навыков и опыта в области информационной безопасности.

Обновления и патчи обычно более доступны с точки зрения финансов. Шифрование и аутентификация/авторизация могут потребовать значительных инвестиций. Мониторинг и обнаружение угроз также могут быть дорогими, но могут сэкономить средства в будущем, предотвращая инциденты.

Обучение пользователей играет важную роль в предотвращении человеческих ошибок, но не исключает их полностью.

Заключение

В данной статье была рассмотрена актуальная и инновационная теа интеграции технологии блокчейн и информационного моделирования зданий (BIM) в строительстве и архитектуре. Переход к BIM-процессам, базирующимся на блокчейн, представляет собой перспективное направление, которое может улучшить передачу, хранение и

защиту данных в данной отрасли. Обеспечение информационной безопасности в среде BIM является сложной задачей, требующая сбалансированного подхода. Каждый метод имеет свои преимущества и ограничения, и оптимальное решение зависит от конкретных потребностей и ресурсов организации. С учетом постоянно меняющейся угрозной среды, важно постоянно обновлять и улучшать методы обеспечения информационной безопасности в BIM, чтобы защитить ценные данные и проекты от потенциальных угроз. Однако, при рассмотрении этой темы, были выявлены несколько важных аспектов и вызовов. Сначала необходимо разработать четкую спецификацию BIM для блокчейн, чтобы определить, какая информация является необходимой для соответствия требованиям строительной отрасли и инструментам блокчейн. Кроме того, совместимость между BIM и блокчейн-технологиями остается важным вопросом, который требует дальнейшего исследования.

Технология блокчейн обладает своими преимуществами и ограничениями. Она эффективна в обеспечении надежности и безопасности данных, но требует внимательного рассмотрения вопроса о ее использовании в контексте BIM. Понимание того, что блокчейн наилучшим образом работает в децентрализованных сетях, важно для принятия решения о его использовании [10].

Несмотря на вызовы, связанные с техническими и организационными аспектами, интеграция блокчейн и BIM представляет потенциал для улучшения эффективности и надежности процессов в строительстве и архитектуре. Она также подчеркивает важность развития кадрового потенциала и подготовки специалистов, способных эффективно работать с этими технологиями.

В завершение, успешная интеграция блокчейн и BIM может стать ключевым фактором для улучшения качества и безопасности строительных проектов, а также оптимизации процессов в данной отрасли.

Список использованных источников

1. BuildingSMART. https://www.buildingsmart.org/ - [Электронный ресурс]. (дата обращения: 27.09.2023).
2. Информационное моделирование зданий https://www.autodesk.com/solutions/bim/benefits - [Электронный ресурс]. (дата обращения: 27.09.2023).
3. Уфимцев, Ю.С. ,сост. Информационная безопасность России Моск. акад. экономики и права; [Ю.С. Уфимцев и др.]. – М.: Экзамен, 2003
4. Ходжаев, А.Г. Оценка стойкости систем защиты информации и скоростные алгоритмы преобразования данных в системах: Автореф. дис. ... канд. техн. наук: 05.13.19 Моск. гос. акад. приборостроения и информатики. – М., 2001
5. Добрынин А. П. и др. Цифровая экономика – различные пути к эффективному применению технологий (BIM, PLM, CAD, IOT, Smart City, BIG DATA и т.д.
6. Букунова О.В., Букунов А.С. Трудности внедрения информационного моделироавния объектов строительства в России // Научный форум: Инновационная наука: сб. ст. по материалам VI междунар. науч.-практ. конф. № 5(6). М.: Изд. «МЦНО», 2017. С. 15-21.
7. Кириллов П. Блокчейн – нервная система умных производств ближайшего будущего // Умное производство. 2017. № 40. выводы [Электронный ресурс] – Режим доступа. – URL: http://www.umpro.ru/index.php?page_id=17&art_id_1=849&group_id_4=24 (дата обращения: 29.09.2023).
8. Алексеенко Н. Блокчейн без посредников: каким будет девелопмент в цифровой экономике // РБК. 26.07.2017. [Электронный ресурс] – Режим доступа. –

URL:https://realty.rbc.ru/news/59788fab9a7947d94ee1ddcb (дата обращения: 29.09.2023).

9. Талапов В.В. Технология BIM: суть и особенности внедрения информационного моделирования зданий. М.: ДМК Пресс, 2015. 410 с.
10. Green BIM. How Building Information Modeling is Contributing to Green Design and Construction. – McGraw-Hill Construction, 2010.

ИССЛЕДОВАНИЕ АЛГОРИТМОВ КОНСЕНСУСА ДЛЯ ОБЕСПЕЧЕНИЯ СТАБИЛЬНОСТИ И НАДЕЖНОСТИ БЛОКЧЕЙН-СИСТЕМЫ

Усатова О.А., Жетыбай Н.Қ.
email: uoa-olga@mail.ru, nazerkezhetybay@gmail.com
Институт информационных и вычислительных технологий КН МНВО РК, Алматы, Республика Казахстан
КазНУ имени аль-Фараби, Алматы, Республика Казахстан

Аннотация. *Данная статья рассматривает сравнительный анализ трех алгоритмов консенсуса в блокчейн технологиях: PoW (Proof of Work), PoS (Proof of Stake) и PBFT (Practical Byzantine Fault Tolerance). Для каждого из этих алгоритмов были рассмотрены их основные преимущества и недостатки, а также их применение в различных сценариях. Было отмечено, что выбор алгоритма консенсуса зависит от конкретного применения и сценария использования. Например, для блокчейнов, где безопасность является приоритетом, может быть более подходящим алгоритм PoW и PBFT, а для блокчейнов, где важны как безопасность, так и энергоэффективность, может быть более подходящим алгоритм PoS. Также были рассмотрены возможные проблемы и уязвимости каждого из алгоритмов и способы их решения.*

Введение

Развитие технологий и появление технологии блокчейн сильно изменили нашу жизнь и повлияли на многие области экономики и бизнеса.

Технология блокчейн, известная также как цепочка блоков, это децентрализованная база данных, которая хранит информацию в виде блоков, связанных между собой по цепочке. Каждый блок содержит информацию о предыдущем блоке, что обеспечивает целостность и надежность всей цепочки. Блокчейн был разработан для поддержки криптовалюты биткоин, но теперь находит применение в других областях, таких как финансы, логистика, здравоохранение, право и многие другие.

Развитие технологий и появление блокчейн-технологии позволили устранить проблему двойных расходов, сделали более прозрачными и безопасными финансовые транзакции, упростили процессы учета и проверки данных в логистике и снабжении, а также увеличили уровень безопасности данных во многих отраслях, включая здравоохранение и право [1].

Блокчейн также открывает новые возможности для создания децентрализованных приложений и платформ, которые не нуждаются в централизованном управлении и могут быть доступны для использования всем пользователям без посредников.

Таким образом, развитие технологий и появление блокчейн-технологии являются важными шагами в направлении более эффективной и безопасной экономики и бизнеса, которая может быть доступна для всех.

Блокчейн-технология стала основой для многих инновационных приложений, таких как криптовалюты, смарт-контракты и децентрализованные приложения. Технология блокчейн имеет некоторые недостатки, такие как низкая скорость транзакций,

высокая стоимость и недостаточная масштабируемость. Для решения этих проблем необходимо усовершенствовать алгоритмы консенсуса, используемые в блокчейн-системах. Алгоритм консенсуса - это механизм, используемый для достижения согласия между участниками сети относительно текущего состояния блокчейн-системы. В блокчейне, основанном на принципе консенсуса, каждый узел сети имеет копию базы данных и должен согласиться с другими узлами по поводу того, какой блок добавить в цепочку [2].

Существует несколько алгоритмов консенсуса, каждый из которых имеет свои достоинства и недостатки. Одним из самых популярных алгоритмов является Proof-of-Work (PoW), используемый в биткоине, требует от узлов решения сложной математической задачи, чтобы добавить новый блок в цепочку. Он основывается на решении сложных математических задач, чтобы создать новые блоки и проверить транзакции. Этот алгоритм обеспечивает высокий уровень безопасности, но требует большого количества вычислительной мощности и энергозатрат [3].

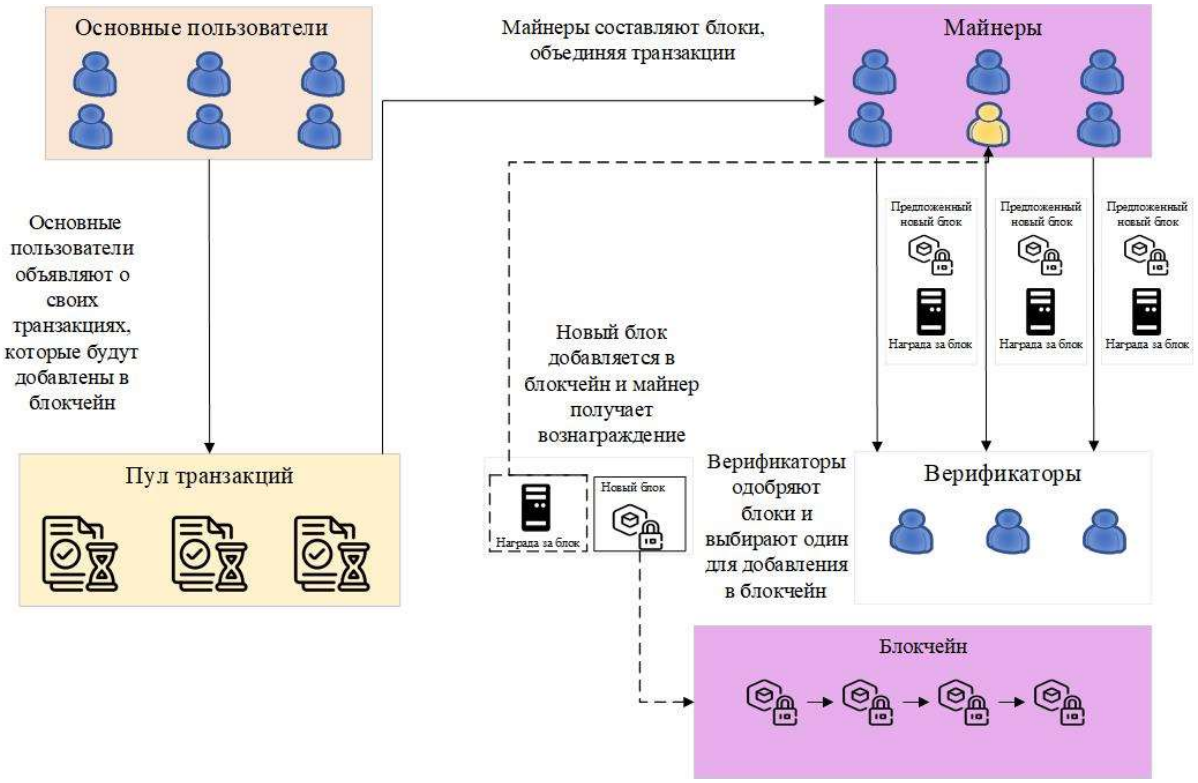


Рис. 1 Классический консенсусный протокол PoW.

Описанный PoW представляет собой очень параллельный процесс. Несколько майнеров одновременно пытаются добавить новый блок (не обязательно один и тот же). Все они итеративно вычисляют значение хеш-функции для разных значений до тех пор, пока не будет удовлетворен один из двух критериев: получено желаемое значение хеш-функции или прибудет потенциально новый блок для проверки. В первом случае соответствующий майнер отправляет свой блок и соответствующее значение nonce другим для проверки, тогда как второй случай означает, что какой-то другой майнер нашел правильное значение и его блок, вероятно, будет принят для включения в Блокчейн. Следовательно, только один майнер работает продуктивно, тогда как все

остальные просто тратят свои ресурсы, время и энергию. Кроме того, транзакции, входящие в их блоки (а не в принятые), должны быть возвращены в пул транзакций и рассмотрены повторно.

Proof-of-Stake (PoS) является алгоритмом консенсуса, который работает на основе доказательства владения активами в сети. Участники сети, у которых есть больше активов, имеют больший шанс стать выбранным для создания нового блока. PoS менее ресурсоемкий, чем PoW и менее безопасный [4]. Это алгоритм используется в блокчейне Ethereum и основывается на выборе стейкеров, которые имеют право создавать новые блоки на основе своей доли в сети. PoW требует меньше вычислительной мощности и энергозатрат, но может быть менее безопасным, если большинство стейкеров решит провести атаку на сеть [5]. Процессы консенсуса PoS представлены на рисунке 2.

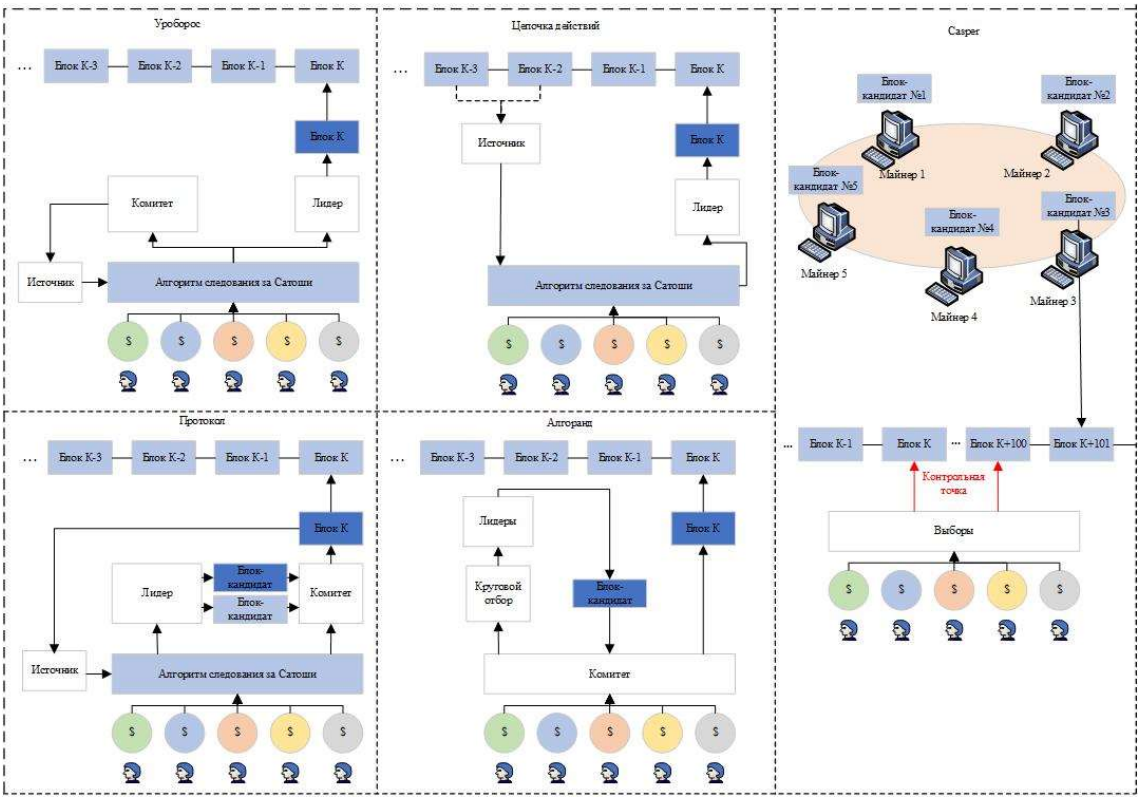


Рис. 2 Процессы консенсуса PoS.

Здесь представлены новые протоколы на основе PoS, которые широко применяются на практике, а именно: Уроборос, Цепочка действий, Casper, Алгоранд и Протокол.

Другое направление исследований - комбинирование различных алгоритмов консенсуса для достижения оптимального баланса между безопасностью, скоростью и масштабируемостью. Например, компания Ripple использует алгоритм консенсуса, основанный на комбинации PoW (Proof of Work) и PBFT (Practical Byzantine Fault Tolerance), который обеспечивает быстрое подтверждение транзакций и высокий уровень безопасности [6].

PBFT используется в распределенных системах, таких как Hyperledger Fabric. Он основывается на голосовании среди узлов сети и позволяет достигать консенсуса даже при наличии нескольких злоумышленников. PBFT обеспечивает высокий уровень

безопасности и быстродействия, но требует большого количества коммуникации между узлами сети [7]. Алгоритм PBFT представляет собой двухслойную систему, которая состоит из клиентов и реплик (представлен на рисунке 3).

Первый слой - слой клиентов, который представляет собой группу узлов, отправляющих запросы на обработку. Каждый клиент отправляет свой запрос на обработку репликам, которые находятся на втором слое.

Второй слой - слой реплик, который состоит из нескольких узлов, которые обрабатывают запросы, отправленные клиентами. Реплики могут быть размещены на разных узлах и даже в разных центрах обработки данных для обеспечения высокой доступности.

Каждый узел реплики имеет свой собственный реестр, в котором хранятся все транзакции, обработанные системой. Реплики синхронизируют свои реестры с помощью протокола PBFT, чтобы обеспечить согласованность данных. Топология PBFT обеспечивает высокую отказоустойчивость и безопасность, так как каждый запрос на обработку проходит через несколько узлов реплик, прежде чем получает подтверждение. Кроме того, система PBFT является устойчивой к атакам типа «византийских», которые могут произойти в распределенных системах [8].

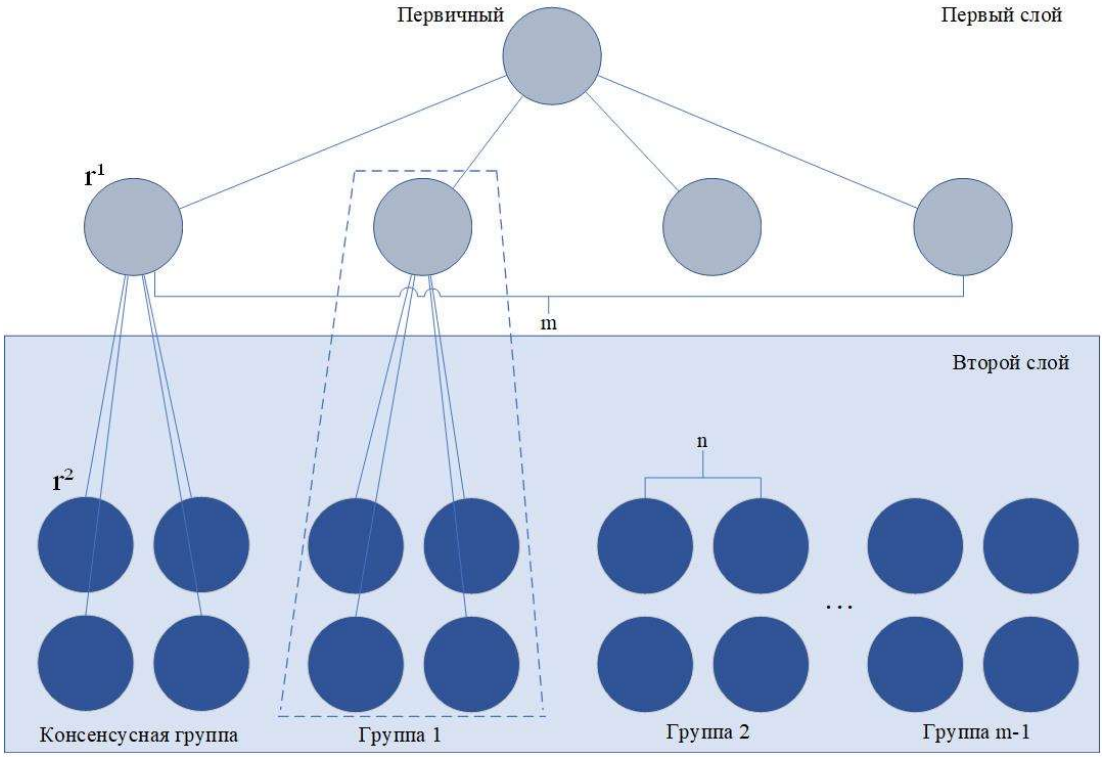


Рис. 3 Топология двухслойной системы PBFT

Исследование алгоритмов консенсуса является важной областью для улучшения стабильности и надежности блокчейн-систем. Новые алгоритмы консенсуса могут помочь решить проблемы масштабируемости и скорости блокчейн-систем, а также снизить затраты на энергию и вычислительные ресурсы, необходимые для обеспечения безопасности сети [9].

Каждый из этих алгоритмов имеет свои преимущества и недостатки, и выбор конкретного алгоритма зависит от требований конкретного проекта. Алгоритмы консенсуса являются ключевыми элементами в распределенных системах и блокчейн технологиях, обеспечивая надежность, безопасность и эффективность работы между узлами системы. Изучение различных алгоритмов консенсуса и их применения может помочь в разработке и улучшении существующих систем и технологий.

Консенсусные алгоритмы для журналирования используются для достижения согласия между несколькими узлами в распределенной системе относительно порядка транзакций в журнале. Ниже приведена таблица, которая сравнивает некоторые из наиболее известных алгоритмов консенсуса для журналирования [10].

Журналирование в блокчейне – это процесс фиксации и хранения записей о транзакциях или других событиях в распределенной базе данных блокчейна.

Каждый блок в блокчейне содержит набор транзакций, которые были проверены и добавлены к блоку. Каждый блок также содержит уникальный идентификатор, называемый хешем, который генерируется на основе содержимого блока. Хеш представляет собой уникальную цифровую подпись блока и используется для связывания блоков в цепочку.

Каждый участник сети блокчейна может увидеть каждый блок, и эти блоки хранятся в распределенной базе данных, что обеспечивает прозрачность и надежность процесса журналирования [11].

Рассмотрим свойства алгоритмов консенсуса, а также их преимущества и недостатки (таблица 1-3).

Таблица 1

Сравнительный анализ свойств алгоритмов консенсуса

Алгоритм	**Тип**	**Описание**	**Преимущества**	**Недостатки**
Бизантинский алгоритм общественного доверия (BFT)	Распределенный	Используется для достижения консенсуса в сети с возможностью присутствия злоумышленников, включая узлы, которые могут сообщать неправильные данные.	Высокая степень устойчивости к атакам	Требуется большая вычислительная мощность, не очень масштабируемый
Proof of Work (PoW)	Нераспределенный	Используется в блокчейне Bitcoin. Майнеры решают сложные математические задачи для добавления блоков в блокчейн.	Простота в реализации, устойчивость к атакам	Очень высокая потребность в энергии, медленная скорость обработки транзакций
Proof of Stake (PoS)	Нераспределенный	Используется в блокчейне Ethereum. Владельцы монет м огут участвовать в создании новых блоков, при этом количество новых блоков, которые могут быть созданы, зависит от количества монет, которыми владеет участник.	Меньшая потребность в энергии, более высокая скорость обработки транзакций, меньшие затраты на оборудование |	Могут возникать проблемы с достижением консенсуса в случае атак на сеть, возможны проблемы с централизацией в случае сильных игроков в сети

Таблица 2

Сравнение структурных свойств

Характеристики	**PoW**	**PoS**	**PBFT**
Требования к участникам	Наличие большого количества вычислительной мощности	Наличие большого количества монет, владение которыми подтверждается	Наличие определенного числа узлов-защитников
Сложность вычислений	Высокая	Низкая	Средняя
Энергопотребление	Высокое	Низкое	Среднее
Скорость транзакций	Низкая	Высокая	Высокая
Степень децентрализации	Средняя	Высокая	Высокая
Скорость принятия решений	Низкая	Высокая	Высокая
Риск атаки на консенсус	Низкий	Низкий	Низкий
Возможность форкинга	Высокая	Низкая	Низкая

Таблица 3

Сравнение свойств безопасности

Алгоритм	**PoW (Proof-of-Work)**	**PoS (Proof-of-Stake)**	**PBFT (Practical Byzantine Fault Tolerance)**
Распределенная атака	Требует большого количества вычислительной мощности для атаки на систему	Требует большого количества монет для атаки на систему	Защищен от распределенных атак, но может быть уязвим к вредоносным узлам
Возможность цензуры	Низкая	Низкая	Низкая
Надежность	Высокая	Высокая	Очень высокая
Необходимость доверия	Низкая	Низкая	Очень высокая
Устойчивость к 51% атакам	Устойчив к 51% атакам	Может быть уязвим к 51% атакам	Не применимо, так как голосование происходит между узлами

Заключение

Консенсусные алгоритмы для журналирования используются для достижения согласия между несколькими узлами в распределенной системе относительно порядка транзакций в журнале.

В целом, алгоритмы консенсуса являются ключевым элементом в распределенных системах и блокчейн технологиях, обеспечивая надежность, безопасность и эффектив-

ность работы между узлами системы. Изучение различных алгоритмов консенсуса и их применения может помочь в разработке и улучшении существующих систем и технологий.

Таким образом, выбор конкретного алгоритма консенсуса зависит от требований конкретного проекта и его целей. Выбор конкретного алгоритма консенсуса для журналирования зависит от целого ряда факторов, включая требования по безопасности, скорости обработки транзакций, энергоэффективности и других. Поэтому перед выбором конкретного алгоритма необходимо провести тщательный анализ требований проекта и решить, какой алгоритм лучше всего подходит для конкретной задачи.

Рассмотрев алгоритмы PoW, PoS и PBFT, можно выделить следующее: PoW обеспечивает высокий уровень безопасности, но требует большого количества вычислительной мощности и энергозатрат. Этот алгоритм подходит для блокчейнов, где безопасность является основным приоритетом, а затраты на энергию не являются критическими; PoS требует меньше вычислительной мощности и энергозатрат, но может быть менее безопасным, если большинство стейкеров решит провести атаку на сеть. Этот алгоритм подходит для блокчейнов, где важны как безопасность, так и энергоэффективность; PBFT обеспечивает высокий уровень безопасности и быстродействия, но требует большого количества коммуникации между узлами сети. Этот алгоритм подходит для распределенных систем, где быстродействие и безопасность одинаково важны.

Данное исследование подготовленно в рамках проекта AP19675957 «Разработка и исследование системы для обеспечения защиты медицинских данных с применением технологии блокчейн и методов искусственного интеллекта», который реализуется в Институте информационных и вычислительных технологий.

Список литературы:

1. Hamza B., Attila K.,Approaches to Overpower Proof-of-Work Blockchains Despite Minority, – 2023. DOI: 10.1109/ACCESS.2023.3234322.
2. Barakova Aliya, Ussatova Olga, Begimbayeva Yenlik, Ibrahim Sogukpinar, «Ensuring Information Security of Web Resources Based on Blockchain Technologies» // (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 14, No. 6, 2023, PP. 834-843 (DOI): 10.14569/IJACSA.2023.0140689.
3. Milan T., Luka M., Dušan R., Tatjana D., Dragan U., Tatjana J.K., Đorđe J., Proof-of-Useful-Work: BlockChain Mining by Solving Real-Life Optimization Problems. – 2022. DOI: https://doi.org/10.3390/sym14091831
4. Cong T. N., Dinh T., Diep N., Dusit N., Huynh T., Eryk D., Proof-of-Stake Consensus Mechanisms for Future Blockchain Networks: Fundamentals, Applications and Opportunities, – 2019. DOI: 10.1109/ACCESS.2019.2925010.
5. Shi Yan, Analysis on Blockchain Consensus Mechanism Based on Proof of Work and Proof of Stake, – 2022. DOI: 10.1109/ICDACAI57211.2022.00098.
6. Bo Gan, QiWu Wu, Xiang Li, Yang Zhou, Classification of Blockchain Consensus Mechanisms Based on PBFT Algorithm, – 2021. DOI: 10.1109/ICCEA53728.2021.00012.
7. Wenyu Li, Chenglin F., Lei Z., Hao Xu, Bin Cao, Muhammad Ali I., A Scalable Multi-Layer PBFT Consensus for Blockchain, – 2020. DOI: 10.1109/TPDS.2020.3042392.
8. J. Preece, J. Easton, To Blockchain or Not to Blockchain, These Are the Questions: A Structured Analysis of Blockchain Decision Schemes, – 2023. DOI: 10.13140/RG.2.2.12104.29449.

9. M.Cimiterra, J.Krafft, Blockchain: Antecedents and Future Challenges, – 2023. DOI: 10.1007/978-3-031-38260-4_9.
10. H.Jangid, P.Meel, Blockchain Protocols: Transforming the Web We Know, – 2023. DOI: 10.1007/978-981-99-3758-5_49.
11. Araz Zirar, Blockchain and human resource management: Emerging themes and research agenda, – 2023. DOI:10.21203/rs.3.rs-3368951/v1.

РЕАЛИЗАЦИЯ КВАНТОВО-УСТОЙЧИВОГО КРИПТОГРАФИЧЕСКОГО АЛГОРИТМА HORS С ХЕШ-ФУНКЦИЕЙ HAS01-256

[1, 2]Лизунов О.А., [1]Алгазы К.Т.
e-mail: o.lizunov@bk.ru, kunbolat@mail.ru
[1]Институт информационных и вычислительных технологий КН МНВО РК, Казахстан
[2]Казахский национальный исследовательский технический университет имени К. И. Сатпаева

Аннотация. *Появление квантовых компьютеров вызвало большой ажиотаж у специалистов в области информационной безопасности. Ученые криптографы доказали, что с помощью квантовых компьютеров можно будет дешифровать зашифрованную информацию за приемлемое время. В связи с этим, возникла необходимость в разработке квантово-устойчивых криптографических алгоритмов и проведении международных конкурсов по выбору наилучших реализаций для стандартизации.*

В данной статье рассматривается постквантовый криптографический алгоритм HORS, основанный на хеш-функциях. Сотрудниками Лаборатории информационной безопасности Института информационных и вычислительных технологий была разработана программная реализация алгоритма HORS на языке программирования C++ в интегрированной среде разработки Qt. В программной реализации алгоритма HORS используется разработанный в Лаборатории информационной безопасности алгоритм хеширования HAS01-256.

Введение

Развитие современных информационно-коммуникационных технологий позволило государственным и частным организациям воплотить в жизнь некоторые теоретические наработки в области квантовых технологий, которые позволили упростить и ускорить выполнение определенных задач. В 2023 году компанией IBM был представлен самый мощный квантовый компьютер Quantum Condor с 433 кубитами, который способен решать определенные задачи за несколько часов, в то время как классическим суперкомпьютерам потребовалось бы нескольких десятков тысяч лет [1]. По оценкам специалистов, широкое практическое использование квантовых вычислений может начаться в 2024-2025 годах с появлением 1000-кубитных систем [2]. По мнению ученых, квантовые компьютеры представляют серьезную угрозу наиболее популярным системам шифрования. В связи с этим, у специалистов в области информационной безопасности возникла обеспокоенность по поводу применения данных технологий с целью проведения криптоанализа. Так, американский учёный Питер Шор в 1994 году разработал квантовый алгоритм для решения задач факторизации целых чисел и дискретного логарифмирования, который позволит взламывать большинство используемых сейчас асимметричных криптографических алгоритмов. В связи с этим, в 2017 году Национальным институтом стандартов и технологий США (NIST) был организован конкурс по выбору квантово-устойчивых криптоалгоритмов.

Основными направлениями синтеза криптографических схем, применяемых в алгоритмах-участниках конкурса NIST являются:

1. Сложные задачи теории целочисленных решеток.
2. Использование кодов, исправляющих ошибки.
3. Схемы цифровой подписи на основе криптографических хеш-функций.
4. Системы многочленов от многих переменных.
5. Сложные задачи, связанные с изогениями суперсингулярных эллиптических кривых.
6. «Экзотические» задачи, такие как проблемы сопряженного поиска, операции в группах кос, алгебра октонионов, многочлены Чебышёва [3].

Существующие схемы цифровой подписи на основе криптографических хеш-функций можно классифицировать следующим образом:

- одноразовые (Лампорта, Винтерница);
- многоразовые с использованием деревьев Меркли;
- многоразовые со снижением стойкости (HORS, FORS, PORS);
- многоразовые с использованием гипердеревьев (SPHINCS, SPHINCS+, Gravity-SPHINCS, XMMS-MT).

С 2017 по 2023 год NIST было проведено 4 раунда по отбору квантово-устойчивых криптографических алгоритмов. В августе 2023 года NIST опубликовал три проекта стандартов на следующие алгоритмы:

- SPHINCS+ (схемы цифровой подписи).
- Dilithium (схемы цифровой подписи).
- Kyber (схема защищенной передачи ключа).

Сотрудниками ЛИБ были рассмотрены существующие схемы цифровых подписей на основе криптографических хеш-функций и в качестве модели построения собственного алгоритма был выбран алгоритм HORS, который является базовой схемой подписи алгоритма SPHINCS.

Алгоритм HORS

HORS (Hash to Obtain Random Subset) – это квантово-устойчивая многоразовая цифровая подпись на основе хеширования.

Алгоритм HORS можно использовать для цифровой подписи сообщений ограниченное количество раз, так как при каждом использовании подписи происходит раскрытие секретных ключей, что снижает безопасность ключа.

Основными составляющими алгоритма HORS являются:

1. Создание секретных ключей с помощью генератора псевдослучайных чисел.
2. Получение открытых ключей с помощью хеширования закрытых ключей.
3. Хеширование сообщения.
4. Получение цифровой подписи на основе хеш-образа сообщения и закрытых ключей.
5. Проверка подписи на основе хеш-образа сообщения, цифровой подписи и открытых ключей [4, 5].

С целью проведения анализа работы алгоритма HORS и созданием на его основе собственного алгоритма, сотрудниками ЛИБ была разработана программная реализация алгоритма HORS с использованием функции хеширования собственной разработки «HAS01-256».

Примечание: алгоритм хеширования «HAS01» разработан в ЛИБ. Он позволяет преобразовать открытый текст произвольной длины в хеш-образ размером 256 («HAS01-256») или 512 («HAS01-512») битов.

Основные параметры алгоритма HORS:

1. m – исходный файл.
2. N – количество частей, равных одному байту в хеш-образе.
3. L – количество бит в одной части хеш-образа.
4. length – длина хеш-образа равна N*L.
5. t – количество закрытых SK = (sk_0, sk_1, …, sk_{t-1}) и открытых ключей PK = (pk_0, pk_1, …, pk_{t-1}) равно 2^L.
6. σ – количество элементов цифровой подписи $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N) = (sk_{m1}, sk_{m2}, \ldots, sk_{mN})$ равно N [6].

В программной реализации алгоритма HORS используются следующие значения:

1. N – 32.
2. L – 8 бит.
3. length – 256 бит.
4. t – 256.
5. σ – 32.

Для демонстрации работы программы был выбран файл «experiment.txt». В результате работы программы были получены хеш-образ файла и цифровая подпись (Рисунок 1), а также три файла, созданные рядом с исходным фалом: файл с закрытыми ключами «experiment.txt.priv» (Рисунок 2), файл с открытыми ключами «experiment.txt.pub» (Рисунок 3) и файл с цифровой подписью «experiment.txt.sig» (Рисунок 4).

Примечание: в связи с тем, что файл с открытыми ключами содержит 256 ключей, на рисунке показаны только первые 33 ключа.

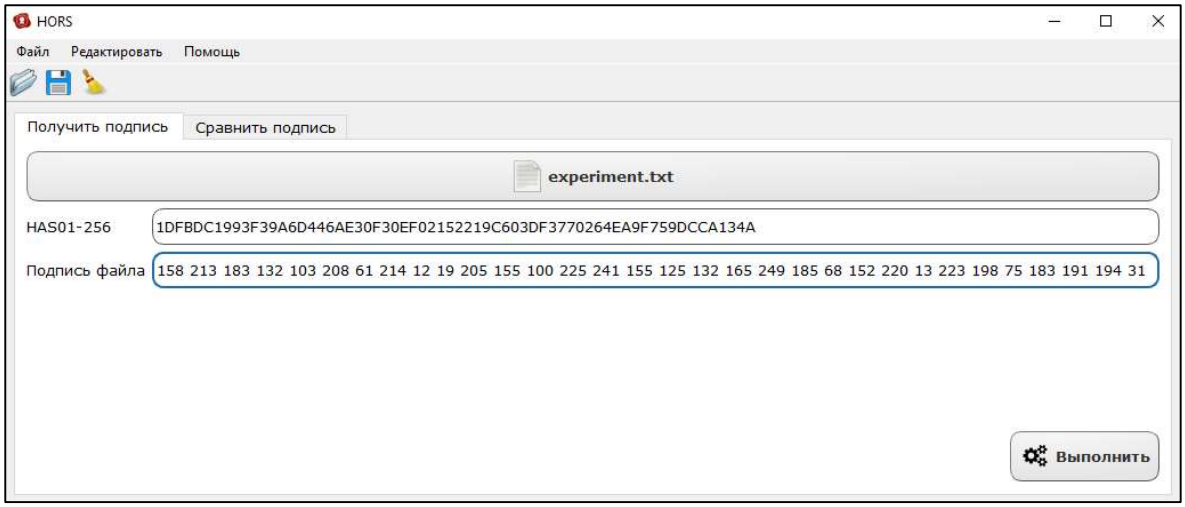


Рис. 1 Получение хеш-образа, цифровой подписи, открытых и закрытых ключей

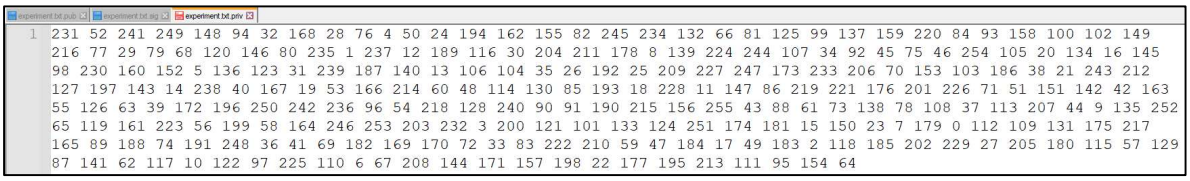


Рис. 2 Файл с секретными ключами

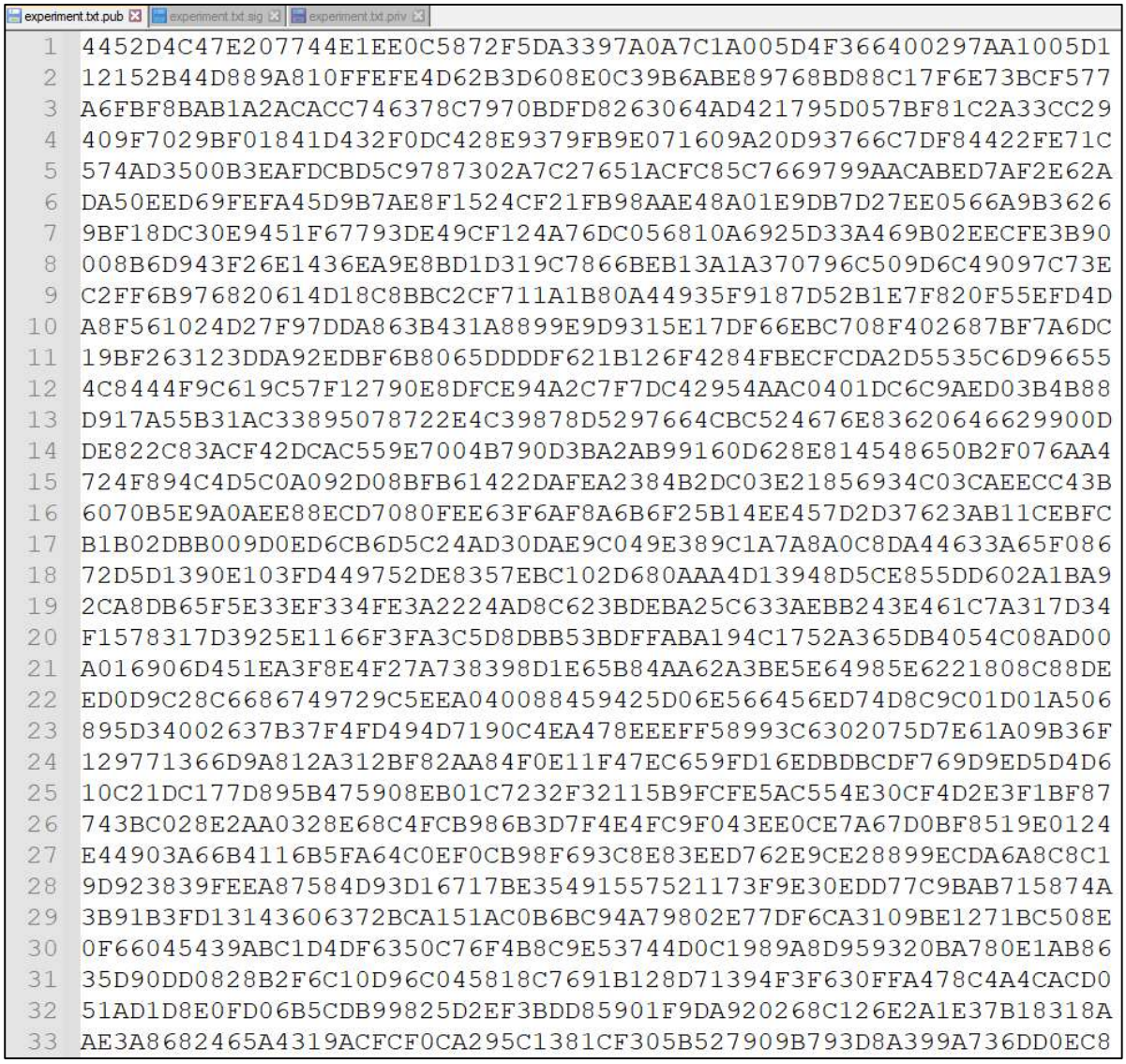


Рис. 3 Файл с открытыми ключами

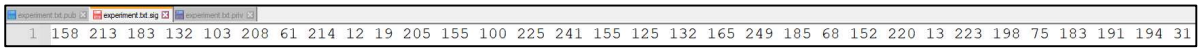


Рис. 4 Файл с цифровой подписью

Для проверки цифровой подписи необходимо выбрать три файла: исходный файл «experiment.txt», файл с открытыми ключами «experiment.txt.pub» и файл

с цифровой подписью «experiment.txt.sig». Так как файлы не подверглись каким-либо изменениям, было получено сообщение, что «подпись верна» (Рисунок 5).

Для имитации создания инцидента, было внесено одно изменение в файл цифровой подписи, т.е. произведена замена первого элемента подписи, который соответствует значению «158» на «157» (Рисунок 6). В результате проверки трех файлов, было получено сообщение, что «подпись неверна» (Рисунок 7).

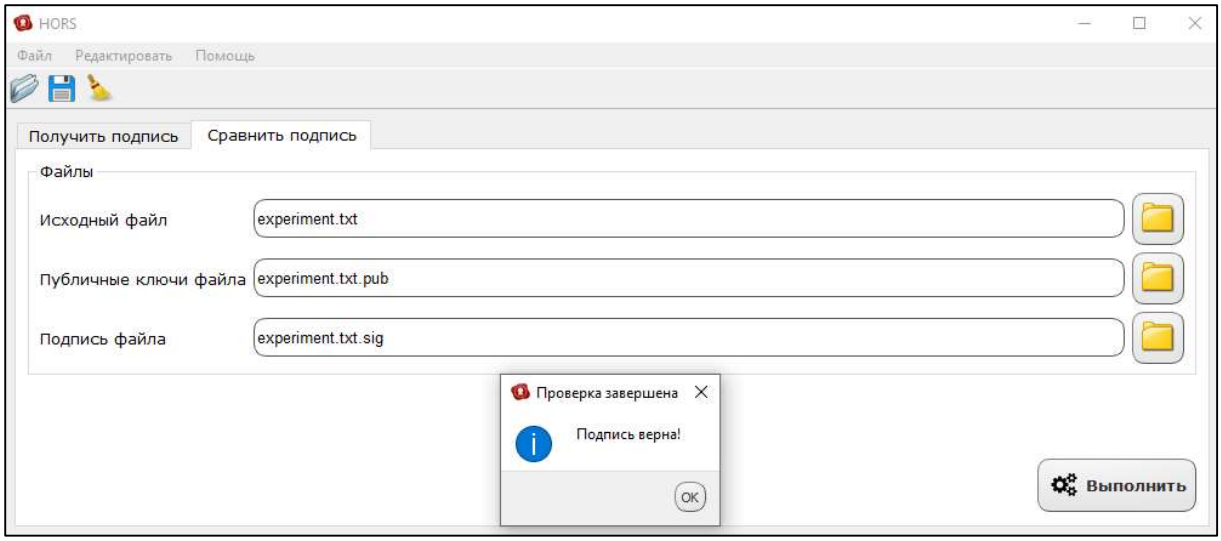


Рис. 5 Проверка цифровой подписи

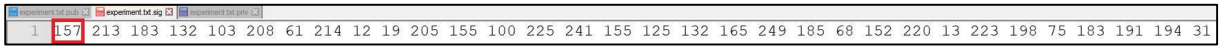


Рис. 6 Файл с измененным значением цифровой подписью

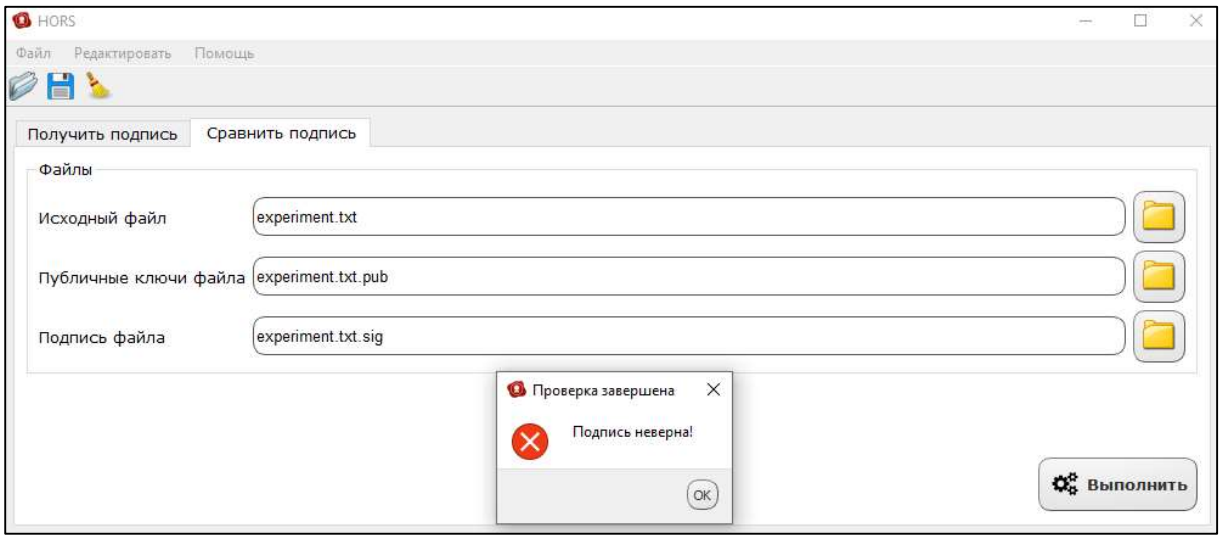


Рис. 7 Проверка цифровой подписи

На основе разработанной программы, были изучены варианты реализации ранее разработанных алгоритмов хеширования с целью дальнейшего их применения в разрабатываемой схеме постквантовой подписи.

Заключение

Таким образом, квантовые технологии являются одними из самых динамичных и конкурентных в сфере высоких технологий. В этой области проводят исследования как традиционные ИТ-гиганты, академические и правительственные организации, так и стартапы. В связи с этим, проведение исследований в области постквантовой криптографии и разработка квантово-устойчивых криптографических алгоритмов играет ключевую роль в обеспечении информационной безопасности.

Сотрудниками ЛИБ на основе рассмотренного постквантового алгоритма HORS и проведенных исследований была разработана общая схема нового постквантового алгоритма цифровой подписи.

Благодарность. Работа выполнена в рамках проекта грантового финансирования AP14870719 «Разработка и исследование алгоритмов пост-квантовой криптографии, основанных на хеш-функциях» КН МНВО РК.

Список литературы:

1. Квантовые компьютеры для всех: уже здесь и сейчас или грядущее науки? - https://habr.com/ru/sandbox/191266/ (25.09.2023).
2. Квантовая криптография против квантовых компьютеров: кто сильнее. - https://rg.ru/2023/07/13/kvantovaia-kriptografiia-protiv-kvantovyh-kompiuterov-kto-silnee.html (26.09.2023).
3. Конкурс NIST по выбору квантово-устойчивых криптоалгоритмов. - https://qapp.tech/help/nist (27.09.2023).
4. *Марухненко А.С. Алгоритмы многоразовой подписи* на основе *хеш-функций.* - https://openarchive.nure.ua/server/api/core/bitstreams/fafb2d77-7e3e-43e0-b98f-10ac815b56e 0/content (25.09.2023).
5. Стандартизация постквантовой криптографии. - https://csrc.nist.gov/Projects/post-quantum-cryptography/post-quantum-cryptography-standardization (27.09.2023).
6. Как работают постквантовые цифровые подписи на основе хэша? (Часть 2) - http://surl.li/loros (26.09.2023).

ӘЛЕУМЕТТІК МЕДИАДАҒЫ ЭКСТРЕМИСТІК МАЗМҰНДЫ АНЫҚТАУ ЖӘНЕ ЗЕРТТЕУ

Ш.Ж.Мусиралиева, Ж.Б.Медетбек , М.А.Болатбек ,
Жумаханова А.Н., Ж.Ы.Елтай
e-mail: bolatbek.milana@gmail.com
Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы

Аңдатпа. *Мақала бойынша жүргізілген зерттеу барысында әлеуметтік желілерден корпусты құру үшін парсер программаларын және API технологиялар көмегімен іздеуді орындау қажет. Қазіргі таңда экстремистік көзқарастар ұлтшылдық, нәсілге бөлу, зорлық зомбылық, діни үгіттеу және т.б. жүзеге асырылады. Экстремизм шектен тыс радикализммен, терроризммен және нигилизммен ұштасады. Экстремизмнің себептері азаматтардың бір бөлігінің әлеуметтік бағдарлануында, тиісті білімнің болмауында, қоғамның дағдарыстық жағдайында, сондай-ақ құқықтық жүйенің тиімсіздігінде жатыр. Экстремизмнің негізгі түрлері: сепаратизм, ксенофобия, ұлтшылдық, шовинизм, расизм/нәсілшілдік, фашизм және діншілдік.*

Корпус құрғаннан соң, құрастырылған корпуста мәтіндерді белгілеу, экстремистік және бейтарап мәтіндердің үлесі шамалас болуы керек; құрастырылған корпус бойынша машиналық оқыту алгоритмдерін оқыту; машиналық оқыту алгоритмдері ішінен жоғары көрсеткіш көрсеткен алгоритмді таңдау; таңдалған үлгі бойынша 2 жаңа ақпараттық ресурсқа талдау; корпус мәліметтеріне қандай препроцессинг алгоритмінің қолданылғандығын көрсету; корпусқа word2vec моделін қолдану; корпусқа bag of words алгоритмін қолдану; корпустағы әр категорияға wordcloud шығару жүзеге асырылды.

Түйінді сөздер*: машиналық оқыту, корпус, VK API, bag of words алгоритмі, wordcloud моделі, word2vec моделі.*

Кіріспе

Экстремизм дегеніміз – қоғамдық ережелер мен нормаларды жоққа шығаратын экстремалды және радикалды іс-әрекеттерді орындау және көзқарастарды ұстану. Бүкіл тарихи даму барысында экстремизм тақырыбы адамның нәсілдік, әлеуметтік, ұлттық және тілдік қатыстылығында көрініс табады. Бұл үрдіс халықтың дінді, тілді және т.б. еркін таңдай алуын, және де мүддесі мен қауіпсіздігін түбегейлі бұзу болып табылады. Экстремизмнің көрінісі бір адамның немесе топтың жалпы бұқараға апаратын белгілі бір идеологияда көрінеді. Экстремистер ерекше, құқықтары мен дамуы жағынан өздерін басқалардан жоғары деп санайды. Бұл белгілі бір әлеуметтік топқа деген жеккөрініш пен өшпенділікті тудырады. Экстремизм әлеуметтік, діни, ұлттық және саяси салаларда жиі кездеседі. Экстремизмнің бірнеше түрі бар:

- Саяси экстремизм
- Діни экстремизм
- Жастар экстремизмі және т.б.

Экстремизмнің негізі – қандай да бір идеялық мазмұнмен және мағынамен толтырылған агрессивтілік. Экстремизмді әртүрлі әлеуметтік немесе мүліктік жағдайы, ұлттық және діни құрамы, кәсіби және жоғары білім деңгейі бар және т.б. бар адамдар жүзеге асыра алады. Қазіргі таңда жастардың экстремистік мінез-құлқының өзіндік ерекшелігі бар: ұлттық, діни және саяси себептермен зорлық-зомбылық жасау. Экстремизм ауқымдылығы бойынша келесідей ерекшеленеді: жеке индивидтің экстремистік көріністері; шағын топтардың (100 адамға дейін) және ірі топтардың (100 адамнан астам) экстремистік көріністері; мемлекеттер мен ұлтаралық бірлестіктердің саясатындағы экстремистік көріністер [1].

Интернет желілерінде экстремисттік мәліметтердің таралуына қарсы тұру қазіргі заманда мемлекеттік органдар мен қоғамның өзекті мәселелерінің бірі болып табылады. Жаһандық мәліметтерді тарату, экстремисттік идеяларды насихаттау, террористік топтарды құру қазіргі кезде интернет желілерін пайдалану көмегімен іске асып жатыр. Террористер бүгінде интернеттің мүмкіндіктерін пайдалану арқылы өзінің жұмысын жүзеге асырып жатыр.

Әдебиеттерге шолу

Бұл мақала жаңа іздеу әдісін ұсынады, ол авторлардың көңіл-күйін талдау арқылы онлайн форумдардағы ең радикалды пайдаланушыларды анықтайды. Әр пайдаланушының хабарламаларының бірнеше сипаттамалары, соның ішінде олардың хабарламаларын орналастыру кезіндегі мінез-құлқы және мазмұны зерттеледі. Мазмұн сөйлеу бөліктерінің тегтерін, көңіл-күйді талдау және "радикалды авторларды көңіл-күйге негізделген сәйкестендіру" деп аталатын жаңа алгоритмді қолдана отырып талданды, ол пайдаланушының орташа көңіл-күй деңгейіне пайыздық ұпайын, теріс хабарламалардың көлемін, теріс хабарламалардың ауырлығын және теріс хабарламалардың ұзақтығын ескереді [2].

Мақалада деректерді, атап айтқанда экстремистік веб-сайттардан жинау үшін дербес веб-сканер құру тәсілін ұсынды. Веб-іздеушінің жетекші компонентіне көңіл-күйге негізделген жіктеу ережелерін қолдану арқылы қол жеткізілді, бұл іздеушіге жүктелген веб-парақтың мазмұны туралы шешім қабылдауға мүмкіндік берді. осы беттерде жиі кездесетін кілт сөздерді іздеу үшін сөйлеу бөліктерін белгілеу (POS) қолданылды. Көңіл-күйді анықтайтын бағдарламалық жасақтаманы жіктеу бағдарламалық жасақтамасымен бірге қолдана отырып, белгілі бір беттің қай сыныпқа кіретінін тиімді анықтай алатын шешім ағашы жасалды. Алынған ағаш төрт сынып арасындағы дифференциация кезінде 80% сәтті және экстремистік беттерді нақты жіктеуде 92% сәтті көрсетті [3].

“Заманауи экстремизм: түрлері мен мәні” атты мақалада автор экстремизмнің түрлерін қарастырады. Экстремизмнің түрлерінің көп болғанына байланысты оның мәнін ашуға және түрлі салдарларын болжауға тырысады. Жәнеде автор мақалада экстремизмнің негізгі әлеуметтік факторларын қарастырады.

https://scholar.google.com/scholar?hl=ru&as_sdt=0%2C5&q=%D0%AD%D0%9A%D0%A1%D0%A2%D0%A0%D0%95%D0%9C%D0%98%D0%A1%D0%A2%D0%A2%D0%86%D0%9A+%D0%9C%D3%98%D0%9B%D0%86%D0%9C%D0%95%D0%A2%D0%A2%D0%95%D0%A0%D0%94%D0%86+%D0%86%D0%97%D0%94%D0%95%D0%A3%D0%94%D0%86%D2%A2+%D0%97%D0%90%D0%9C%D0%90%D0%9D%D0%90%D0%A3%D0%98+%D3%99%D0%B4%D1%96%D1%81%D1%82%D0%B5%D1%80%D1%96&btnG=

“Қазақстандағы ақпараттық технологиялар арқылы діни экстремизмнің насихатталуы” атты мақалада мақалада исламтану және дінтану саласына қатысты

ақпараттық технологиялардың Қазақстан аумағындағы діни экстремизмнің таралуына әсері, күші қарастырылады. Қазақстанның қазіргі кездегі діни сауаттылық жағдайына жасалған талдау діни мәліметтерді ақпараттық технологиялар көздерінен алудың қауіптілігіне тұжырым жасауға мүмкіндік береді. Зерттеу кезінде қол жеткізген нәтижелері: қазіргі таңда адамдар кез келген ақпарат алу үшін ақпараттық технологиялардың көмегіне жүгінеді деген тұжырым. Бірақ әлеуметтік желілердегі және сайттардағы ақпараттардың барлығының бірдей қауіпсіз әрі сенімді емес екендігіне көз жеткіземіз делінген.

https://bulletin-religious.kaznu.kz/index.php/relig/article/view/221

“ Интернет-коммуникациядағы кибербуллинг: пайда болуы, өзекті мәселелері” атты мақалада кибербуллингтің құрбаны болмауы үшін ата-аналар тарапынан қандай бақылау керектігі туралы кеңестер беріледі. Сондай-ақ, ата-аналар арасында сауалнама жүргізіліп, статистика ұсынылды. Қазіргі жаһандану заманында шеті көрінбейтін Интернет әлемі үлкен ақпараттық банкке айналды, оны пайдаланушылар саны да күн сайын өсіп келеді. Интернет желісі қолданушыларға үлкен мүмкіндік берді, бірақ сонымен қатар бір-біріне залалын тигізуі мүмкін әртүрлі ақпаратты жасауда шексіз еркіндікке жол ашты делінген. Сондай ақ бұл мақалада интернет-коммуникацияның жаңа құбылысы – кибербуллинг әлеуметтік проблема ретінде қарастырылып, оны шешу жолдары қарастырылған.

https://scholar.google.com/scholar?hl=ru&as_sdt=0%2C5&q=%D0%91%D1%83%D0%BB%D0%BB%D0%B8%D0%BD%D0%B3%D1%82%D1%96+%D1%82%D0%B0%D0%B1%D1%83+&btnG=

Мәліметтерді жинақтау

Кейбір әлеуметтік желілерде өздерінің API – лары бар. Ол арқылы біз өзімізге қажетті мәліметтерді оңай әрі еркін жинай аламыз. Басқаша айтқанда, әлеуметтік желілерімен жұмыс жасауға арналған интерфейс.

Әлеуметтік желілерде және интернет желісінде экстремисттік мәліметті іздеуді үшін парсерлерді және әлеуметтік желілердің API технологияларын пайдалу көмегімен іске асырылды.

VK API – ВКонтакте әлеуметтік желісімен жұмыс істеуге арналған арнайы интерфейс. VK API көмегімен ВКонтакте серверіне арнайы запростарды жіберу арқылы өзімізге қажетті мәліметті алуға болады. VK API-мен жұмыс істеу үшін бізге арнайы токен қажет. VK API көмегімен құрылған сұраныс мысалы:
https://api.vk.com/method/users.get?user_id=12345678&v=5.95
Серверден келген жауаптар JSON форматта қайтарылады (1-сурет).

```
Ввод [1]: import vk
          import csv
          import pandas as pd

Ввод [5]: access_token = 'ce7320345a00c808a072d6f69bc721bbce78fc3c68188e32e6ddfa228f36491b66858fadce5426d1ff374'
          v = '5.95'
          keyword_city = "Алматы"
          keywords_groups = ['сирия', 'ислам', 'талиб', 'бомба', 'война']

Ввод [6]: session = vk.Session(access_token=access_token)
          api = vk.API(session, v=v)
```

1-сурет VK API көмегімен экстремисттік топтарды анықтау.

Қажетті кітапханаларды импорт жасаймыз және VK API-мен жұмыс істеуге арналған токенді енгіземіз. ВКонтакте серверлерімен жұмыс істеуге арналған сессияны анықтаймыз.

Әлеуметтік желіден **“сирия”, “ислам”, “талиб”, “бомба” және “война”** кілттік сөздер көмегімен топтарды іздейміз. Топтарды Алматы қаласынан іздейміз [4].

Алдымен топтардың идентификаторларын алу үшін groups әдісі көмегімен аламыз. Алынған идентификаторларды groups.getById әдісінде қолданып, топтар жайлы негізгі мәліметтерді аламыз: топтын атауы, қатысушылар саны, топтың типі, байланыс информация, топтың суреті және т.б. Осы мәліметтерді csv форматты файлға жазып аламыз. Аталған процессті барлық кілттік сөздер үшін қолданып шығамыз (2-3сурет).

```
for keyword in keywords_groups:
    group_ids = [g['id'] for g in api.groups.search(city_id=city_id, q=keyword, sort=2, count=500)['items']]
    Groups = api.groups.getById(fields='members_count,contacts', group_ids=group_ids)
    filename = city_name + " - " + keyword + ".csv"
    f = open(filename, 'w', encoding='utf-8')
    fieldnames = ['sort2', 'id', 'type', 'is_closed', 'screen_name', 'name', 'members_count', 'contacts', 'photo_50', 'photo_100'
    writer = csv.DictWriter(f, fieldnames=fieldnames)
    writer.writeheader()
    i = 1
    for group in Groups:
        group['sort2'] = i
        writer.writerow(group)
        i+=1
```

2-сурет Кілттік сөздер

```
pd.read_csv('Алматы - ислам.csv')
```

	sort2	id	type	is_closed	screen_name	name	members_count	contacts	photo_50	
0	1	81413715	page	0	islam.akikat	ИСЛАМ - АҚИҚАТ☝	91731.0	NaN	https://sun2.dataix-kz-akkol.userapi.com/s/v1/...	akl
1	2	79682163	page	0	alhamdulillah__1	Alhamdulillah ~ Ислам	276664.0	[{'user_id': 121587618, 'desc': 'Реклама '}, {...	https://sun2.dataix-kz-akkol.userapi.com/s/v1/...	akl
2	3	36506095	group	0	thequranicislam	Коранический Ислам	7508.0	[]	https://sun1.dataix-kz-akkol.userapi.com/s/v1/...	akl
3	4	117848149	page	0	islam_dini_alla_zholyinda	Ислам діні. Алла жолында.	9572.0	[]	https://sun1.dataix-kz-akkol.userapi.com/s/v1/...	akl
4	5	84126080	group	0	club84126080	Ислам тағылымдары	1920.0	[]	https://sun1.dataix-kz-akkol.userapi.com/s/v1/...	akl
...	...	...	...	...	...	...	...	...	...	
213	214	58398885	group	0	dostarisl	ISL Dostar	40.0	[]	https://sun2.dataix-kz-akkol.userapi.com/s/v1/...	akl
214	215	91918648	group	0	club91918648	♡♥islam♡♥	14.0	[{'user_id': 268923526, 'desc': '·····♥·····'}, ...	https://sun2.dataix-kz-akkol.userapi.com/s/v1/...	akl
215	216	66738089	group	0	club66738089	Dinim ISLAM	12.0	[{'user_id': 203853355}]	https://sun2.dataix-kz-akkol.userapi.com/s/v1/...	akl
216	217	28342524	group	0	club28342524	islam	13.0	[{'user_id': 135197740}, {'user_id': 52215560}]	https://sun9-60.userapi.com/c9736/g28342524/e_...	60.userapi.com/
217	218	5455962	group	0	club5455962	ISL "DOSTAR" THE CHAMPIONS (football 2008)	33.0	[{'user_id': 11862966, 'desc': 'smotritel)'}, ...	https://sun9-73.userapi.com/c298/g5455962/c_d2...	73.userapi.com/

218 rows × 11 columns

3-сурет Қолданылған процесс нәтижелері

YouTube API қолдану

YouTube API – Google-дің YouTube әлеуметтік бейне-хостинг желісімен жұмыс істеуге арналған интерфейс. YouTube API көмегімен желідегі бейнелер жайлы толық информацияны алуға болады: жалпы мәліметтер, көру саны, лайктер және дизлайктер саны, пікірлер, бейненің ұзындығы және т.б. Интерфейсті қолдану үшін арнайы токенді және әзірлеуші кілтін алу керек.

YouTube API көмегімен **“сирия”, “ислам”, “талиб”, “бомба” және “война”** сияқты кілттік сөздер бойынша табылған бейнелердің пікірлерін және жауаптарын аламыз [5].

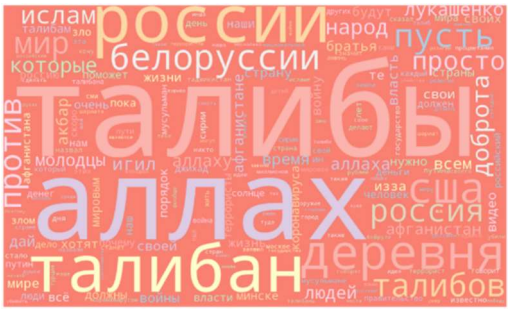
Қажетті кітапханаларды жүктеп, орнатамыз. Токенді, әзірлеуші кілтін және бейне идентификаторын анықтаймыз. Әрі қарай пікірлерді алуға және олардың жауаптарын алуға арналған әдістерді құрамыз. Пікірлер commentThreads() және comments() әдістері көмегімен алынады. YouTube API сұраныстарын орындау үшін execute() әдісі шақырылады.API көмегімен алынған мәліметтерді қолданып, корпусты құрамыз.

Корпус дегеніміз – лингвистикада қандай да бір әдістермен таңдалған және өңделген тексттер жиынтығы. Корпустағы тексттар саны 2000. Әрбір текст экстремисттік және бейтарап (status) белгіленген. Корпустағы экстремисттік және бейтарап тексттер саны шамалас. Корпустағы экстремисттік және бейтарап тексттер қатынасы шамалас.

Bag of words дегеніміз – тексттегі сөздердің кездесуін сипаттайтын мәтіннің көрінісі. Bag of words құру үшін sklearn кітапханасының CountVectorize() әдісін қолданамыз. Мысал ретінде корпустан кез келген екі сөйлемді қарастырамыз.

Word2vec дегеніміз – тексттегі сөздерді сандар векторына айналдыру әдісі. Word2vec екі архитектурасы бар: CBOW және Skip gram.
Word2vec құру үшін genism кітапханасын қолданамыз. Алдымен корпусты токендерге бөлеміз, сосын Word2Vec әдісі көмегімен word2vec моделін құрамыз.

Wordcloud дегеніміз – текстте сөздердің кездесуіне байланысты оларды визуализациялау әдісі [6]. Wordcloud құру үшін wordcloud кітапханасын қолданамыз (4-сурет).



4-сурет Нәтижесі

Экстремистік мәтіндерді анықтау үшін машиналық оқыту әдістерін пайдалану

Машиналық оқыту деп – оқуға қабілетті алгоритмдерді құруды айтамыз. Біздің жағдай да біз машиналық оқыту көмегімен текстті класификациялау есебін шешуге арналған машиналық оыту әдістерін қолданамыз.
Текстті классификацияладың бірнеше әдістерін қарастырамыз

Алдымен корпусты үйрету және тексеру датасеттеріне бөлу керек. Ол үшін sklearn кітапханасының train_test_split әдісін қолданамыз [7]. Корпустың 1500 текстерін модельді үйретуге және 501 мәтінді тексеруге қолданамыз (5-сурет).

```
from sklearn.model_selection import train_test_split

x_train, x_valid, y_train, y_valid = train_test_split(x, y, test_size = 0.25, random_state = 42)

print(x_train.shape)
print(x_valid.shape)
print(y_train.shape)
print(y_valid.shape)

(1500, 2500)
(501, 2500)
(1500,)
(501,)
```

5-сурет train_test_split әдісін қолдану

RandomForrestClassifier – мәліметтер жиынының әртүрлі ішкі үлгілеріндегі бірқатар шешім ағашының жіктеуіштеріне сәйкес келетін және болжамдық дәлдікті жақсарту және артық сәйкестікті бақылау үшін орташалауды пайдаланатын бағалаушы.

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix
from sklearn.metrics import f1_score

model = RandomForestClassifier()
model.fit(x_train, y_train)

y_pred = model.predict(x_valid)

print("Үйрету дәлдігі :", model.score(x_train, y_train))
print("Тексеру дәлдігі :", model.score(x_valid, y_valid))

Үйрету дәлдігі : 0.9986666666666667
Тексеру дәлдігі : 0.9021956087824351
```

6-сурет RandomForrestClassifier әдісі

LogisticRegression - классификацияның сызықтық үлгісі. Бұл модельде бір сынақтың ықтимал нәтижелерін сипаттайтын ықтималдықтар логистикалық функцияның көмегімен модельденеді.

```
from sklearn.linear_model import LogisticRegression

model = LogisticRegression()
model.fit(x_train, y_train)

y_pred = model.predict(x_valid)

print("Үйрету дәлдігі :", model.score(x_train, y_train))
print("Тексеру дәлдігі :", model.score(x_valid, y_valid))

Үйрету дәлдігі : 0.9986666666666667
Тексеру дәлдігі : 0.8842315369261478
```

7-сурет LogisticRegression әдісі

DecisionTreeClassifier классификация және регрессия үшін пайдаланылатын параметрлік емес бақыланатын оқыту әдісі болып табылады [8].

```
from sklearn.tree import DecisionTreeClassifier

model = DecisionTreeClassifier()
model.fit(x_train, y_train)

y_pred = model.predict(x_valid)

print("Үйрету дәлдігі :", model.score(x_train, y_train))
print("Тексеру дәлдігі :", model.score(x_valid, y_valid))

Үйрету дәлдігі : 0.9986666666666667
Тексеру дәлдігі : 0.8582834331337326
```

8-сурет DecisionTree lassifier әдісі

Үш әдістің ішінде ең үлкен нәтижені көрсеткен әдіс – RandomForrestClassifier. Үйрету дәлдігі – 99.8%. Тексеру дәлдігі – 90%.

Болжау әдісін тексеру үшін қосымша тексеру арналған корпусты жинаймыз. Корпусты жинау үшін VK API, YouTube API және т.б. әдістерді қолданамыз.

Қорытынды. Қорытындылай келе, жұмысты орындау барысында әлеуметтік желілерден қажетті мәліметтерді алу үшін API технологияларын, парсер программаларын қолдануды қарастырдық. Әлеуметтік желілерде кілттік сөздерді көмегімен алынған мәліметтердің көмегімен корпусты құрдық. Корпуста бейтарап және экстремисттік мәліметтерді қарастырдық. Корпус мәліметтеріне токенизация, пунктуациядан тазалау, көп кездесетін сөздерден тазалау, стоп сөздерден тазалау сияқты препроцессинг алгоритмдерін қолдандық. Корпус статистикасын және визуализациясын қарастырдық.

Word2vec, bag of words және n-gram алгоритмдері мен модельдерін қолдандық және корпусты машиналық оқытуға дайындадық. Машиналық оқыту барысында бірнеше алгоритмдерді қарастырып, ең жоғары нәтиже көрсеткен алгоритм көмегімен болжау әдісін құрастырдық. Болжау әдісі көмегімен қосымша тестілеу корпусындағы экстремисттік тексттер үлесін анықтадық.

Берілген зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті қаржыландыратын “Қазақ тіліндегі кибер экстремизмнің идеологиялық бағыттарын жасанды интеллект әдістері көмегімен мультиклассификациялау” жобасы аясында орындалды (грант AP19676342, жоба жетекшісі Мусиралиева Ш.Ж.).

Пайдаланылған әдебиеттер тізімі

1. https://analyticsindiamag.com/hands-on-implementation-of-tf-idf-from-scratch-in-python/ 20.09.2023
2. https://kavita-ganesan.com/tfidftransformer-tfidfvectorizer-usage-differences/ 18.09.2023

3. https://www.geeksforgeeks.org/python-word-embedding-using-word2vec/ 23.09.2023
4. https://stackabuse.com/classification-in-python-with-scikit-learn-and-pandas/ 23.09.2023
5. https://www.learndatasci.com/glossary/binary-classification/ 19.09.2023
6. https://www.thepythoncode.com/article/text-classification-using-tensorflow-2-and-keras-in-python 19.09.2023
7. https://github.com/aquibjaved/Deep-learning-for-text-classification 24.09.2023
8. https://scikit-learn.org/stable/modules/linear_model.html 24.09.2023

LWC ШИФРЛАУ АЛГОРИТМІНЕ ДИФФЕРЕНЦИАЛДЫҚ КРИПТОТАЛДАУ

Дюсенбаев Д.С., Хомпыш А.
e-mail: dimash_dds@mail.ru, ardabek@mail.ru
ҚР ҒжЖБМ «Ақпараттық және есептеуіш технологиялар институты»

Аңдатпа. *Құпия ақпараттарды қорғауға арналған блокты шифрлау алгоритмдерінің криптоберіктілігін анықтау маңызды жұмыстардың бірі. Криптоберіктілікті бағалау үшін зерттеліп отырған алгоритм мүмкіндігінше бірнеше әдістер арқылы сыналуы тиіс. Бұл мақалада жаңа құрылған LWC шифрлау алгоритміне дифференциалдық криптоталдау жүргізу кезеңдері мен алынған нәтижелер баяндалған. Орындалған зертеу жұмысын қортындылай келе, LWC шифрлау алгоритмнің дифференциалдық талдауға криптографиялық тұрақтылығы жоғары екендігі анықталды.*

Кілт сөздер: *жеңіл салмақты шифрлау алгоритмі, криптографиялық тұрақтылық, дифференциалдық криптоталдау.*

Кіріспе

Қазіргі заманда криптографиялық шифрлау алгоритмдерінің ішінде симметриялы блокты шифрлар кең қолданысқа ие. Бірақ, мұндай шифрлау алгоритмдерін есептеу ресурстары қатаң шектеулі құрылғыларда қолдану тиімсіз болып саналады. Оның бір себебі, блокты алгоритмдардың шифрлау, кілт жасау процедураларының күрделілігінде, сондай ақ аппаратты түрде жүзеге асыру барысында қолданылатын операциялардың көптігіне байланысты аз көлемдегі шекті ресурстарға қолдану тиімсіз.

Мұндай шекті ресурстардың талаптарына сай қолданылатын ақпаратты криптографиялық қорғаудың жаңа бағыты жеңілсалмақты криптография екендігі белгілі. Жеңілсалмақты криптография пайда болуының нәтижесінде шектеулі есептеу ресурстармен қамтылған интернетке кіру мүмкіншілігі бар әр түрлі смарт құрылғыларда мәліметтерді криптографиялық қорғау мәселесі шешімін тапты. Қазіргі уақытта жеңіл салмақты криптографиялық алгоритмдер құрумен айналысатын ғалымдардың саны өте көп, солардың ішінде ерекше атап өтетін жұмыстарға Present [1,2], Clefia [3], Katan [4], Simon [5,6], Speck [6,7], Secure IoT(SIT) [8] және т.б. алгоритмдері жатқызуға болады. Бұл жұмыстарда авторлар алгоритмдердің құрылымында қолданылған криптографиялық түрлендіру әдістеріне кеңінен тоқтала отырып, алгоритмдердің сенімділігін зерттеген. Осы бағыттағы ғылыми жұмыстарға талдау жүргізе келе, басты сұлбасы мен примитивтерін негізге ала отырып жаңа LWC жеңілсалмақты шифрлау алгоритмі құрылды [9].

Бұл мақалада LWC шифрлау алгоритміне дифференциалдық криптоталдау әдісі бойынша криптографиялық тұрақтылығына бағалау жүргізілді. Айтылған әдістердің барлығы сызықтық емес функцияларға негізделгендіктен, алдымен сызықтық емес функцияларға жеке-жеке талдау жүзеге асырылды. Сондықтан берілген алгоритмнің схемасынан сызықтық емес функцияларды қалауымызша жіктеп немесе біріктіріп талдауға қолайлы болатындай бөліп аламыз және ескеретін жағдай, ол талдаудың нәтижесіне әсер етпейтіндей таңдалуы тиіс.

S түрлендіруі екі төрт биттік S-блоктан тұрғандықтан бір сегіз биттік S-блок ретінде қарауға болады, бірақ біз екі төрт биттік S-блок ретінде қарастырдық, өйткені талдаудың нәтижесі бірдей болады. SP түрлендіруін төрт төрт биттік S-блок ретінде қарастырамыз (кесте 1-6).

1-кесте. S1 – блогі

x	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
Sbox1	2	E	D	6	8	A	B	1	5	3	4	9	0	F	7	C

2-кесте. S2 – блогі

x	0	1	2	3		4	5	6	7	8	9	A	B	C	D	E	F
Sbox2	F	5	D	8		C	2	4	7	0	9	6	A	1	3	E	B

3-кесте. S1P түрлендіруінің бірінші S-блогі

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	7	B	5	2	6	E	8	9	C	1	A	0	F	D	3

4-кесте. S2P түрлендіруінің екінші S-блогі

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
F	9	B	2	3	4	1	D	0	A	5	6	8	C	7	E

5-кесте. PS1 түрлендіруінің үшінші S-блогі

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2	5	8	0	E	3	A	F	D	4	B	7	6	9	1	C

6-кесте. PS2 түрлендіруінің төртінші S-блогі

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
F	0	C	1	5	9	2	3	D	6	4	E	8	A	7	B

LWC шифрлау алгоритміне дифференциалдық криптоталдау

Дифференциалды криптоталдау - симметриялы блокты шифрлардың криптоталдау әдісі. Бұл әдісті алғаш рет 1990 жылы Израиль мамандары Эли Бихам мен Ади Шамир ұсынған. Дифференциалды криптоталдау әртүрлі шифрлау раундарында шифрланатын мәндер арасындағы айырмашылықты түрлендіруге негізделген. Шабуыл жасау үшін белгілі бір айырмашылықпен байланысты ашық мәтіндердің жұптары қолданылады. LWC алгоритмының схемасына сәйкес сызықтық емес функцияларды сызықтық талдаудағыдай алты блоктың негізінде қарастырамыз. Сондықтан алдымен сол блоктарға кіріс пен шығыс айырымдарына қатысты төмендегідей алты айырымдары бойынша кестесін аламыз [10,11,12] (кесте 7-12).

7-кесте. S1-блогі үшін кіріс және шығыс айырымдар кестесі

Δ	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	2	0	0	0	2	0	0	0	2	4	2	2	0	2
2	0	2	0	4	0	0	0	2	2	0	2	2	0	0	0	2
3	0	2	0	2	2	0	0	2	2	2	0	0	4	0	0	0
4	0	0	0	2	2	4	2	2	0	0	2	0	2	0	0	0
5	0	0	0	2	0	0	2	0	4	0	2	0	2	2	2	0
6	0	0	2	0	2	2	2	0	0	2	0	0	2	0	0	4
7	0	0	0	2	2	2	0	2	0	4	0	2	0	0	2	0
8	0	0	0	0	0	2	0	2	2	2	0	0	2	4	0	2
9	0	2	2	0	2	0	2	4	0	0	2	2	0	0	0	0
A	0	0	0	0	0	2	4	2	2	0	0	2	0	0	2	2
B	0	2	0	2	4	0	0	0	0	0	2	2	0	2	2	0
C	0	2	2	0	0	0	0	0	2	2	4	0	0	2	0	2
D	0	4	2	0	0	2	0	0	0	0	0	2	0	2	2	2
E	0	0	4	2	0	2	0	0	0	2	0	0	2	2	2	0
F	0	2	2	0	2	0	2	0	2	2	0	0	0	0	4	0

8-кесте. S2 үшін кіріс және шығыс айырымдар кестесі

Δ	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	2	2	0	4	0	0	0	2	2	0	2	0	2	0
2	0	0	2	2	0	2	2	0	4	0	0	0	0	2	0	2
3	0	0	0	0	0	0	2	2	2	0	4	2	0	2	0	2
4	0	4	0	2	0	0	0	2	2	2	2	0	0	0	0	2
5	0	0	0	2	2	0	0	0	2	2	2	0	2	4	0	0
6	0	2	4	0	0	0	0	2	0	2	2	2	0	0	2	0
7	0	2	0	0	2	2	0	2	2	0	0	4	0	0	0	2
8	0	2	2	0	0	0	0	0	0	0	2	2	4	2	0	2
9	0	0	0	2	0	2	2	2	0	2	0	0	0	0	2	4
A	0	2	2	0	2	2	0	0	0	4	0	0	0	2	0	2
B	0	0	0	2	2	2	2	4	2	0	0	0	2	0	0	0
C	0	0	2	4	0	0	2	0	0	0	0	2	2	2	2	0
D	0	2	2	0	2	2	4	0	0	0	0	0	2	0	2	0
E	0	2	0	0	2	0	0	0	2	0	2	2	2	0	4	0
F	0	0	0	0	4	0	2	2	0	2	0	2	0	2	2	0

9-кесте. S1P түрлендіруінің бірінші S-блогі үшін кіріс және шығыс айырымдар кестесі

Δ	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	2	2	0	4	0	0	0	2	2	0	2	0	2	0

2	0	0	2	2	0	2	2	0	4	0	0	0	0	2	0	2
3	0	0	0	0	0	0	2	2	2	0	4	2	0	2	0	2
4	0	4	0	2	0	0	0	2	2	2	2	0	0	0	0	2
5	0	0	0	2	2	0	0	0	2	2	2	0	2	4	0	0
6	0	2	4	0	0	0	0	2	0	2	2	2	0	0	2	0
7	0	2	0	0	2	2	0	2	2	0	0	4	0	0	0	2
8	0	2	2	0	0	0	0	0	0	0	2	2	4	2	0	2
9	0	0	0	2	0	2	2	2	0	2	0	0	0	0	2	4
A	0	2	2	0	2	2	0	0	0	4	0	0	0	2	0	2
B	0	0	0	2	2	2	2	4	2	0	0	0	2	0	0	0
C	0	0	2	4	0	0	2	0	0	0	0	2	2	2	2	0
D	0	2	2	0	2	2	4	0	0	0	0	0	2	0	2	0
E	0	2	0	0	2	0	0	0	2	0	2	2	2	0	4	0
F	0	0	0	0	4	0	2	2	0	2	0	2	0	2	2	0

10-кесте. S2P түрлендіруінің екінші S-блогі үшін кіріс және шығыс айырымдар кестесі

Δ	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	2	2	0	2	2	0	4	2	0	2	0	0	0
2	0	0	4	0	2	2	0	0	0	2	0	2	2	0	0	2
3	0	0	2	0	0	2	4	0	0	0	0	2	0	2	2	2
4	0	0	2	0	0	0	2	0	4	0	2	0	2	2	0	2
5	0	2	2	2	0	0	2	0	0	0	2	4	2	0	0	0
6	0	0	0	0	4	0	2	2	2	0	2	0	0	2	2	0
7	0	2	2	0	0	0	0	0	2	2	0	0	0	2	4	2
8	0	0	0	4	2	0	2	0	2	0	0	2	0	0	2	2
9	0	0	0	0	0	2	0	2	0	2	2	0	2	2	0	4
A	0	2	0	0	2	0	0	0	2	2	4	2	0	0	0	2
B	0	2	2	2	0	2	0	0	0	2	0	0	2	4	0	0
C	0	0	0	2	2	2	0	2	0	0	0	2	4	0	2	0
D	0	2	0	2	2	4	0	2	2	2	0	0	0	0	0	0
E	0	2	2	2	0	0	2	4	2	0	0	0	0	0	2	0
F	0	4	0	0	0	2	0	2	0	0	2	2	0	2	2	0

11-кесте. PS1 түрлендіруінің үшінші S-блогі үшін кіріс және шығыс айырымдар кестесі

Δ	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	2	0	2	2	2	0	0	2	4	0	2
2	0	0	0	2	2	4	2	2	0	0	2	0	2	0	0	0
3	0	2	2	0	0	0	0	0	2	2	4	0	0	2	0	2
4	0	0	2	0	0	0	2	0	0	0	2	4	2	2	0	2
5	0	2	2	0	2	0	2	4	0	0	2	2	0	0	0	0
6	0	0	0	2	0	0	2	0	4	0	2	0	2	2	2	0

7	0	4	2	0	0	2	0	0	0	0	0	2	0	2	2	2
8	0	2	0	4	0	0	0	2	2	0	2	2	0	0	0	2
9	0	0	0	0	0	2	4	2	2	0	0	2	0	0	2	2
A	0	0	2	0	2	2	2	0	0	2	0	0	2	0	0	4
B	0	0	4	2	0	2	0	0	0	2	0	0	2	2	2	0
C	0	2	0	2	2	0	0	2	2	2	0	0	4	0	0	0
D	0	2	0	2	4	0	0	0	0	0	2	2	0	2	2	0
E	0	0	0	2	2	2	0	2	0	4	0	2	0	0	2	0
F	0	2	2	0	2	0	2	0	2	2	0	0	0	0	4	0

12-кесте. PS2 түрлендіруінің төртінші S-блогі үшін кіріс және шығыс айырымдар кестесі

Δ	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	2	2	0	0	0	0	0	0	0	2	2	4	2	0	2
2	0	4	0	2	0	0	0	2	2	2	2	0	0	0	0	2
3	0	0	2	4	0	0	2	0	0	0	0	2	2	2	2	0
4	0	0	2	2	0	4	0	0	0	2	2	0	2	0	2	0
5	0	0	0	2	0	2	2	2	0	2	0	0	0	0	2	4
6	0	0	0	2	2	0	0	0	2	2	2	0	2	4	0	0
7	0	2	2	0	2	2	4	0	0	0	0	0	2	0	2	0
8	0	0	2	2	0	2	2	0	4	0	0	0	0	2	0	2
9	0	2	2	0	2	2	0	0	0	4	0	0	0	2	0	2
A	0	2	4	0	0	0	0	2	0	2	2	2	0	0	2	0
B	0	2	0	0	2	0	0	0	2	0	2	2	2	0	4	0
C	0	0	0	0	0	0	2	2	2	0	4	2	0	2	0	2
D	0	0	0	2	2	2	2	4	2	0	0	0	2	0	0	0
E	0	2	0	0	2	2	0	2	2	0	0	4	0	0	0	2
F	0	0	0	0	4	0	2	2	0	2	0	2	0	2	2	0

Осы жоғарыдағы кестелердегі мәндер кіріс пен шығыс айырымдар қатынасының сәйкес санын дифференциал салмағы деп, $[\Delta x, \Delta y]$ белгілейміз, мұндағы $\Delta x = x \oplus x' = (x_1, x_2, x_3, x_4) \oplus (x_1', x_1', x_1', x_1') = (x_1 \oplus x_1', x_2 \oplus x_2', x_3 \oplus x_3', x_4 \oplus x_4')$ және $\Delta y = y \oplus y' = (y_1, y_2, y_3, y_4) \oplus (y_1', y_1', y_1', y_1') = (y_1 \oplus y_1', y_2 \oplus y_2', y_3 \oplus y_3', y_4 \oplus y_4')$.

$$\{[\Delta x, \Delta y] = 4 | [1,14], [2,12], [3,3], [4,9], [5,2], [6,15], [7,10], [8,11], [9,13], [10,5], [11,1], [12,6], [13,8], [14,4], [15,7]\}$$

жиыны $\frac{4}{16} = \frac{1}{4}$ ықтималдылықпен бірінші кестеден кездесетін айырдардың дифференциал салмағы. Осы сияқты қалған айырым кестелерінде салмағы осындай әр кестеде он бестен кездеседі. Осы айырым кестелері мен алгоритмнің сұлбасын ескеріп әрбір цикл үшін дифференциалдық талдау нәтижелерін алуға арналған бағдарлама жазылды, ол нәтижелер төменде көрсетілген.

Әрбір і-ші циклдегі кіріс және шығыс мәтіндерін сәйкесінше $u_i = u_{i0}, u_{i1}, u_{i2}, \dots, u_{i15}$, $v_i = v_{i0}, v_{i1}, v_{i2}, \dots, v_{i1}$. олардың айырымдары $[\Delta u_i, \Delta v_i]$, мұндағы $\Delta u_i = u_i \oplus u_i{}', \Delta v_i = v_i \oplus v_i{}'$.

1-мысал.

Δu = 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 – ашық мәтіндердің айырымы болсын.

1-ші циклдың кіріс айырымы:

Δu_1 = 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

<<< 5 цикл бірінші бөлік үшін айырымды Δw_i деп белгілесек:

Δw_1 = 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0

1-ші циклдың шығыс айырымы:

Δv_1 = 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 - $\left(\frac{1}{4}\right)^2 = 2^{-4}$

Δu_2 = 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0

Δw_2 = 2 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0

Δv_2 = 5 0 9 12 5 0 0 12 0 0 0 12 0 0 9 12 - $\left(\frac{1}{4}\right)^4 = 2^{-8}$

Δu_3 =5 0 9 12 5 0 0 12 0 0 0 12 0 0 9 12

Δw_3 = 1 3 8 10 5 0 0 12 0 0 0 12 0 0 9 12

Δv_3 = 3 13 5 9 3 2 5 9 3 2 6 14 3 13 6 9 - $\left(\frac{1}{4}\right)^9 = 2^{-18}$

Δu_4 =3 13 5 9 3 2 5 9 3 2 6 14 3 13 6 9

Δw_4 = 10 11 2 7 3 2 5 9 3 2 6 14 3 13 6 9

Δv_4 = 9 7 1 2 3 10 2 10 5 6 13 5 1 11 15 13 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$

Δu_5 =9 7 1 2 3 10 2 10 5 6 13 5 1 11 15 13

Δw_5 = 14 2 5 2 3 10 2 10 5 6 13 5 1 11 15 13

Δv_5 = 3 0 14 9 0 10 4 10 7 1 2 4 1 8 3 8 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$

Δu_6 =3 0 14 9 0 10 4 10 7 1 2 4 1 8 3 8

Δw_6 = 1 13 2 6 0 10 4 10 7 1 2 4 1 8 3 8

Δv_6 = 13 3 15 13 9 6 9 2 5 13 5 14 8 10 14 14 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$

Δu_7 = 13 3 15 13 9 6 9 2 5 13 5 14 8 10 14 14

Δw_7 = 7 15 11 10 9 6 9 2 5 13 5 14 8 10 14 14

Δv_7 = 3 0 1 12 11 15 15 8 2 2 1 10 5 8 15 4 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$

Δu_8 = 3 0 1 12 11 15 15 8 2 2 1 10 5 8 15 4

Δw_8 = 0 3 8 6 11 15 15 8 2 2 1 10 5 8 15 4

Δv_8 = 9 8 5 10 15 5 7 14 3 13 6 0 1 11 12 1 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$

Δu_9 = 9 8 5 10 15 5 7 14 3 13 6 0 1 11 12 1

Δw_9 = 0 11 5 3 15 5 7 14 3 13 6 0 1 11 12 1

Δv_9 = 3 11 7 8 0 8 8 10 7 3 13 0 5 7 7 7 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$

Δu_{10} = 3 11 7 8 0 8 8 10 7 3 13 0 5 7 7 7

Δw_{10} = 6 15 0 7 0 8 8 10 7 3 13 0 5 7 7 7

Δv_{10} = 9 14 14 14 11 14 10 6 0 6 1 15 3 7 14 15 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$

Δu_{11} = 9 14 14 14 11 14 10 6 0 6 1 15 3 7 14 15

Δw_{11} = 13 13 13 3 11 14 10 6 0 6 1 15 3 7 14 15

Δv_{11} = 3 15 2 4 15 15 1 10 4 3 5 8 1 15 6 15 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$

Δu_{12} = 3 15 2 4 15 15 1 10 4 3 5 8 1 15 6 15

$\Delta w_{12} =$ 14 4 8 7 15 15 1 10 4 3 5 8 1 15 6 15
$\Delta v_{12} =$ 12 2 1 6 7 13 12 11 3 6 8 4 6 7 13 14 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{13} =$ 12 2 1 6 7 13 12 11 3 6 8 4 6 7 13 14
$\Delta w_{13} =$ 4 2 13 8 7 13 12 11 3 6 8 4 6 7 13 14
$\Delta v_{13} =$ 15 15 0 14 15 7 12 8 4 8 2 9 13 3 1 3 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{14} =$ 15 15 0 14 15 7 12 8 4 8 2 9 13 3 1 3
$\Delta w_{14} =$ 14 1 13 15 15 7 12 8 4 8 2 9 13 3 1 3
$\Delta v_{14} =$ 1 3 11 13 14 5 3 15 4 4 9 4 13 6 4 14 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{15} =$ 1 3 11 13 14 5 3 15 4 4 9 4 13 6 4 14
$\Delta w_{15} =$ 7 7 10 2 14 5 3 15 4 4 9 4 13 6 4 14
$\Delta v_{15} =$ 7 2 3 10 13 0 0 15 15 1 14 10 6 7 2 13 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{16} =$ 7 2 3 10 13 0 0 15 15 1 14 10 6 7 2 13
$\Delta w_{16} =$ 4 7 4 14 13 0 0 15 15 1 14 10 6 7 2 13
$\Delta v_{16} =$ 14 2 7 15 0 4 8 10 11 11 14 9 11 12 14 13 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$

1-мысалдың кілтті анықтаудағы он алты раундтағы дифференциалдық талдаудың ықтималдылығы $2^{-4} \cdot 2^{-8} \cdot 2^{-18} \cdot (2^{-32})^{13} = 2^{-446}$ тең.

2-мысал.
$\Delta u_1 =$ 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
$\Delta w_1 =$ 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
$\Delta v_1 =$ 0 0 5 0 0 0 0 0 0 0 0 0 0 0 0 0 - $\frac{1}{4} = 2^{-2}$
$\Delta u_2 =$ 0 0 5 0 0 0 0 0 0 0 0 0 0 0 0 0
$\Delta w_2 =$ 0 10 0 0 0 0 0 0 0 0 0 0 0 0 0 0
$\Delta v_2 =$ 0 0 10 0 0 0 10 0 0 0 10 0 0 0 10 0 - $\left(\frac{1}{4}\right)^{2} = 2^{-4}$
$\Delta u_3 =$ 0 0 10 0 0 0 10 0 0 0 10 0 0 0 10 0
$\Delta w_3 =$ 1 4 0 0 0 0 10 0 0 0 10 0 0 0 10 0
$\Delta v_3 =$ 15 0 8 8 0 0 8 8 0 0 8 14 15 0 8 14 - $\left(\frac{1}{4}\right)^{4} = 2^{-8}$
$\Delta u_4 =$ 15 0 8 8 0 0 8 8 0 0 8 14 15 0 8 14
$\Delta w_4 =$ 1 1 1 14 0 0 8 8 0 0 8 14 15 0 8 14
$\Delta v_4 =$ 6 13 7 3 13 5 7 3 13 11 9 14 6 13 9 14 - $\left(\frac{1}{4}\right)^{11} = 2^{-22}$
$\Delta u_5 =$ 6 13 7 3 13 5 7 3 13 11 9 14 6 13 9 14
$\Delta w_5 =$ 10 14 6 13 13 5 7 3 13 11 9 14 6 13 9 14
$\Delta v_5 =$ 10 10 3 13 14 8 9 2 8 7 7 5 13 11 13 10 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_6 =$ 10 10 3 13 14 8 9 2 8 7 7 5 13 11 13 10
$\Delta w_6 =$ 4 7 11 5 14 8 9 2 8 7 7 5 13 11 13 10
$\Delta v_6 =$ 14 11 12 12 5 14 7 8 2 15 14 9 5 6 1 5 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_7 =$ 14 11 12 12 5 14 7 8 2 15 14 9 5 6 1 5
$\Delta w_7 =$ 7 9 9 13 5 14 7 8 2 15 14 9 5 6 1 5
$\Delta v_7 =$ 13 12 1 14 4 15 2 1 6 7 15 10 9 15 14 12 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$

$\Delta u_8 =$ 13 12 1 14 4 15 2 1 6 7 15 10 9 15 14 12
$\Delta w_8 =$ 8 3 13 11 4 15 2 1 6 7 15 10 9 15 14 12
$\Delta v_8 =$ 5 13 5 8 0 9 5 9 4 14 6 11 14 2 12 6 - $\left(\frac{1}{4}\right)^{16} = 2^{-3}$
$\Delta u_9 =$ 5 13 5 8 0 9 5 9 4 14 6 11 14 2 12 6
$\Delta w_9 =$ 10 11 0 11 0 9 5 9 4 14 6 11 14 2 12 6
$\Delta v_9 =$ 9 4 5 10 8 12 15 15 0 14 13 5 7 1 2 5 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{10} =$ 9 4 5 10 8 12 15 15 0 14 13 5 7 1 2 5
$\Delta w_{10} =$ 8 11 5 2 8 12 15 15 0 14 13 5 7 1 2 5
$\Delta v_{10} =$ 1 7 3 1 2 8 0 8 7 1 13 11 15 7 14 0 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{11} =$ 1 7 3 1 2 8 0 8 7 1 13 11 15 7 14 0
$\Delta w_{11} =$ 14 6 2 2 2 8 0 8 7 1 13 11 15 7 14 0
$\Delta v_{11} =$ 0 11 0 7 14 15 4 6 5 1 4 4 5 2 7 6 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{12} =$ 0 11 0 7 14 15 4 6 5 1 4 4 5 2 7 6
$\Delta w_{12} =$ 6 0 14 1 14 15 4 6 5 1 4 4 5 2 7 6
$\Delta v_{12} =$ 10 14 0 14 1 4 2 6 9 12 0 15 15 6 1 11 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{13} =$ 10 14 0 14 1 4 2 6 9 12 0 15 15 6 1 11
$\Delta w_{13} =$ 12 1 13 5 1 4 2 6 9 12 0 15 15 6 1 11
$\Delta v_{13} =$ 13 4 15 8 6 1 14 13 4 15 9 6 14 3 11 8 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{14} =$ 13 4 15 8 6 1 14 13 4 15 9 6 14 3 11 8
$\Delta w_{14} =$ 9 15 1 10 6 1 14 13 4 15 9 6 14 3 11 8
$\Delta v_{14} =$ 10 3 5 0 7 0 13 9 13 2 1 13 14 6 7 7 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{15} =$ 10 3 5 0 7 0 13 9 13 2 1 13 14 6 7 7
$\Delta w_{15} =$ 6 10 1 4 7 0 13 9 13 2 1 13 14 6 7 7
$\Delta v_{15} =$ 4 6 12 5 15 7 1 6 13 5 10 15 12 7 10 14 - $\left(\frac{1}{4}\right)^{16} = 2^{-32}$
$\Delta u_{16} =$ 4 6 12 5 15 7 1 6 13 5 10 15 12 7 10 14
$\Delta w_{16} =$ 13 8 10 8 15 7 1 6 13 5 10 15 12 7 10 14
$\Delta v_{16} =$ 15 10 5 10 4 2 13 14 15 8 3 8 2 11 1 12 - $\left(\frac{1}{4}\right)^{16} = 2^{-3}$

2-мысалдың кілтті анықтаудағы он алты раундтағы дифференциалдық талдаудың ықтималдылығы $2^{-2} \cdot 2^{-4} \cdot 2^{-8} \cdot 2^{-22} \cdot (2^{-32})^{12} = 2^{-420}$ тең.

Қорытынды

Ұсынылып отырған жеңіл салмақты LWC шифрлау алгоритмі статистикалық талдаулар, оның ішінде НИСТ жинағы бойынша алдынған нәтижелер, оң статистикалық бағалауларға ие болатын. Ал бұл мақалада алгоритмнің сенімділігін анықтаудың екінші бір маңызды талдауы - дифференциалдық криптоталдау әдісі бойынша зерттелді. Алынған нәтижелер LWC шифрлау алгоритмі дифференциалдық криптоталдауға тұрақты екендігін көрсетті.

Алғыс

Жұмыс AP09259570 «Шектелген ресурстар үшін отандық жеңілсалмақты шифрлау алгоритмін құру және зерттеу» гранттық қаржыландыру ғылыми жобасы аясында жүргізілді.

Әдибеттер

1. Tang Zh., Cui J., Zhong H., Yu M., A Random PRESENT Encryption Algorithm Based on Dynamic S-box International [Journal of Security and Its Applications], 2016, vol. 10, no. 3, pp.383-392 http://dx.doi.org/10.14257/ijsia.2016.10.3.33
2. Suhail A., Mir N., Mehvish A., Ishfaq S., Tariq B. M. FPGA Implementation of PRESENT Block Cypher with Optimised Substitution Box [2022 Smart Technologies], Communication and Robotics (STCR), 2022, pp.1-6. doi: 10.1109/STCR55312.2022.10009366.
3. Shirai T., Shibutani T., Akishita K., et al. The 128-bit blockcipher CLEFIA [FSE 2007. LNCS.], 2007. v. 4593. pp. 181–195.
4. Qatan F. M., Damaj I. W. High-speed KATAN ciphers on-a-chip [Computer Systems and Industrial Informatics (ICCSII)], 2012 International Conference on, 2012, pp. 1–6.
5. Aysu E., Gulcan P., Schaumont. SIMON says: Break area records of block ciphers on FPGAs, [IEEE Embed Syst Lett], 2014, vol.6 , pp. 37–40. https://doi.org/10.1109/les.2014.2314961
6. Yustiarini B.Y., Dewanta F., Nuha H. H. A Comparative Method for Securing Internet of Things (IoT) Devices: AES vs Simon-Speck Encryptions [2022 1st International Conference on Information System & Information Technology (ICISIT)], Indonesia, 2022, pp. 392-396, doi: 10.1109/ICISIT54091.2022.9872666.
7. Beaulieu R., Treatman-Clark S. D., Shors B., Weeks J., Smith and Wingers L. The SIMON and SPECK lightweight block ciphers [2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)], San Francisco, USA, 2015, pp. 1-6, doi: 10.1145/2744769.2747946.
8. Muhammad U., Irfan A., Imran M., Shujaat Kh., Usman A, Sh. SIT: A Lightweight Encryption Algorithm for Secure Internet of Things, [(IJACSA) International Journal of Advanced Computer Science and Applications], 2017, vol. 8, no. 1, pp 402-411.
9. Khompysh, A., Kapalova, N., Lizunov, O., Dyusenbayev, D., Sakan, K., (2023). Development of a New Lightweight Encryption Algorithm. International Journal of Advanced Computer Science and Applications (IJACSA), 14 (5), 452-459. https://doi.org/10.14569/IJACSA.2023.0140548
10. Biham E., Shamir A. Differential Cryptanalysis of the Full 16-round DES // Crypto'92. – Springer-Velgar, 1998. – P. 487.
11. Biham E., Shamir A. Differential Cryptanalysis of DES-like Cryptosystems Extended Abstract // Crypto'92. - Springer-Velgar, 1998. - P.21.
12. Бабенко Л.К., Ищукова Е.А. Современные алгоритмы блочного шифрования и методы их анализа. – М.: Гелиос АРВ, 2006. – 376 с.

ПРИМЕНЕНИЕ SAT-РЕШАТЕЛЕЙ ДЛЯ ОЦЕНКИ СТОЙКОСТИ КРИПТОГРАФИЧЕСКИХ АЛГОРИТМОВ

Варенников А.В., Сақан Қ.С.
e-mail: avarennikov@ipic.kz, kairat_sks@mail.ru
Институт информационных и вычислительных технологий КН МНВО РК, Казахстан

Аннотация. *В данной статье представлена методология применения современных SAT-решателей для оценки стойкости криптографических алгоритмов и приведены экспериментальные данные, полученные при исследовании криптостойкости разработанной хеш-функции HBC-256 с использованием современных SAT-решателей.*

Введение

В области информатики SAT-решатели представляют собой алгоритмы, а также реализующие их компьютерные программы, предназначенные для решения задачи выполнимости булевых формул (или просто SAT-задачи). В качестве входных данных SAT-решатель получает некоторую формулу над пропозициональными переменными, принимающими значение «истина» или «ложь», например, такую как: $(x \vee y) \wedge (x \vee \overline{y})$, и выводит ответ, является ли эта формула выполнимой, то есть существуют ли такие значения x и y, которые делают эту формулу истинной, или невыполнимой, что означает, что таких значений x и y не существует. В приведённом примере формула выполнима, если x имеет значение «истина», поэтому SAT-решатель должен вернуть значение SATISFIABLE или сокращенно SAT. Если формула невыполнима, SAT-решатель должен вернуть значение UNSATISFIABLE или UNSAT.

В соответствии с теоремой Кука [1], проблема выполнимости булевых формул, в общем случае, является NP-полной. Это означает, что алгоритмы поиска решения задачи выполнимости имеют экспоненциальную сложность. Тем не менее, в настоящее время разработаны эффективные алгоритмы для SAT, которые позволяют решать задачи, включающие десятки тысяч переменных и миллионы ограничений.

SAT-решатели начинают свою работу с упрощения булевой формулы, приведённой к конъюнктивной нормальной форме и представленной с использованием стандартного формата, разработанного Центром дискретной математики и теоретической информатики (DIMACS) [2]. Чаще всего они основываются на алгоритме поиска с возвратом Дэвиса-Патнема-Лонгемана-Лавленда (DPLL) [3, 4], но включают ряд расширений и дополнений. Большинство SAT-решателей поддерживают ограничения по времени, что позволяет им завершиться в разумные сроки даже, если они не могут найти решение, выдав соответствующее сообщение (например, UNKNOWN). Очень часто SAT-решатели не просто выводят ответ, но и предоставляют дополнительную информацию, например, значения для x, y и т.д. в случае, если формула выполнима, или, если формула невыполнима, минимальный набор выражений, входящих в состав этой формулы, которые не могут быть выполнены.

Современные SAT-решатели оказали значительное влияние на такие области как верификация программного обеспечения, анализ программ, решение задач с заданными ограничениями, искусственный интеллект, автоматизация электронного проектирования и исследование операций.

1. Основы логического криптоанализа

В статье Ф. Массаччи и Л. Марраро [5] было показано, что низкоуровневые свойства современных криптографических алгоритмов можно свести к поиску решения SAT-задачи, а затем с помощью эффективных SAT-решателей сделать необходимые выводы. Данный подход получил название логического криптоанализа.

Основная идея логического криптоанализа состоит в том, что битовые последовательности, рассматриваются как последовательности пропозициональных переменных, в которых каждая переменная истинна, когда соответствующий бит равен 1, и ложна, когда соответствующий бит равен 0. Теперь нам нужно закодировать свойства криптографического алгоритма в виде логической формулы $\mathcal{E}(\mathcal{P},\mathcal{K},\mathcal{C})$, которая принимает истинное значение тогда и только тогда, когда для соответствующих последовательностей битов выполняется условие $\mathrm{C} = \mathrm{E_K}(\mathrm{P})$, где P, K, C – битовые последовательности, представляющие открытый текст, ключ шифрования и шифртекст, а $\mathcal{P}$, $\mathcal{K}$, $\mathcal{C}$ – соответствующие наборы пропозициональных переменных.

Чтобы смоделировать поиск ключа, обозначим через $v_{\mathcal{C}}$ значения истинности пропозициональных переменных, соответствующих битам известного шифртекста. Тогда, для нахождения ключа с применением атаки на основе шифртекста достаточно найти модель для $\mathcal{E}(\mathcal{P},\mathcal{K},v_{\mathcal{C}})$, то есть значения истинности остальных используемых в данной формуле пропозициональных переменных, при которых эта формула принимает истинное значение. При атаке на основе открытого текста нам помимо шифртекста также известны значения $v_{\mathcal{P}}$ пропозициональных переменных $\mathcal{P}$. В этом случае нам найти модель для $\mathcal{E}(v_{\mathcal{P}},\mathcal{K},v_{\mathcal{C}})$. При атаке алгоритма хеширования $\mathrm{C} = \mathrm{H}(\mathrm{P})$ методом нахождения прообраза нам нужно найти модель для $\mathcal{H}(\mathcal{P},v_{\mathcal{C}})$ (при условии, что ключ не используется).

2. Пропозициональное кодирование

Формирование пропозициональной формулы $\mathcal{E}(\mathcal{P},\mathcal{K},\mathcal{C})$ или $\mathcal{H}(\mathcal{P},\mathcal{C})$, описывающей логические характеристики криптографического алгоритма, по существу, представляет собой операцию обратного проектирования. Основная идея заключается в том, чтобы пройтись по всему алгоритму, попутно генерируя формулы, соответствующие каждой выполняемой операции. Там, где это возможно, операции выполняются непосредственно над пропозициональными переменными, представляющими входные биты. Например, для операции перестановки пропозициональное кодирование не применяется – вместо этого выполняется перестановка входных битов и на выходе получаем переставленные пропозициональные переменные.

Поскольку нет смысла выполнять такое кодирование вручную, на практике преобразование криптографических алгоритмов в булевы формулы осуществляется с использованием специальных инструментов, которые позволяют записать криптографический алгоритм на высокоуровневом языке программирования, а затем преобразовать полученную программу в соответствующую булеву формулу в конъюнктивной нормальной форме, и сохранить её в файле, имеющем стандартный формат DIMACS для последующего использования в качестве входных данных любыми современными SAT-решателями.

При подготовке экспериментальных данных для этой статьи в качестве такого инструмента использовался программный комплекс Transalg [6]. Данный комплекс предоставляет возможность преобразования в булевы формулы произвольных алгоритмов, вычисляющих всюду определенные дискретные функции, то есть функции вида $f\colon \{0,1\}^* \to \{0,1\}^*$, записанных на специально разработанном Си-подобном языке программирования TA, синтаксис которого описан в [7].

В качестве тестового криптографического алгоритма для проведения испытаний с использованием современных SAT-решателей был выбран алгоритм хеширования данных HBC-256, разработанный сотрудниками лаборатории информационной безопасности Института информационных и вычислительных технологий, преобразующий открытый текст, разбитый на 384-битовые блоки, в 256-битовое хеш-значение с использованием в качестве функции сжатия блочного шифра [8].

Для того, чтобы преобразовать этот алгоритм в булеву формулу, на специализированном языке программирования TA была составлена программа, реализующая все выполняемые им операции, в том числе, побайтовая перестановка, побитовый циклический сдвиг, процедура замены с использованием четырёх 16-байтовых S-блоков, побайтовая операция XOR и сложение элементов выбранных строк и столбцов матрицы. Затем текст этой программы передается на вход программному комплексу Transalg и обрабатывается модулем синтаксического разбора, результатом работы которого является дерево синтаксического разбора, описывающее внутреннее представление этой программы. Обход дерева синтаксического разбора и построение булевых уравнений осуществляет модуль преобразования. При выводе итогового пропозиционального кода осуществляется переход к конъюнктивной нормальной форме при помощи преобразований Цейтина [9]. В результате пропозиционального кодирования алгоритма хеширования HBC-256 была получена булева формула, содержащая 196 864 пропозициональных переменных и состоящая из 987 120 выражений, каждое из которых представляет собой дизъюнкцию из одного или нескольких литералов. Общее количество литералов, то есть пропозициональных переменных или их отрицаний, равно 3 322 520.

После того, как было получено пропозициональное представление алгоритма хеширования HBC-256, необходимо задать значения управляющих переменных $v_{\mathcal{P}-i}$ и $v_{\mathcal{C}}$, и выполнить ряд тестов с применением современных SAT-решателей, чтобы получить информацию о времени выполнения вычислений, использовании оперативной памяти и других экспериментальных данных. В данном случае $v_{\mathcal{C}}$ представляет собой набор из 256 пропозициональных переменных, соответствующих 256-битовом набору известного хеш-значения, а $v_{\mathcal{P}-i}$ представляет собой набор из 384 пропозициональных переменных, соответствующих 384-битовом набору блока открытого текста, в котором неизвестными являются первые i битов. При $i = 0$ открытый текст и хеш-значение известны, и нужно выполнить проверку на их соответствие друг другу (проверка хеш-значения). При $i = 384$ открытый текст неизвестен, то есть нужно найти открытый текст по известному хеш-значению (атака нахождения прообраза).

3. Экспериментальные данные

В таблице 1 приведено время поиска заданного количества i неизвестных битов открытого текста с применением последовательных и параллельных SAT-решателей. Различие между последовательными и параллельными SAT-решателями заключается в том, что последовательные SAT-решатели выполняются в одном потоке, в то время как параллельные SAT-решатели разбивают задачу на подзадачи, которые выполняются в

нескольких параллельных потоках. В общем случае не существует SAT-решателя, который для всех задач показывал бы лучшие результаты по сравнению с другими решателями. Отдельно взятый алгоритм может лучше справляться с задачами, которые с трудом преодолевают другие, и в то же время оказаться гораздо хуже при решении других задач.

Тесты проводились на сервере HP Enterprise/DL380 с двумя процессорами Intel Xeon Gold 6242R 3.10 ГГц и оперативной памятью 128 Гб, под управлением операционной системы Linux Ubuntu Server 22.04.3 LTS. Параметр, задающий количество потоков, используемых параллельными SAT-решателями, был установлен в значение 32. Незаполненные ячейки таблицы 1 означают, что был превышен установленный лимит времени, равный 48 часам (172 800 секунд).

Таблица 1

Время выполнения вычислений (в секундах)

Кол-во битов, i	Последовательные SAT-решатели			Параллельные SAT-решатели		
	CaDiCaL	Lingeling	Kissat	Parkissat	Plingeling	Treengeling
0	0,22	0,05	0,17	0,29	0,70	0,37
4	0,50	24,69	936,94	1,84	12,30	77,50
8	1,99	-	11 505,03	2,84	11,80	41,68
12	43,01	-	45 120,32	8,91	14,70	150,48
16	29,36	-	-	30,88	22,20	1 078,64
20	-	-	-	387,09	144,40	-
24	-	-	-	2 250,77	22 145,10	-
28	-	-	-	48 048,77	-	-
32	-	-	-	-	-	-

Заключение

Результаты, приведённые в таблице 1, показывают, что ни один из используемых SAT-решателей не преодолел отметку в 32 бита, поэтому нахождение всех 384 битов открытого текста не представляется возможным. Таким образом, можно утверждать, что алгоритм хеширования HBC-256 обладает высокой стойкостью к атаке поиска прообраза.

Несмотря на то, что логический криптоанализ уступает по эффективности традиционным методам криптоанализа, описанная методология может быть применена для оценки стойкости любых криптографических алгоритмов.

Данная работа была выполнена в рамках проекта AP14870419 «Разработка средства криптографическом защиты информации для защиты переговоров по КВ радиосвязи» Министерства науки и высшего образования Республики Казахстан.

Литература

1. *Cook S.* The complexity of theorem-proving procedures // Proceedings of the Third Annual ACM Symposium on Theory of Computing, 1971. – pp. 151-158.

2. DIMACS. Satisfiability Suggested Format. 1993. – http://www.domagoj-babic.com/uploads/ResearchProjects/Spear/dimacs-cnf.pdf (дата обращения: 27.09.2023).
3. *Davis, M., Putnam, H.* A Computing Procedure for Quantificational Theory // Journal of the ACM, vol. 7(3), 1960. – pp. 201-215.
4. *Davis, M., Longemann, G., Loveland, D.* A Machinery Program for Theorem-Proving // Communications of the ACM, vol. 5(7), 1962. – pp. 394-397.
5. *Massacci F., Marraro L*. Logical Cryptanalysis as a SAT Problem // Journal of Automated Reasoning. 2000. – vol. 24, no. 1-2. – pp. 165-203.
6. *Отпущенников И. В., Семенов А. А*. Инструментальное средство трансляции алгоритмов вычисления дискретных функций в выражения исчисления высказываний Transalg 1.0. – https://gitlab.com/transalg/transalg (дата обращения: 26.09.2023).
7. *Отпущенников И. В.* Методы и средства преобразования процедурных описаний дискретных функций в булевы уравнения: дисс. … канд. техн. наук: 05.13.11. – Иркутск, 2011. – 128 с.
8. *Сақан Қ. С.* Разработка алгоритмов хеширования на основе итеративных блочных шифров и исследование их криптостойкости: дисс. … док. филос. (PhD): 27.09.23. – Алматы, 2023. – 103 с.
9. *Цейтин Г. С.* О сложности вывода в исчислении высказываний // Записки научных семинаров ЛОМИ АН СССР. 1968. Т. 8. – С. 234–259.

DEVELOPMENT OF A USER AUTHENTICATION USER AUTHENTICATION METHOD FOR WEB RESOURCES

[1]Beknazarov S.I., [2]Bidakhmet Zh.,
e-mail: sundetnet105@gmail.com
Al-Farabi Kazakh National University, Almaty, Kazakhstan[1, 2]

Annotation. The improvement of user authentication for web resources is the focus of this article's research. In order to address the issue of the shortcomings of current authentication techniques, the paper suggests a novel strategy based on two-factor authentication with the Django framework. The development of a scalable global two-factor authentication system is then suggested, consisting of one-time passwords and password authentication in addition to a unique test after signing in. The Django framework is the primary tool used to put the suggested strategy into practice. The study's findings demonstrate that the created method enables getting around the drawbacks of current authentication techniques and ensuring the security, adaptability, and usability of web resources.

I. Introduction

Overview of user authentication on Web resources

User authentication is an important aspect of ensuring the security and confidentiality of user information on web resources. This allows web applications to verify the identity of users and control their access to available features and data.

Since Web Applications are widely used for both business and personal purposes, the authentication system of such applications has a high importance.

This review discusses various aspects of user authentication on web resources, in particular, taking into account the use of the Django framework. We'll look at key authentication methods, password-based authentication, social media authentication, and two-factor authentication, with a focus on their implementation in Django.

Multiple factor authentication is recommended to increase security, such as combining passwords with biometric data or smart cards. A new form of authentication gaining popularity is cookie authentication, which uses an HTTP cookie to identify a user to a web service. This is convenient for users, as they don't need to remember their login details every time they visit the website. However, if cookies are not encrypted properly, they can be stolen and pose security risks[1].

Authentication security is crucial for protecting web resources and user accounts from unauthorized access. It ensures user privacy, mitigates unauthorized access attempts, builds user trust, ensures regulatory compliance, and prevents account takeovers. By implementing strong authentication measures, organizations create a secure environment, safeguard sensitive data, and provide a positive user experience.

The purpose of this review is to identify a wide range of opinions and recommendations for developing and testing new authentication methods within the Django platform.

Traditional authentication methods

Password-based authentication is a process where users are required to provide their login information, including a username and password, to validate their identity.

Two-factor authentication (2FA) is implemented to enhance the protection of both a user's login information and the resources they can access. It offers a higher level of security compared to single-factor authentication (SFA), where users provide only one form of verification, typically a password or passcode.

Biometric authentication. This uses user's biological characteristics, such as fingerprints, facial recognition, or voice, etc. to verify identity.

Hardware Key authentication. Using a physical device, such as a USB key or a hardware token, to authenticate the user.

Authentication by email or SMS. The user receives a unique authentication code via email or SMS and enters it to confirm his identity.

Authentication via social networks. The user can use the credentials of his social network account to log in to other platforms or services.

Certificate authentication. The use of electronic certificates to verify and verify the identity of the user.

Single Sign-On (SSO) is an authentication method that allows users to securely access multiple applications and websites using a single set of credentials. This method is based on a trust relationship established between the service provider (an application) and an identity provider[1].

II. RelatedWork

S. Vaithyasubramanian and others [2] propose a two-factor authentication scheme using an array password generated by a Petri net model. The scheme requires the user to enter a combination of characters and select predefined cells in a 2D grid, which serves as the second factor. The scheme is evaluated for security, efficiency, and usability, and compared with other existing schemes.

Various techniques, such as facial and fingerprint recognition, have been proposed for healthcare monitoring to address biometric recognition problems.

Ortiz B.L. and others [3] propose a user authentication system that utilizes a Long Short-Term Memory (LSTM) model to recognize users based on their typing patterns. This is a novel Bi-LSTM-based architecture for smartphone PPG biometrics classification and demonstrated its performance through multiple experiments on raw- and feature-based data. The system captures and analyzes user keystroke dynamics to determine whether the user is authentic or not.

Carrillo-Torres, D and others[4] propose a Multi-Factor Authentication (MFA) based on a non-biometric mechanism that does not require additional hardware. The novelty of the proposed mechanism lies in a two-factor authentication algorithm which requires a user to identify specific images out of a set of randomly selected images, then the user is required to establish a self-pre-configured relation between two given images to complete authentication. A functional prototype of the proposed system was developed and deployed. The proposed system was tested by users of different backgrounds achieving 100% accuracy in identifying and authenticating users, if authentication elements and credentials were not forgotten [4].

The novelty of the proposed mechanism lies in a two-factor authentication algorithm which requires a user to identify specific images out of a set of randomly selected images, then the user is required to establish a self-pre-configured relation between two given images to complete authentication.

R. F. Olanrewaju and others[5] propose a new method for user authentication in web-based premium applications. The proposed method involves the use of a one-time password (OTP) generated by the user's mobile device, which is then verified by the server. The authors

claim that this method is more secure and user-friendly than traditional authentication methods such as passwords or biometrics.

Yu, L and others[6] propose a blockchain-based authentication and authorization scheme for distributed mobile cloud computing services. The authors aim to address the security and privacy issues in the existing authentication and authorization schemes.

Critical factors of successful method

Making a list of critical factors for a good method requires careful analysis and a systematic approach. In such a list, the following factors should be taken into account:

- Security is provided by reliable encryption mechanisms;
- Easy to use;
- Scalability;
- Interface.

Table 1.

Evaluation of the advantages and disadvantages of the proposed methods

Research Paper	Advantages	Disadvantages
Blockchain-Based Authentication and Authorization Scheme for Distributed Mobile Cloud Computing Services	Detailed explanatory of the proposed scheme and the use of blockchain technology to ensure secure authentication and authorization	Lack of discussion on the scalability of the proposed scheme
User Authentication Recognition Process Using Long Short-Term Memory Model	The use of a deep learning model like LSTM allows for a more accurate and efficient user authentication process compared to traditional methods. The system is able to continuously adapt and learn from user behavior, improving its accuracy over time.	Focuses only on user authentication. The system relies solely on keystroke dynamics, The system requires the user to perform a typing task every time they log in
Two Factor Authentication for Secured Login Using Array Password Engender by Petri net	The scheme uses a 2D grid as the second factor, which could be more resistant to attacks like keylogging or shoulder surfing	Outdated, and may not account for newer security threats. The scheme may not be scalable for applications that require authentication for a large number of users, as the array password needs to be pre-generated for each user
Novel Multi-Factor Authentication Algorithm Based on Image Recognition and User Established Relations	The use of image recognition as a factor in the authentication process may provide additional security, the use of user-established relations may help prevent attacks based on social engineering	No detailed analysis of the effectiveness of the proposed algorithm in real-world scenarios
A Frictionless and Secure User Authentication in Web-Based Premium Applications	Provides an additional layer of security by using OTPs that expire quickly, making it more difficult for attackers to gain unauthorized access. Also, it is user-friendly designed and convenient enough	Limited to premium web-based applications only, may not be applicable to other contexts

Weaknesses of existing methods

An article titled "A Frictionless and Secure User Authentication in Web-Based Premium Applications" will be used as the main source work for further development. This work serves as an important research source on the basis of which a new idea will be developed and presented.

To solve the weakness of the existing method, expressed in its limited applicability on a large scale, it is proposed to create a scalable universal method of two-factor authentication on a web resource using the Django framework.

As a result, based on the work "A Frictionless and Secure User Authentication in Web-Based Premium Application", a new idea will be developed that will serve as an extension and improvement of existing authentication methods, and make an important contribution to the field of user authentication in premium web applications.

III. Development of a new authentication method

The Django authentication system handles both authentication and authorization. Briefly, authentication verifies a user is who they claim to be, and authorization determines what an authenticated user is allowed to do. Here the term authentication is used to refer to both tasks[7].

The auth system consists of:

- Users
- Permissions: Binary (yes/no) flags designating whether a user may perform a certain task.
- Groups: A generic way of applying labels and permissions to more than one user.
- A configurable password hashing system
- Forms and view tools for logging in users, or restricting content
- A pluggable backend system [7]

Django provides a robust user authentication system, with support for user accounts, user groups, permissions, and cookie-based user session processing.

In Django, an authentication backend is a Python class that provides a way to authenticate users. The default authentication backend uses usernames and passwords to authenticate users, but you can create your own backend to implement alternative authentication methods [8].

Authentication backends work by intercepting the login process and checking the user's credentials against some external source. This could be a database, an LDAP server, or any other authentication provider. If the credentials are valid, the backend returns a user object that Django can use to authenticate the user and grant access to protected resources [8].

Description of the idea of a new method

Versatile, scalable Django web-framework, two-factor authentication system with the first part of the authentication methods of Password and second part of OTP via email, or special test. The test is a fast-paced test after login, which users can set manually, and if user fails to pass this test, authentication will be cancelled. Time for salvation of the can be set by user, it will be muscle memory case, which might be not comfortable at first, but after some training, it will be an easy additional step, which can even be set as a reaction test game, if user wants so, which in some cases, can be considered as an advantage.

Differences of the new method from the existing

The main difference is: After-login authentication is can be modified by users. When an unsuccessful attempt of additional testing occurs, the actual user receives an email notification, after three attempts by users after authorization in the IP system from which the device authenticates, it is blocked for a minute, and each subsequent unsuccessful attempt will increase

the delay. It is important to note that delays and notifications will not occur if the first part of password authentication fails. You can disable the second authentication factor.

IV. Conclusion

Main results and conclusions of the study

A novel user authentication concept was created throughout the research and has to be further developed. The project highlights the benefits of utilizing Django for user authentication, including automatic authentication processing and session management that guarantee the system's dependability and scalability.

The Django web framework-based universal and scalable solution with two-factor authentication was put forth. The approach consists of two parts: the first portion relies on password authentication, and the second part uses one-time passwords (OTPs), which can be delivered via email or through a unique test. This exam can be manually configured by users and is an accelerated test. The user's authentication will be revoked if they fail the test.

The study's findings demonstrate that the created method enables getting around the drawbacks of current authentication techniques and ensuring the security, adaptability, and usability of web resources. The suggested method has the potential to be improved upon and used in numerous high-class online applications.

Prospects for the development and implementation of the new authentication method

There are opportunities for further development and integration of the new authentication method into different web services, for now, it will be widely used Django-based web-sites. In addition to greater security provided by two-factor authentication, it also offers end users flexibility and customizability. The new method may become more functional and the level of user data protection may rise with further development.

Wherever a higher level of security and user data protection is needed, the proposed authentication mechanism can be successfully used. It is particularly helpful in systems where maintaining information security is important or where illegal access to an online resource could have negative effects. The application of this technique enhances user experience and builds system trust.

Overall, the established authentication mechanism is a noteworthy addition to the field of information security and a potential approach for guaranteeing the security of web resources.

References:

1. Bucko, A.; Vishi, K.; Krasniqi, B.; Rexha, B. Enhancing JWT Authentication and Authorization in Web Applications Based on User Behavior History. Computers 2023, 12, 78. https://doi.org/10.3390/computers12040078.
2. S. Vaithyasubramanian, A. Christy, D. Lalitha, Two Factor Authentication for Secured Login Using Array Password Engender by Petri net, Procedia Computer Science, Volume 48, 2015, Pages 313-318, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.04.187.
3. Ortiz, B.L.; Gupta, V.; Chong, J.W.; Jung, K.; Dallas, T. User Authentication Recognition Process Using Long Short-Term Memory Model. Multimodal Technol. Interact. 2022, 6, 107. https://doi.org/10.3390/mti6120107.
4. Carrillo-Torres, D.; Pérez-Díaz, J.A.; Cantoral-Ceballos, J.A.; Vargas-Rosales, C. A Novel Multi-Factor Authentication Algorithm Based on Image Recognition and User Established Relations. Appl. Sci. 2023, 13, 1374. https://doi.org/10.3390/app13031374.
5. R. F. Olanrewaju, B. U. I. Khan, M. A. Morshidi, F. Anwar and M. L. B. M. Kiah, "A Frictionless and Secure User Authentication in Web-Based Premium Applications," in IEEE Access, vol. 9, pp. 129240-129255, 2021, doi: 10.1109/ACCESS.2021.3110310.

6. Yu, L.; He, M.; Liang, H.; Xiong, L.; Liu, Y. A Blockchain-Based Authentication and Authorization Scheme for Distributed Mobile Cloud Computing Services. Sensors 2023, 23, 1264. https://doi.org/10.3390/s23031264.
7. URL: https://docs.djangoproject.com/en/3.2/topics/auth/ (date of application: 21.06.2023).
8. URL: https://reintech.io/blog/writing-custom-authentication-backend-django (date of application: 21.06.2023).

ЖАҢА ХЕШТЕУ АЛГОРИТМІНІҢ СТАТИТИКАЛЫҚ ҚАУІПСІЗДІК ҚАСИЕТТЕРІН ЗЕРТТЕУ

Сақан Қ.С., Хомпыш А.
e-mail: kairat_sks@mail.ru, ardabek@mail.ru
ҚР ҒжЖБМ «Ақпараттық және есептеуіш технологиялар институты»

Аңдатпа. *Мақалада ақпараттық қауіпсіздік зертханасы ұсынған HBC-256 хештеу алгоритмінің статистикалық қауіпсіздігі NIST статистикалық сынақтар жиынтығы арқылы бағаланады. Сонымен бірге, осы сынақтарға жекелей тоқтала кетіп, қысқаша мағлұмат беріледі. Жүргізілген эксперименталды жұмыстар үшін 19 200 000 байт болатын файл *.zip форматындағы кездейсоқ архив файл алынды. Алынған нәтижелер әр сынақ бойынша бағаланып, кіріс ақпарат сынақтан өткен-өтпегендігін анықтайды. Зерттеу барысында алынған көрсеткіштерге сүйене отырып, есептелген хеш-мәндер тізбегі жалғанкездейсоқтыққа өте жақын екендігін расталды.*

Кілт сөздер: *хеш-функция, шифрлау алгоритмі, статистикалық сынақтар.*

Кіріспе.

Жаңа құрылған хештеу алгоритмінің оның тәжірибедегі тиімділігімен қоса қауіпсіздігі қасиеттерін жан-жақты зерттеу жұмыстарын жүргізу міндетті талаптардың бірі болып саналады. Талдау жүргізудің жан-жақты және объективті болуы ғылыми тұрғыдан негізделіп, сенімділіктің нақты бағалануына үлкен назар аударған жөн. Қауіпсіздік пен тиімділікті бағалау барысында заманауи ақпараттық технологиялардың жетістіктері мен мүмкіндіктері, сондай-ақ ең кемінде, олардың алдағы жақын қысқа мерзімдегі даму қарқынын ескеру маңызды [1].

Хеш функциялар криптография және цифрлық қолтаңбалар қосымшаларында дұрыс және қауіпсіз жұмыс істейтініне көз жеткізу үшін оның сапасын бағалау – міндетті атқарылатын зерттеу жұмыстардың бірі болып саналады [2]. Хеш функцияның сапасын бағалау критерийлері ретінде оған қойылатын жоғарыда айтылған талаптар мен қасиеттерін қарастыруға болады, атап айтқанда [3-4]:

Коллизияға төзімділік: жақсы хеш функция бірдей хеш-мән беретін екі кірісті табуды қиындатуы керек.

Бірінші және екінші прототипке төзімділік: жақсы хеш функциясы көрсетілген хеш нәтижесін беретін кірістерді табуды қиындатуы керек.

Лавиндік эффекті: хабарламаның шамалы өзгеруі хеш-мәннің айтарлықтай өзгеруіне әкелуі керек.

Диффузия: хеш функция хабарламаны шығыс биттеріне біркелкі таратуы керек.

Тиімділік: жақсы хеш функциясы есептеуде жылдам болуы керек және аз операция орындалуы қажет етеді.

Әмбебаптық: жақсы хеш функциясы кірістер мен кірістердің кең ауқымы үшін жақсы жұмыс істеуі керек.

Анықталғандық: әрқашан бірдей кіріс үшін хеш функциясының шығысы бірдей болуы керек.

Хеш-мәннің ұзындығы: шығыс ұзындығы қолданбаға сәйкес келуі керек және коллизия шабуылдарының алдын-алу үшін жеткілікті үлкен болуы керек.

Криптографиялық қауіпсіздік қасиеттері: криптографиялық қауіпсіздікті қажет ететін қосымшалар үшін хеш функция берік математикалық негізге ие болуы немесе дәлелденген криптографиялық примитивтерден жасалуы керек және мұқият талдау мен тестілеу арқылы қауіпсіз түрде дәлелденуі керек.

Ашықтық: хеш функциясында шабуылдаушы қолдана алатын жасырын функциялар немесе әлсіз жерлер болмауы керек.

Криптографиялық және статистикалық талдауларға негізделген (дифференциалды, сызықтық, алгебралық және т.б.) шабуылдарға төзімділік: жақсы хеш функциясы аталған шабуылдарға осал болмауы керек, алгоритм сұлбасы негізінде шабуылдаушы кірістер (хабарламалар, кілттер, инициализациялық векторлар) мен шығыстар (хеш-мәндер) арасындағы айырмашылықтарды талдай отырып, кіріс туралы ақпаратты ала алмауы керек.

Бұл критерийлер хеш мүмкіндіктерінің қауіпсіз және нақты қолданбаларда пайдалануға жарамды екеніне көз жеткізуге көмектеседі.

Хеш функциясына шабуыл оның сапасы мен қауіпсіздік қасиеттерін бағалап қана қоймай, оны одан әрі жетілдірудің бірден-бір жолы ретінде қарастыруға болады. Хеш функциясына шабуыл – бұл хеш функцияларының қауіпсіздік қасиеттері мен талаптарының бірінің немесе бірнешеуінің дұрыс орындалмауын дәлелдеуге бағытталған әс-әрекеттер.

HBC-256 хештеу алгоритмінің статистикалық қауіпсіздік қасиеттерін зерттеу

Құрылған HBC-256 хештеу алгоритмін қауіпсіздігі қасиеттерін зерттеу барысында алынған хеш-мәндерді танымал статистикалық сынақтар арқылы псевдакездейсоқ тізбектердің сипаттарына ие бола алу мүмкіндігіне тексереміз. Аталған хештеу алгоритмінің сипаттамалары мен оның қауіпсіздік қасиеттеріне қатысты жүргізілген басқа да ғылыми зерттеулерді [5] мақаладан танысуға болады. Зерттеудің осы бағытында біз неғұрлым көп сынақ жұмыстарын жүргізетін болсақ, соғұрым дәлдікке жақын нәтиже алатын боламыз. Талдау барысында жасалған хештеу алгоритмі арқылы алынған хеш-мәндердің үлкен көлемдегі жиынын қарастыратын боламыз. Егер жиындағы хеш-мәндер алдын-ала болжамды болса, онда бұл қолданылатын ең күшті хештеу алгоритмінің өзі де осал болып шығады – шабуылдаушы ықтимал хеш-мәндер кеңістігі тарыла түскендігін пайдаланып, оларды теру (немесе «сұрыптау») арқылы кейбір ақпаратты, дәлірек айтқанда, шабуылдаушы үшін коллизияны тауып ала алу мүмкіндігі арта түседі. Сол себепті, есептелінетін хеш-мән псевдокездейсоқ тізбекке неғұрым жақын болғаны дұрыс. NIST ұсынатын сынақтардың әрқайсысы кіріс тізбек ретінде хеш-мәндер жиынын қабылдайды. Әрі қарай, осы тізбектің белгілі бір қасиетін сипаттайтын статистика есептеледі – бұл бір мән немесе мәндер жиынтығы болуы мүмкін. Осыдан кейін бұл статистика эталондық статистикамен немесе математикалық тәсілдермен алынған «нөлдік гипотезамен» салыстырылады. Бәрімізге белгілі – іс жүзіндегі статистика біз қанша мықты, идеалды алгоритм пайдалансақ та, ол мән еш уақытта эталондық мәнмен 100% үйлесе бермейді. Сондықтан, 5% ауытқуды шекара ретінде аламыз. Егер іс жүзіндегі статистика мәні осы шекарадан аспаса, тәжірибе сынақтан өтті деп ұйғаратын боламыз [6].

Кез келген жаңа криптографиялық құрылым сияқты, HBC-256 алгоритмі де оның криптографиялық қасиеттерін, атап айтқанда: қайтымсыздығы мен бірінші және екінші тектегі коллизияларға төзімділігін растау үшін мұқият зерттеуді қажет етеді.

Хеш-мәндердің статистикалық қасиеттерін бағалау

Кез-келген еркін ұзындықтағы *M* хабарламасы үшін жасалған HBC-256 хештеу алгоритмінің *h(M)* хеш-мәні жалған кездейсоқ тізбектердің қасиеттерін қанағаттандыруы міндетті шарттардың бірі болып саналады. Бұл хештеу алгоритмдеріне қойылатын негізгі талаптардың бірі, яғни хеш функциясына негізделген жалғанкездейсоқ сандар генераторын кездейсоқ сандар генераторынан ажырату қиын болуы керек.

NIST статистикалық сынақтар жиынтығы арқылы бағалау.

Неғұрлым сенімді статистикалық бағалау үшін Ұлттық стандарттар және технологиялар институтының (бұдан әрі – NIST) статистикалық сынақтар жиынтығын қолдана отырып, HBC-256 хеш функциясының кездейсоқтық қасиетін бағалау қажет. NIST ұсынған статистикалық сынақтар жиынтығы құрамына 15 статистикалық тест кіреді, бірақ осы 15 сынақтың бірнешеуі кіріс параметрлерге (күйлерге) байланысты бірнеше сынақтардың жиынтығы болып табылады, сондықтан іс жүзінде жүргізілетін сынақтар саны 15-тен әлдеқайда көп. Сынақтардың мақсаты берілген тізбектің шынайы кездейсоқ тізбектерден ауытқушылық өлшемін анықтау болып табылады. Бұл сынақтар шынайы кездейсоқ тізбектерге тән әртүрлі статистикалық қасиеттерге негізделген. Сынақтар екілік санау жүйесіндегі тізбектерге (екілік тізбектерге) жүргізіледі [7].

Әр сынақ кездейсоқтықты белгілі бір критерийлер бойынша бағалап, сәйкесінше, функцияны *p*-мәндері бойынша бағалады. *p*-мәні – бұл есептелген тексеру статистикасы H_0 нөлдік гипотезадан бас тартуға әкелетін маңыздылық деңгейінің ең кіші мәні (яғни ақиқат гипотезадан бас тарту ықтималдығы). Біздің жағдайда, егер сынақ жүргізу нәтижесінде есептелген $p > \alpha=0,01$ мәніне ие болса, бұл зерттелетін екілік тізбек 99% сенімділікпен кездейсоқ болады дегенді білдіреді. Мұндағы $p \in[0,1]$, α – маңыздылық деңгейі, яғни H_0 нөлдік гипотезаны қабылдамау ықтималдығы [54].

HBC-256 алгоритмінің хеш-мәндердің кездейсоқтығын талдау үшін келесі сынақтар жүргізілді:

1) Жиілік (монобиттік) сынақ: Бұл сынақтың мақсаты – екілік тізбектегі нөлдер мен бірліктердің ара-қатынасын өлшеу. Өту критерийлері бірлік пен нөлдердің сандарының (жиілігінің) жақындығымен анықталады, олар шамамен бір-біріне жақын болуы керек.

2) Блок ішіндегі жиіліктер сынағы: бұл сынақ жиілік сынағының жалғасы болып табылады, онда бірлік пен нөлдің арақатынасы *M*-биттік блоктың өлшемін қолдану арқылы анықталады. Сынақтан өту критерийі болып *M*-биттік блоктағы бірліктер мен нөлдердің жиілігі *M*/2-ге жақын болуы саналады.

3) Бірдей биттердің тізбектері сынағы: бұл сынақтың мақсаты – әртүрлі ұзындықтағы тек бірліктер немесе тек нөлдерден тұратын тізбекшелердің саны кездейсоқ тізбектегі санына сәйкес келетіндігі туралы қорытынды жасау. Атап айтқанда, қарастырылатын тізбектегі тек бірліктер мен тек нөлдер тізбекшелері өзара жиі немесе сирек ауысып отыратынын тексереді.

4) Блоктағы бірліктердің ең ұзын тізбекшесі сынағы: бұл сынақ «Бірдей биттердің тізбегі сынағының» жалғасы болып табылады. Сынақ мақсаты – мұндай бірлік тізбекшенің ұзындығы қарастырылған екілік тізбек ұзындығындай кездейсоқ екілік тізбектегі бірліктердің ең ұзын бірлік тізбегінің ұзындығына сәйкес келетіндігін-келмейтіндігін анықтау.

5) Екілік матрицаның рангін анықтау сынағы: Мұнда қарастырылған екілік тізбектен алынған қиылыспайтын ішкі екілік матрицалардың рангтері есептеледі. Бұл сынақтың мақсаты – алғашқы тізбекті құрайтын тұрақты ұзындықтағы ішкі тізбекшелердің өзара сызықтық тәуелділігін тексеру.

6) Дискретті Фурье түрлендіру сынағы (спектрлік сынақ): Бұл сынақтың мақсаты – алғашқы екілік тізбектің периодтық қасиеттерін анықтау, мысалы, бір-біріне жақын орналасқан қайталанатын үзінділерді анықтау.

7) Қабаттаспайтын үлгілерге сәйкестік сынағы: Бұл сынақ алғашқы тізбекте табылған алдын-ала енгізілген үлгілердің негізінде нөлдер мен бірліктердің бірқалыпты үлестірілуін бағалайды.

8) Қабаттасатын үлгілерге сәйкестік сынағы: Бұл бір-біріне сәйкес келмейтін үлгіні тексеруге ұқсас, тек айырмашылығы - үлгіні тапқан кезде, тор келесі іздеуді жалғастырмас бұрын бір битке ауысады.

9) Маурердің «әмбебап статистикалық» сынағы: Мұнда алғашқы екілік тізбектегі бірдей үлгілер арасындағы биттер саны анықталады. Соның негізінде тізбектің айтарлықтай қысылғанын немесе қысылмағанын тексереді. Айтарлықтай қысылған реттілік кездейсоқ емес деп анықталады.

10) Сызықтық күрделілік сынағы: бұл сынақ кездейсоқтықты тексеру үшін берілген екілік тізбектің сызықтық күрделілігін бағалайды. Сынақ кері байланысты жылжудың сызықтық регистрінің жұмыс істеу принципі негізінде жүргізіледі.

11) Сериялық сынақ: бұл сынақ берілген екілік тізбектегі барлық мүмкін болатын қабаттасатын үлгілерді анықтайды. Сынақ мақсаты – берілген биттік тізбектегі *M*-биттік ұзындықтағы қабаттасатын 2^M үлгілерінің пайда болу саны шынымен кездейсоқ биттік тізбектегімен бірдей ме, жоқ па, соны анықтау.

12) Жуықтау энтропиялық сынақ: Алдыңғы сериялық сынақ сияқты бұл сынақ алғашқы биттік тізбек бойынша *M*-биттік ұзындықтағы үлгілердің барлық ықтимал қабаттасу жиілігін санауға бағытталған. Сынақтың мақсаты – алғашқы биттік тізбектің ұзындығы *M* және *M+1* бит болатын екі тізбектескен блоктарының қабаттасу жиіліктерін кездейсоқ биттік тізбектің сынақ аналогіндегі ұқсас блоктардың қабаттасу жиіліктерімен салыстыру.

13) Жиынтық қосынды сынағы (Cusums сынағы): бұл сынақ кездейсоқтықты тексеру кезінде ішінара тізбекшелердің жиынтық қосындысы тым аз немесе тым үлкен екенін анықтайды, нәтижені кездейсоқ екілік тізбектегі ішінара тізбекшелердің жиынтық қосындысымен салыстырады. Мұны тікелей сынақ немесе сынақ арқылы бағалауға болады.

14) Еркін ауытқулар сынағы: Бұл сынақтың мақсаты – кездейсоқтықты өлшеу үшін қандай да бір ауытқу орын алатын немесе болмайтын циклдегі белгілі бір күйге бару санын анықтау. Іс жүзінде бұл сынақ циклдың сегіз (-4, -3, -2, -1, +1, +2, +3, +4) күйінің әрқайсысы үшін жүргізілген сегіз тәуелсіз сынақтан тұрады.

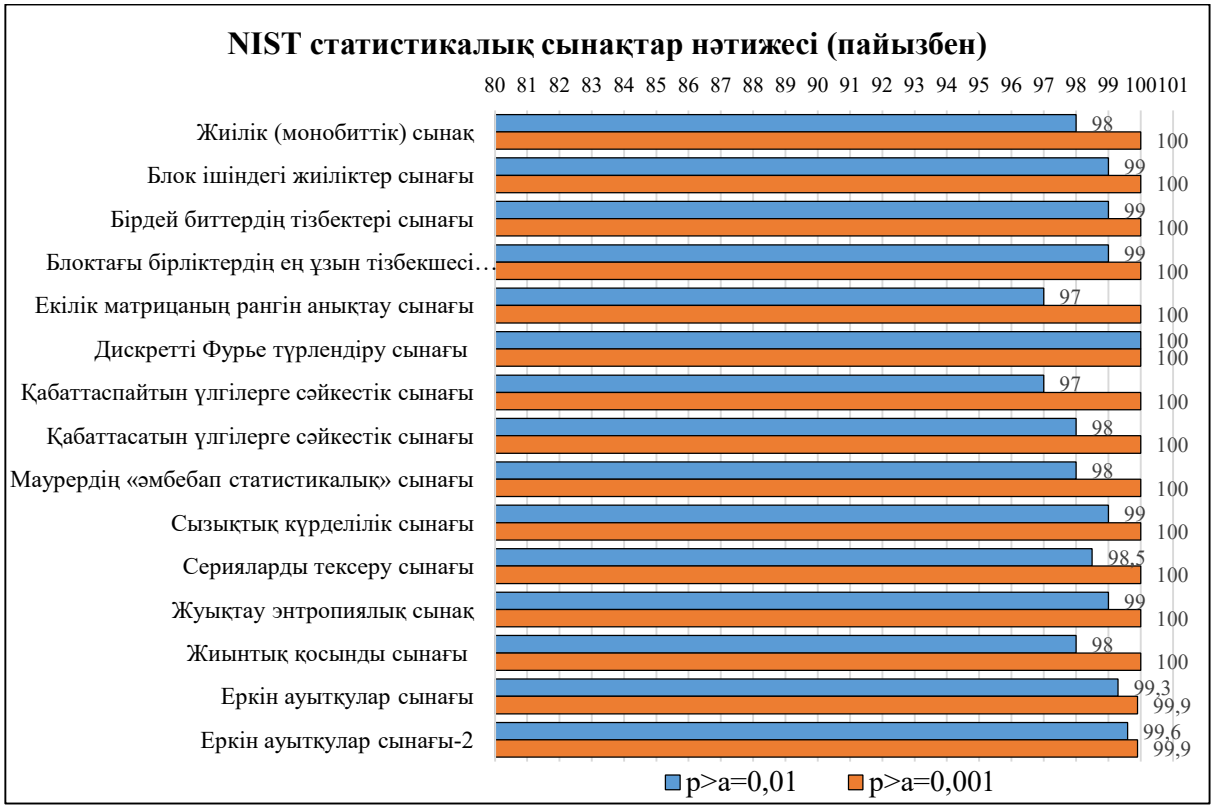
15) Еркін ауытқулар сынағы-2: Алдыңғы сынаққа ұқсас. Ерекшелігі – әртүрлі күйлерге барудың күтілетін санынан ауытқуларды ерікті түрде айналып өту кезіндегі жалпы санын анықтау. Сынақ әр жағдайға арналған 18 сынақтан тұрады: -9, -8, ..., -1, +1, +2, ... , +9.

Талдау үшін NIST ұсынған статистикалық сынақтардың жиынтығы ресми дереккөзден алынған бағдарламалық жасақтама пайдаланылды.

Ұсынылған хештеу алгоритмін NIST ұсынған статистикалық сынақтардан өткізу төмендегідей жүргізілді. Тестілеу үшін көлемі 19 200 000 Б болатын файл *.zip форматындағы кездейсоқ архив файл алынды. Бұл форматтағы файл таңдау себебі, файл мазмұнында қайталаулар санын минимизициялау болды. Алгоритмнің сипаттамасына сәйкес файлдағы кезекті әрбір 48 байт ақпарат хештеу алгоритміне хабарлама ретінде енгізіліп, нәтижесінде соған сәйкес 32 байттағы хеш-мән болып белгілі бір файлға тізбектеліп жазылып отырады, яғни 400 мың хабарламаға сәйкес 400 мың хеш-мән

жазылған. Шығыс нәтижелер – хеш-мәндер жазылған *.hash форматындағы файлдың көлемі 12 800 000 байтты құрады. Шығыс файлды әрқайсысы 12500 байт болатын 100 файлға (тізбекке) бөлінді. Әрі қарай, алынған файлдарға NIST статистикалық сынақтар жиынтығы қолданылды.

Төменгі Сурет-1-де NIST статистикалық сынақтар жиынтығы арқылы құрылған хештеу алгоритмімен алынған хеш-мәндер тізбегінің псевдокездейсоқтыққа қатысты талдау көрсетілген. Көрсетілген суретте α – маңыздылық деңгейінің екі жағдайы қарастырылған және пайыздық көрсеткішпен сынақтан өту дәрежесі бейнеленген.



Сурет 1 –NIST статистикалық сынақтар нәтижелері

Алынған нәтиже бойынша барлық сынақтар сәтті өтті. NIST ұсыныстары бойынша, егер сынақтан өткен тізбектердің ең аз саны 100 тізбектің 96 тізбегі болса, сынақ сәтті өтті деп саналады. Сынақ нәтижелері HBC-256 хештеу алгоритмі арқылы алынған тізбектерде ауытқулардың жоқ екендігін көрсетті. Осылайша, алгоритм статистикалық қауіпсіздіктің жоғары деңгейіне ие.

ҚОРЫТЫНДЫ

Жұмыста блоктық шифрлау алгоритмі негізінде қауіпсіздігі қасиеттері мен өнімділігі жағынан жоғары, бағдарламалы-аппараттық жүзеге асыруға және параллельдік есептеуге икемделген жаңа хештеу алгоритмі HBC-256 хештеу алгритміне NIST статистикалық сынақтар жиынтығы арқылы зерттеу жұмыстары жүргізілді. нәтижелерге сүйене отырып, алынған хеш-мәндер тізбегі жалғанкездейсоққа өте жақын екендігін расталды.

Алғыс

Жұмыс AP14870719 «Хеш-функцияларға негізделген пост-кванттық криптографиялық алгоритмдерді жасау және зерттеу» гранттық қаржыландыру ғылыми жобасы аясында жүргізілді.

Пайдаланылған әдебиеттер тізімі

1. Rajeev S., Geetha G. Cryptographic Hash Functions: A Review. International Journal of Computer Science Issues // ISSN (Online):1694-0814. – 2012. – Vol. 9. – P. 461-479.
2. Preneel B., Govaerts R., Vandewalle J. Hash Functions Based on Block Ciphers: A Synthetic Approach // Crypto'93. – Vol. 773 of LNCS. – P. 368-378. Springer-Verlag, 1993.
3. Al-Kuwari, S., Davenport, J., Bradford, R. Cryptographic Hash Functions: Recent Design Trends and Security Notions // IACR. https://eprint.iacr.org/2011/565.pdf 12.10.2021.
4. Andrew R., Juan S. A statistical Test Suite foe Random and Pseudo Random Number Generators for Cryptographic Applications // https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-22r1a.pdf
5. Sakan K.S., Nyssanbayeva S.Y., Kapalova N.A., Algazy K.T., Khompysh A., Dyusenbayev D.S. Development and analysis of the new hashing algorithm based on block cipher // Eastern-European Journal of Enterprise Technologies. – 2022. – № 2/9 (116). – P. 60-73. doi:10.15587/1729-4061.2022.252060.
6. Nysanbayeva S.E., Kapalova N.A., Dyusenbayev D.S., Algazy K.T., Sakan K.S. Investigation of the stastical security of a pseudo-random sequence generator // International Conference “Computer Data Analysis and Modeling: Stochastics & Data Science” (CDAM-2022). – Minsk. September 06-10, 2022. – P. 137-144.
7. Bassham L., Rukhin A., Soto J., Nechvatal J., Smid M., Leigh S., Levenson M., Vangel M., Heckert N., Banks D. A Statistical test suite for random and pseudorandom number generators for cryptographic applications, special publication (NIST SP), National institute of standards and technology, Gaithersburg, MD, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=906762 23.01.2023.

ИССЛЕДОВАНИЕ И РЕАЛИЗАЦИЯ ПРОТОКОЛА КВАНТОВОГО РАСПРЕДЕЛЕНИЯ КЛЮЧЕЙ E91

Е.Е. Бегимбаева[1,2], Т.М. Жаксалыков[1,2*], М.Т. Аршидинова[1,3]
e-mail: enlik_89@mail.ru, zhaxalykov8@gmail.com
[1]*Казахский Национальный Исследовательский Технический Университет имени К.И. Сатпаева, Казахстан*
[2] *Казахстанско - Британский Технический Университет, Казахстан*
[3] *Институт Информационных и Вычислительных Технологий КН МНВО РК, Казахстан*

Аннотация. *В статье приведена реализация квантового протокола распределения ключей E91, который основан на фундаментальных принципах квантовой механики. Данный протокол выделяется как многообещающий метод обеспечения информационной безопасности. Протокол E91, используя свойства квантовой неделимости и индетерминизма, обладает потенциалом обеспечить надежную и гарантированную передачу данных. Для реализации протокола E91 было использовано математическое и компьютерное моделирование. Был проведен обзор литературы и анализ предыдущих исследований в области квантовой криптографии и методов распределения ключевой информации. Данное исследование может способствовать усовершенствованию протокола E91 и играет важную роль в обеспечении информационной безопасности в современных коммуникационных системах.*

Введение

Квантовая криптография представляет собой важную сферу информационной безопасности, применяющую основы квантовой механики. Она обеспечивает надежную защиту от атак на шифрование и эффективно служит средством обеспечения безопасности данных в процессе их передачи через сети связи и вычислительные системы [1]. Среди разнообразных протоколов квантового распределения ключей, протокол E91 выделяется как один из самых перспективных и вдохновляющих. Протокол E91, представленный Артуром Экертом в 1991 году, представляет инновационный способ обеспечения безопасного обмена ключами, используя принципы квантовой запутанности и измерения. Существенной характеристикой данного протокола является его способность эффективно сопротивляться атакам и обеспечивать высокий уровень безопасности [2].

В данной статье рассматривается исследование, посвященное реализации протокола квантового распределения ключей E91. Основной целью данного исследования является программная реализация протокола квантового распределения ключа E91. Ожидается, что результаты данного исследования способствуют дальнейшему совершенствованию протокола E91 и будут представлять новое понимание для применения квантовой криптографии в реальных информационных системах.

1. Исследование и реализация протокола E91

В 1991 году математик Артур Экерт представил метод генерации секретных кодов, используя принципы квантовой запутанности, что привело к появлению протокола E91. Исследование Экерта уделяет особое внимание процессу формирования "ловушек" состояний квантовой природы, которые играют ключевую роль в протоколе [2]. Протокол начинается с того, что внешний участник, именуемый Чарли, генерирует специальное квантовое состояние "ловушки". После этого Чарли отправляет одну базовую единицу квантовой информации Алисе и Бобу, используя специализированные каналы связи, разработанные для квантовых вычислений. Алиса и Боб, получив квантовые биты информации, производят предсказание ориентации кубитов - квантовых битов. Затем, как и в предыдущих этапах протокола, участники сравнивают измерения, которые были проведены в выбранном направлении, посредством открытых каналов связи. Путем анализа запутанных результатов, Алиса и Боб получают секретный пароль.

В начальной фазе процесса, Алиса и Боб получают квантовые состояния в форме синглетов $|\psi_S\rangle$, которые были сгенерированы Чарли [3]. Для моделирования, требуются регистры, включающие два квантовых бита и четыре классических бита (Рисунок 1).

```
# Creating registers
qr = QuantumRegister(2, name="qr")
cr = ClassicalRegister(4, name="cr")
```

Рис. 1 Инициализация регистров

Предположим, что кубиты с метками qr[0] и qr[1] принадлежат Алисе и Бобу соответственно. Формирование Чарли синглетного состояния (Рисунок 2):

```
singlet = QuantumCircuit(qr, cr, name='singlet')
singlet.x(qr[0])
singlet.x(qr[1])
singlet.h(qr[0])
singlet.cx(qr[0],qr[1])
```

Рис. 2. Инициализация синглетов

Два кубита qr[0] и qr[1] теперь связаны вместе и влияют друг на друга. Чарли передает одну часть информации Алисе, а другую — Бобу.

Второй этап процедура измерения начинается с подготовки измерительных операций, которые будут применяться Алисой и Бобом. Установим, что $A(\vec{a}_i)= \vec{a}_i\cdot\vec{\sigma}$ и $B(\vec{b}_j)= \vec{b}_j\cdot\vec{\sigma}$ представляет собой обсервируемые спиновые проекции, используемые для их измерений. Для выполнения этих операций стандартный базис Z может быть повернут в нужное положение по мере необходимости. На рисунке 3 представлено обозначение для возможных измерений Алисы и Боба. Блоки, расположенные слева, могут быть рассмотрены как детекторы, применяемые участниками для измерения наблюдаемых величин X, W, Z и V.

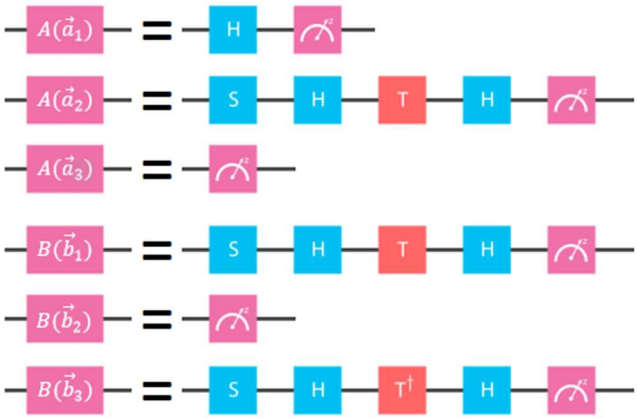


Рис. 3 Измерения Алисы и Боба

Допустим, Алиса и Боб стремятся сгенерировать конфиденциальный ключ, используя N синглетных состояний, созданных Чарли (Рисунок 4).

```
# Define the number of singlets N
numberOfSinglets = 500
```

Рис. 4 Определение количества синглетов

Участники обязаны определить направления, в которых они собираются проводить измерения спиновых проекций своих кубитов. Для этой цели Алиса и Боб формируют последовательности b и b', содержащие случайно сгенерированные элементы (Рисунок 5). На рисунке 6 представлен результат объединение всех цепей.

```
aliceMeasurementChoices = [random.randint(1, 3) for i in range(numberOfSinglets)] # string b of Alice
bobMeasurementChoices = [random.randint(1, 3) for i in range(numberOfSinglets)] # string b' of Bob
```

Рис. 5 Генерация строк Алисы и Боба

```
circuits = [] # the list in which the created circuits will be stored

for i in range(numberOfSinglets):
    # create the name of the i-th circuit depending on Alice's and Bob's measurement choices
    circuitName = str(i) + ':A' + str(aliceMeasurementChoices[i]) + '_B' + str(bobMeasurementChoices[i])

    # create the joint measurement circuit
    # add Alice's and Bob's measurement circuits to the singlet state curcuit
    # singlet state circuit # measurement circuit of Alice # measurement circuit of Bob
    circuitName = singlet + aliceMeasurements[aliceMeasurementChoices[i]-1] +
    bobMeasurements[bobMeasurementChoices[i]-1]

    # add the created circuit to the circuits list
    circuits.append(circuitName)
```

Рис. 6 Объединение всех цепей

Суть заключается в том, чтобы эмулировать каждый этап формирования синглетных состояний, распределения кубитов между участниками и измерения спиновой проекции в выбранном направлении протокола E91. Это достигается путем одновременного выполнения каждой из указанных схем в одном цикле [5].

На третьем этапе фиксация результатов, проведено выполнение схем на виртуальной модели (Рисунок 7).

```
backend=BasicAer.get_backend('qasm_simulator')
result = execute(circuits, backend=backend, shots=1).result()
#print(result) # uncomment for detailed result
```

Рис. 7 Подключение симулятора

Рассмотрим результат выполнения первой схемы, представленный четырьмя цифрами. Важно отметить, что Алиса и Боб регистрируют результаты измерений в классических битах cr[0] и cr[1] (последние две цифры). В симуляции процесса генерации секретного ключа без вмешательства стороннего наблюдателя, классические биты cr[2] и cr[3] всегда остаются нулевыми. Также стоит обратить внимание, что на выходе формируется словарь, где ключами служат полученные результаты, а значения представляют их количество.

Четвертый этап: обнаружение совпадений. На предыдущем этапе, мы сохраняем результаты измерений Алисы и Боба в списках "aliceResults" и "bobResults" соответственно (обозначенных как "a" и "a'). В данный момент участники проводят сопоставление своих последовательностей "b" и "b'" через доступный классический канал. Если Алиса и Боб проводят измерения спиновых проекций своих кубитов для i-го синглета в одном и том же направлении, то Алиса записывает результат "a_i" как бит строки "k", а Боб записывает результат "-a_i" как бит строки "k'" (Рисунок 8).

```
ceKey = [] # Alice's key string k
Key = [] # Bob's key string k'

omparing the stings with measurement choices
 i in range(numberOfSinglets):
 # if Alice and Bob have measured the spin projections onto the a_2/b_1 or a_3/b_2 directions
 if (aliceMeasurementChoices[i] == 2 and bobMeasurementChoices[i] == 1) or
 (aliceMeasurementChoices[i] == 3 and bobMeasurementChoices[i] == 2):
     aliceKey.append(aliceResults[i]) # record the i-th result obtained by Alice as the bit of the secret key k
     bobKey.append(- bobResults[i]) # record the multiplied by -1 i-th result obtained Bob as the bit of the secret
     #key k'

Length = len(aliceKey) # length of the secret key
```

Рис. 8 Получение ключа

Теперь значения ключей, "k" и "k'", сохраняются в списках "aliceKey" и "bobKey" соответственно.

Поскольку строки "k" и "k'" остаются секретными, Алиса и Боб не обладают информацией о различиях в битах своих ключей. Для определения числа ошибок участники могут провести тест на случайной выборке. Алиса случайным образом выбирает δ битов из своего секретного ключа и передает информацию о выбранных битах Бобу. После этого Алиса и Боб сравнивают значения этих контрольных битов. При достаточно большом значении δ количество ошибок в контрольных битах будет тесно соответствовать количеству ошибок в остальных битах [6].

2. Применение теста значения корреляции CHSH

Для оценки надежности и безопасности протокола, Алиса и Боб рассчитывают корреляционный тест CHSH - показатель, который характеризует отношение измерений. Если значение дисбаланса CHSH сильно отклоняется от значения -2√2, это указывает на

нарушение в расположении состояний, созданных Чарли. В случае обнаружения подобных отклонений, сеанс связи прерывается для последующего анализа и выяснения причин ошибок [2].

Алиса и Боб стремятся обеспечить отсутствие помех в процессе связи. Для этой цели они вычисляют значение корреляции CHSH, используя результаты, полученные в результате измерения спиновых проекций на направления a_1/b_1, a_1/b_3, a_3/b_1 и a_3/b_3. Важно подчеркнуть, что это аналогично измерению наблюдаемых $X \otimes W$, $X \otimes V$, $Z \otimes W$ и $Z \otimes V$ соответственно [7]. Так как Алиса и Боб раскрыли свои последовательности "b" и "b'", им известно, какие измерения они выполнили и какие результаты они получили. Используя эту информацию, участники вычисляют предполагаемые значения.

3. Результаты

На рисунке 9 приведены все значимые результаты.

```
corr = chsh_corr(result) # CHSH correlation value

# CHSH inequality test
print('CHSH correlation value: ' + str(round(corr, 3)))

# Keys
print('Length of the key: ' + str(keyLength))
print('Number of mismatching bits: ' + str(abKeyMismatches) + '\n')
```

```
CHSH correlation value: -2.612
Length of the key: 112
Number of mismatching bits: 0
```

Рис. 9 Получение результатов

В заключение, Алиса и Боб обладают конфиденциальными ключами "k" и "k'" (представленными как "aliceKey" и "bobKey"). Теперь они имеют возможность использовать метод одноразового блокнота для зашифровки и расшифровки сообщений.

Поскольку было проведено моделирование протокола E91 без участия стороннего лица Евы, значение корреляции CHSH должно стремиться к -2,828 [8]. Кроме того, ключи Алисы и Боба не должны содержать несовпадающих битов. Важно отметить, что имеется 9 потенциальных комбинаций измерений, которые Алиса и Боб могут провести, но только 2 из них позволяют получить результаты для создания секретных ключей [9,10]. Таким образом, соотношение длины ключей к количеству синглетных состояний "N" должно стремиться к значению 2/9.

4. Заключение

В данной статье приведены результаты исследования и реализации протокола квантового распределения ключей E91. Были рассмотрены механизмы создания синглетных состояний участниками протокола и проанализирован процесс измерения спиновых проекций в различных направлениях. В ходе исследования было обнаружены важные особенности, способствующие созданию безопасных ключей, а также выявить аспекты, которые требуют дополнительного внимания и усовершенствования. Также подчеркнули важность вычисления корреляции CHSH в контексте проверки наличия помех и степени безопасности протокола. Результаты подтверждают, что при отсутствии подслушивателя значение корреляции CHSH должно приближаться к значению -2,828, что является ключевым индикатором надежности протокола.

Данное исследование подготовлено в рамках проекта AP19675961 «Разработка и исследование протоколов распределения ключей на основе квантовых свойств»,

который реализуется в Казахском национальном исследовательском техническом университете имени К.И.Сатпаева.

Список литературы

1. Холево А. С. Квантовые системы, каналы и информация. — Москва: МЦНМО, 2010.
2. Холево А. С. Математические основы квантовой информатики. Лекционные курсы НОЦ. т. 30. — Москва: МИАН, 2018. — с. 3—117.
3. Ekert A. Quantum cryptography based on Bell's theorem. // Phys. Rev. Lett. — 1991. — т. 67. — с. 661.
4. Нильсен М., Чанг И. Квантовые вычисления и квантовая информация. — Москва: Мир, 2006.
5. P. Stavroulakis M. S. Handbook of Information and Communication Security. — Berlin, Heidelberg: Springer, 2010.
6. Кронберг Д. А., Ожигов Ю. И., Чернявский А. Ю. Квантовая криптография. Учебное пособие. — Москва: МГУ имени М. В. Ломоносова, факультет ВМК.
7. Huttner B., Imoto N., Gisin N. Quantum Cryptography: Public Key Distribution and Coin Tossing // Physical Review A. — 1995. — с. 11—15.
8. Scarani V., Acin A., Ribordy G. Quantum Cryptography Protocols Robust against Photon Number Splitting Attacks for Weak Laser Pulse Implementations // Phys. Rev. Lett. — 2004. — с. 92.
9. Lo H., Ma X., Chen K. Decoy state quantum key distribution. // Phys. Rev. Lett. — 2005. — с. 94.
10. Erven C. On Free Space Quantum Key Distribution and its Implementation with a Polarization-Entangled Parametric Down Conversion Source. // University of Waterloo. — Ontario, Canada, April 2007. — с. 15—17.

ӘЛЕУМЕТТІК ЖЕЛІЛЕРДЕГІ СУРЕТТЕР МЕН БЕЙНЕЛЕРДІ ӨҢДЕУДІҢ ЕРЕКШЕЛІКТЕРІ

Ш.Ж. Мусиралиева[1], К.Б. Багитова[1*],
И.А. Терейковский[2], А.М. Усманова[1]
**E-mail: kbbagitova@gmail.com*
[1]Әл – Фараби атындағы Қазақ ұлттық университеті, Қазақстан, Алматы
[2]Игорь Сикорский атындағы Киев политехникалық институты, Украина, Киев

Аңдатпа. *Соңғы жылдары әлеуметтік желілер зорлық-зомбылық мазмұнды контенттерді таратудың ұясына айналды. Facebook, Twitter және YouTube секілді платформалардың пайда болуымен қолданушылар күн сайын көптеген зорлық-зомбылыққа толы суреттер мен бейнелердің куәгерлері болып келеді. Бұл мәселемен күресу үшін зерттеушілер мен жасақтаушылар әлеуметтік желілердегі зорлық-зомбылық мазмұнды анықтау және белгілеу үшін нейрондық желіні талдауға жүгі-неді.*

Нейрондық желілер, әсіресе терең оқыту үлгілері, сурет пен бейнені талдауда жоғары тиімділік танытты. Зорлық-зомбылық және зорлық-зомбылықсыз мазмұн таңбаланған үлкен деректер жиынын оқыту арқылы бұл модельдер зорлық-зомбылыққа қатысты ерекшеліктер мен мүмкіндіктерді тануды үйренеді. Бұл оларға зорлық-зомбылық бейнелері мен бейнелерді жоғары дәлдікпен автоматты түрде анықтауға және жіктеуге мүмкіндік береді. Бұл зерттеуде біз зорлық-зомбылықты анықтау әдістерін зерттеп, талдаймыз және зорлық-зомбылықты анықтауды зерттеудегі жаңа үрдістерді болжаймыз.

Кілттік сөздер: *әлеуметтік желілер, саяси экстремизмді анықтау, суреттер, контрастты түзету, вейвлеттік талдау.*

Кіріспе. Көп жағдайда әлеуметтік желілердегі графикалық мазмұны монохромды немесе түрлі-түсті кескіндерден, сондай-ақ монохромды немесе түсті форматта ұсынылған моно-арналық бейнеден тұрады [1,2]. Статикалық бейнелер мен бейне ақпараттарды нейрондық желіні талдау саласындағы практикалық тәжірибе мен зерттеу нәтижелері [2,3 және 8] көрсеткендей, мұндай талдаудың тиімділігін арттырудың перспективалы жолдарының бірі алдын ала өңдеуді жүзеге асыру болып табылады. Ең алдымен, мұндай өңдеудің мақсаты кескіндер мен бейне ағындарының параметрлерін берілген типтегі нейрондық желі моделінің енгізу өрісінің параметрлеріне қойылатын талаптарға жеткізу. Осылайша, нейрондық желі моделінің кіріс өрісінің өлшемі мен талданатын кескіннің өлшемі арасындағы елеулі айырмашылық кескінді масштабтау үшін арнайы процедураны қолдануды талап етеді, өйткені тривиальды масштабтау процедурасы кескінді шамадан тыс бұрмалауы мүмкін. Бұрмалау әсіресе кескіннің мүмкін болатын пропорционалды емес масштабтауы жағдайында байқалады, бұл қажеттілік нейрондық желі моделінің кіріс өрісінің арақатынасы мен талданатын кескіннің арақатынасы арасындағы сәйкессіздік жағдайында туындауы мүмкін.

Сонымен қатар, кейбір желідегі әлеуметтік желілерде кескіндер мен бейнелер сығылған түрде орналастырылғанын ескеру қажет, бұл оларды нейрондық желіні талдау құралдарының енгізуіне жіберу кезінде қиындықтар тудыруы мүмкін. Сондай-ақ, көптеген жағдайларда кіріс сигналын алдын ала өңдеу жаттығулар деректерінде көрсету қиын шу мен бұрмалауды теңестіру арқылы нейрондық желі моделінің шығыс сигналының дәлдігін айтарлықтай арттыруы мүмкін. 1 – суретте көрсетілгендей, бейне ағыны жеке кескін кадрларынан тұрады. Сондықтан бейне ағынын өңдеу бұл процесті нақты уақыт режимінде жүзеге асыру қажеттілігін ескере отырып, жеке суреттерді өңдеу ретінде қарастырылуы мүмкін.

1 – сурет. Бейне ағынын жеке кадрлар ретінде көрсету

Қазіргі уақытта онлайн әлеуметтік желілердің саны өте көп болғандықтан, беделді Profit.kz сайтында 2023 жылдың 27 сәуірінде жарияланған «Қазақстандағы әлеуметтік желілердің рейтингі және олардан сатып алушылар не күтеді» атты мақаласы негізге алынды. Сауалнамаға қатысқандардың 56%-ы әлеуметтік желіде бейнелерді көріп, музыка тыңдайтыны, респонденттердің 44%-ы жаңалықтар оқитыны, 26%-ы бұрыннан танитын адамдармен араласатыны, 26%-ы блогерлерді бақылайтыны атап өтілген. Бұл ретте контенттің негізгі тұтынушылары 14 пен 18 жас аралығындағы жастар. Қазақстан Республикасындағы онлайн әлеуметтік желілерге қатысуды сипаттайтын мақалада анықталған негізгі көрсеткіштер 1.1-кестеде келтірілген.

1.1 – кесте.

Қазақстан Республикасындағы әлеуметтік желілерді қолдану көрсеткіші

Әлеуметтік желі атауы	**Жас көрсеткіші**	**Танымалдылығы, %**
Instagram	14-35 жас	71
TikTok		59
Telegram		57
YouTube		57
ВКонтакте		27
Facebook		13
Одноклассники	25-35 жас	8
Twitter		3

Практикалық тәжірибені ескере отырып, 1 – кестеде көрсетілген онлайн әлеуметтік желілер тізіміне қосымша LinkedIn, Viber және WhatsApp платформалары қосылды. Көрсетілген тізімдегі әлеуметтік желілер суреттер мен бейне ағындарды алдын ала өңдеуге байланысты жоғарыда айтылған мәселелерді шешу жолдарын анықтау тұрғысынан зерттелді. Белгілі бір әлеуметтік желідегі ең көп таралған кескін және бейне параметрлері ғана ескерілді.

Instagram желісіндегі сурет: пост үшін квадрат сурет 1080 × 1080 px, пост үшін тігінен сурет 1080 × 1350 px, Stories үшін сурет 1080 × 1920 px, IGTV үшін мұқаба1080 × 1920 px, Reel үшін мұқаба 1080 x 1920 px, профиль суреті 110 × 110 px кем емес. Файлдар форматтары: png, jpg, gif. Сурет жақтауларының қатынасы: 1:1; 1.91:1; 4:5.

Instagram желісіндегі бейнелер. Тік бейнелер үшін ажыратымдылығы 600 × 750 px – ден бастап 1080 × 1350 px дейін. Жақтаулар қатынасы: 4:5. Көлденең бейнелер үшін ажыратымдылық - 600 × 315 px басталып, 1080 × 566 px дейін, жақтаулар қатынасы 1,91:1 тең. Форматтар: MP4, MOV, FLV, WMV. Қолданылатын кодек H.264.

TikTok желісіндегі сурет: профиль фото 100 × 100 px, ұлғайту мүмкіндігімен 200 × 200 px, бейне иллюстрациялар үшін 1080 × 1920 px. Аватар кескіндерін пайдалану үшін 20 × 20 px. Файлдар форматтары: png, jpg.

TikTok желісіндегі бейнелер. TikTok желісінде ұсынылатын бейне ажыратымдылығы 1080 × 1920 px. TikTok бейне арақатынасы 9:16 немесе 1:1 бүйірлік панельдерімен болуы керек. Сонымен қатар, TikTok бейнелер өрісі оңынан және солынан 64 пиксель және жоғарыдан және төменнен 150 пиксель болуы қажет. TikTok көлденең бейнелерді қабылдайтынына қарамастан тік бейнелерді қолдану ұсынылады. Файлдар форматы: mov, mp4, avi, mpeg, 3pg. Қолданылатын кодек H.264.

Telegram желісіндегі сурет: профиль немесе канал суреті 640 × 640 px; пост мұқабасы: 640 × 480 px, 800 × 400 px немесе 800 × 800 px. Суреттің максималды өлшемі: ені 5000 px – ге дейін, биіктігі 3000 px – ге дейін. Суреттер файлдарының форматтары: jpg, gif, png, tiff, webp, bmp.

Telegram желісіндегі бейнелер. Telegram арнасындағы пост үшін ұсынылатын бейне өлшемдері тік бейнелер үшін 3:2 немесе 2:3 арақатынасында болуы керек, яғни 1080 × 720 пиксель немесе 720 × 1080 px, 480 × 320 px немесе 320 × 480 px. H.264 және MPEG-4 кодектер мен контейнер ретінде пайдаланылады. Бейне файлының негізгі пішімдері: mp4, mov, wmv, avi, avchd, flv, f4v, swf, mkv, webm немесе html5.

Viber суреттері. Viber хабарламасында пайдалануға болатын ең үлкен кескін өлшемі 1600 × 1200 px, стандартты өлшемі - 600 × 600 px. Кескін файл пішімі: jpg, gif, png.

Viber бейне. Ұсынылатын ажыратымдылық 640×480 px, 1280×720 px немесе 1920×1080 px. Қолайлы бейне пішімдері - mp4, avi немесе mov. H.264 кодегі бар MP4 ұсынылады. Бейненің арақатынасы жоспарланған көру әдісін ескере отырып таңдалады. Viber екі негізгі арақатынасын қолдайды: 16:9 және 4:3. Жалпы бейне жазбалар үшін 16:9 арақатынасын пайдалануды ұсынамыз, себебі бұл көптеген заманауи құрылғылар үшін стандартты қатынас.

WhatsApp суреттері. WhatsApp jpg, jpeg және png кескін пішімдерін қолдайды. Оңтайлы көрсету үшін баннерлер өлшемі кемінде 600 × 600 px, ең үлкен өлшемі 1600 × 1200 px болуы керек. WhatsApp профиль суреті: экраны 4 дюймден аз смартфондар үшін 192 × 192 px немесе 140 × 140 px. Топтық диалог профилі үшін кескін өлшемі 192 × 192 px.

WhatsApp Stories үшін ұсынылатын өлшем - 750 × 1334 px. Кескін пішімі jpg және png санаттарында болуы керек. Суреттер шаршы пішінді.

WhatsApp бейне. WhatsApp-та mp4, avi, flv, avi, wmw, mkv, 3gp, mov, vob кеңейтімі бар бейнелерді ойнатуға болады. Бейнелерді mp4 форматында жіберу ұсынылады. Қысқа бейнелер үшін ұсынылатын кадр биіктігі - 360 px немесе 480 px, ұзын бейнелер үшін - 240 px. Максималды ажыратымдылық 1280 × 720 px.

YouTube желісіндегі суреттер. Бейненің жоғарғы жағындағы сурет: 1280 × 720 px, арна мұқабасының минималды өлшемі 2048 × 1152 px, арна мұқабасының максималды өлшемі 2560 × 1440 px, профиль фотосы 800 × 800 px, Shorts мұқабасы 1920 × 1080 px. Пікірлер арақатынасы 16:9. Ұсынылатын файл пішімі - jpg және png.

YouTube желісіндегі бейне. Сатуға немесе жалға алуға арналған бейнелердің ажыратымдылығы кемінде 1920x1080 px, арақатынасы 16:9 болуы керек. Жарнамасы бар немесе жоқ мазмұн үшін YouTube ең төменгі ажыратымдылықты орнатпайды, бірақ 16:9 арақатынасы үшін кемінде 1280x720 px және 4:3 арақатынасы үшін кемінде 640x480 px ұсынады. YouTube 8 бейне форматтарын қолдайды: mp4, avi, wmv, mov, 3gp, flv, mpeg-1,2, webm. Стандартты 16:9 арақатынасы бар бейнелер үшін келесі ажыратымдылықтар пайдаланылады: 7680 × 4320 px, 3840 × 2160 px, 2560 × 1440 px, 1920 × 1080 px, 1280 × 720 px, 804 × 64 px, 854 × 60 px, 426 × 240 px.

ВКонтакте желісіндегі суреттер. Ұсынылатын өлшемдер. Сыртқы сілтемеге арналған сурет: 510 × 228 px; қауымдастық мұқабасы: 1590 × 530 px; жазбалар үшін сурет: 510 × 510 px (тікбұрышты кескінмен, арақатынасы 3:2); ВКонтакте тарихына арналған сурет: 1080 × 1920 px; клиптерге арналған мұқаба: 1080 × 1920 px; профиль фотосы: 200 × 500 px; айналмалы галерияға арналған суреттер (карусель): 400 × 400 px кем емес; әмбебап посттарға арналған сурет: 510 × 510 px (тікбұрышты кескінмен, арақатынасы 3:2); Сайттың сол жағындағы жарнамаға арналған сурет: мәтіні бар сурет 145 × 85 px, «үлкен сурет»: 145 × 165 px; қауымдастық немесе қолданба логотипі: 145 × 145 px; веб-сайтты жарнамалауға арналған сурет: логотиптің минималды өлшемі 256 × 256 px; үзінді үшін ең төменгі өлшем: 1080 × 607 px, арақатынасы 16:9); сюжеттерде жарнамаға арналған сурет: арақатынасы 9:16, өлшемі 720 × 1280 px кем емес. Кескін файл пішімі: jpg, png немесе gif (анимациясыз).

ВКонтакте желісіндегі бейне. Пост үшін ұсынылатын бейне ажыратымдылығы 1280 × 720 px немесе 1920 × 1080 px арақатынасы 16:9, ал бейнеклиптер үшін ұсынылатын ажыратымдылық - бірдей арақатынастағы 720 × 1280 px. 4K дейінгі бейне ажыратымдылығына қолдау көрсетіледі. Жарнамаға арналған бейнежазбалардың параметрлері: арақатынасы - 1:1, 4:5, 16:9, 9:16; ең аз ені - 600 px; бейне форматы - mp4, mpeg, avi, mov; бейне кодек H.264, ең төменгі ажыратымдылық 640x360, ұсынылатын ажыратымдылық 1280x720 немесе 1920x1080. Қолдау көрсетілетін пішімдері: avi, mp4, 3gp, mpeg, mov, mp3, flv, wmv, mkv, ts, vob.

LinkedIn суреттері. Сыртқы сілтемелерге арналған сурет 1200 × 627 px, постқа арналған сурет 1080 × 1920 px, мұқаба 1584 × 396 px, профиль суреті: 400 × 400 px. 16:9 қатынасы.

LinkedIn бейне. LinkedIn Native бейнелері тік немесе көлденең болуы мүмкін. Тік бейнелер автоматты түрде жаңалықтар арнасындағы шаршыға қиылады. Бастапқы бейне үшін қолайлы арақатынасы 1:2,4 немесе 2,4:1. Ажыратымдылығы 256×154 px - ден 4096×2304 px дейін. Қолдау көрсетілетін бейне форматтары: avi, mp4, mkv, flv, mpeg-4, mpeg-1, asf, webm, quicktime, wmv2, wmv3, vp8, vp9. Ұсынылатын кодек – H.264.

Одноклассникидегі суреттер. Жазбаның максималды кескін өлшемдері 1680 × 1680 px, мұқаба суреті 1944 × 600 px, профиль фотосы: 190 × 190 px. 900 × 600 px немесе 1200 × 800 px рұқсаты бар төртбұрышты пішімдегі жазбалар үшін кескіндерді пайдалану ұсынылады. Мақсатты жарнама: жарнамалық сілтеме өлшемдері - 640 × 320 px - ден

(2:1); карусельдер - 600 × 600 px; көп форматта - 1080 × 607 px (1,78: 1); баннер - 240 × 400 px (3: 5); тизер - 90 × 75 px (6: 5). Қолдау көрсетілетін пішімдері: jpg, png.

Одноклассникидегі бейне. Рұқсат етілген бейне ажыратымдылығы - 640×360 px, 1280×720 px немесе 1920×1080 px. Әлеуметтік желі барлық танымал бейне файл пішімдерін қолдайды: mpeg, webm, mov, mp4, mxf, flv, svcd, vob, m2ts, dat, 3g2, m2v, mpe, avi, wm, rmvb, m4v, mpg, ts, 3gp, mkv, asf, vcd dv, wmv, m2p, evo, divx, mcf, qt, f4v, ogg, rm, ogm.

Facebook суреттері. Ұсынылатын кескін өлшемдері: пост үшін 1200 × 630 px; сыртқы сілтеме суреті: 1080 × 1080 px; Facebook парақшасының мұқабасы: - 820 × 312 px; Facebook Stories үшін: 1080 × 1920 px; профиль фотосы: 170 × 170 px; оң жақ бағандағы баннер және кіріс хабарламалар: 1200 × 1200 px; Facebook, Facebook Marketplace және жарнамалық хабарламалардағы жылдам мақалалар үшін: 1200 × 1200 px. Кескіннің арақатынасы: Facebook арнасының жарнамасы үшін: 1,91:1-ден 4:5-ке дейін; Сыртқы сілтеме суреті: 1,91:1 - 1:1; оң жақ бағандағы баннер және кіріс хабарламалар: 16:9-дан 1:1-ге дейін; Facebook Instant Articles, Facebook Marketplace және демеуші жазбалар үшін: 9:16-дан 16:9-ға дейін; сілтемемен жарияланған кезде: 1:1; Facebook Stories үшін: 1.91-ден 9:16-ға дейін. Facebook желісінде қолданылатын суреттер үшін кескіннің шамамен 14% бос қалдыру ұсынылады. Ол үшін суреттің жоғарғы және төменгі жағына мәтінді, логотипті және суреттің басқа негізгі элементтерін қоймау керек. Кескін файл пішімдері: jpg, png, gif.

Facebook бейне. Бейне параметрлері оның локализациясының орнына байланысты: Facebook арнасы, Facebook Marketplace, Facebook желісіндегі In-stream бейне жарнамасы, Facebook Stories.

Facebook арнасына арналған бейне параметрлері: ұсынылатын бейне өлшемдері - альбомдық және портреттік бағдар үшін 1280 × 720 px; ең төменгі ені - 600 px (ұзындығы арақатынасына байланысты) пейзаждық және портреттік бағдарлау үшін; ландшафт арақатынасы - 16:9; портрет арақатынасы 9:16 (егер бейнеде сілтеме болса, арақатынасы 16:9).

360 форматындағы бейне: максималды өлшемдері - 4096 × 2048 px; ең аз ені 600 px (ұзындық арақатынасына байланысты); арақатынасы 2:1.

Бейне параметрлері Жұмыс үстелі жаңалықтар арнасының сілтемесі Бейне (пейзаж, шаршы және оң жақ баған): ұсынылатын өлшемдер - 1280 × 720 px (пейзаж, шаршы); минималды өлшемдер 600 × 315 px (1,9:1 пейзаж) немесе 600 × 600 px (шаршы); ландшафт арақатынасы - 16:9.

Карусельдегі бейне параметрлері: ұсынылатын өлшемдер - пейзаж және шаршы үшін 1080 × 1080 px; ең аз өлшемдер 600 × 315 px (1,9:1 пейзаж) немесе 600 × 600 px (шаршы).

Бейне жарнамалар жинағы: ұсынылатын бейне өлшемдері - пейзаж және шаршы үшін 1200 × 628 px; минималды өлшемдер 600 × 315 px (1,9:1 пейзаж) немесе 600 × 600 px (шаршы); ландшафт арақатынасы - 16:9.

Facebook желісіндегі слайд-шоу: ұсынылатын өлшемдер – пейзаж және портреттік бағдарлау үшін 1200 × 720 px; ландшафт арақатынасы - 16:9; Портрет арақатынасы 9:16 (егер бейнеде сілтеме болса, арақатынасы 16:9).

Facebook Stories: ең төменгі өлшемдері - 500 × 500 px; арақатынастары - 1,9:1-ден 9:16-ға дейін. Ұсынылатын бейне пішімдері - mp4 және mov.

Twitter желісіндегі суреттер. Tweet суреті: 1024 × 512 px, мұқаба суреті: 1500 × 500 px, профиль суреті: 400 × 400 px.

Твиттердегі бейне. Twitter үшін көлденең бейнелерге арналған техникалық сипаттамалар: mp4 пішімі, ұсынылатын өлшемі 1280 × 1024 px, ең аз өлшемі 32 × 32 px, максималды өлшемі 1920 × 1200 px, арақатынастары: 1:2,39 – 2,39:1. Ұсынылатын өлшем 1280 × 720 px. Ұсынылатын файл пішімдері mp4 немесе mov.

Қорытынды. Зерттеу нәтижелері көрсеткендей, ең танымал әлеуметтік желілер jpg, png, gif, bmp форматындағы суреттерді пайдаланады. Ең көп таралған бейне пішімдері mp4 және mov, ал пайдаланылатын пішімдердің тізіміне мыналар кіреді: mpeg, webm, mov, mp4, mxf, flv, svcd, vob, m2ts, dat, 3g2, m2v, mpe, avi, wm, rmvb, m4v, mpg, ts, 3gp, mkv, asf, vcd dv, wmv, m2p, evo, divx, mcf, qt, f4v, ogg, rm, ogm. Сондай-ақ кескіндер мен бейнелер үшін максималды және ең төменгі ажыратымдылық мәндерінің тізімі, кескіндер мен бейнелер үшін рұқсат етілген арақатынас диапазоны анықталды.

Осылайша, зерттеу нәтижесінде:

– Интернеттегі ең танымал әлеуметтік желілерде орналастырылған графикалық контенттің параметрлері анықталды: сурет сақталатын файл түрі, форматы, ажыратымдылығы және кескіннің арақатынасы;
– Саяси экстремизмді анықтау үшін онлайн-әлеуметтік желілерде суреттерді алдын ала өңдеудің негізгі міндеттерінің тізімі негізделген:

1. Қысылған кескінді қалпына келтіру.
2. Кескіндерді олардың өлшемін нейрондық желі моделінің енгізу өрісінің өлшеміне келтіру үшін масштабтау.
3. Суреттерді берілген түс пішіміне келтіру.
4. Интернеттегі әлеуметтік желілердегі саяси экстремизм саласына тән типтік араласулар мен бұрмалауларды теңестіру.

Пайдаланылған әдебиеттер тізімі

1. https://unesdoc.unesco.org/ark:/48223/pf0000260382/PDF/260382eng.pdf.multi
2. https://www.ivint.org/the-rise-of-digital-extremism-how-social-media-eroded-americas-political-stability/
3. https://www.sciencedirect.com/science/article/abs/pii/S0736585320300046
4. https://www.sciencedirect.com/science/article/pii/S2666307423000025
5. https://www.sciencedirect.com/science/article/abs/pii/S0736585323000357
6. https://www.sciencedirect.com/science/article/abs/pii/S1568494623002557
7. https://www.worldscientific.com/doi/10.1142/S021964922350003X
8. S. Toliupa, I. Tereikovskyi, L. Tereikovska, S. Mussiraliyeva and K. Bagitova, "Deep Neural Network Model for Recognition of Speaker's Emotion," *2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T)*, Kharkiv, Ukraine, 2020, pp. 172-176, doi: 10.1109/PICST51311.2020.9468017.
9. Shynar Mussiraliyeva, Kalamkas Bagitova and Daniyar Sultan, "Social Media Mining to Detect Online Violent Extremism using Machine Learning Techniques" International Journal of Advanced Computer Science and Applications(IJACSA), 14(6), 2023. http://dx.doi.org/10.14569/IJACSA.2023.01406146
10. Mussiraliyeva, Shynar & Bolatbek, Milana & Omarov, Batyrkhan & Bagitova, Kalamkas. (2020). Detection of Extremist Ideation on Social Media Using Machine Learning Techniques. 10.1007/978-3-030-63007-2_58.

USING MULTI-FACTOR AUTHENTICATION IN ENSURING THE SECURITY OF MEDICAL DATA

Olga Ussatova[1,3], Shakirt Makilenov[2,3*], Yerulan Kurbanbek[3],
Saule Amanzholova[1,3], Yenlik Begimbayeva[1,4]
e-mail: uoa_olga@mail.ru, shakirt.makilenov@gmail.com,
yerulan02@gmail.com, shokataeva@gmail.com
[1]*Institute of Information and Computational Technologies, Kazakhstan*
[2]*al-Farabi Kazakh National University, Kazakhstan*
[3]*International University of Information Technologies, Kazakhstan*
[4] *K.I. Satbayev Kazakh National Research Technical University*

Annotation. *In a healthcare environment where protecting patient data is a priority, multi-factor authentication provides a powerful means of securing and managing data between physical devices, users and information flows. This article explores the life cycle of the MFA and its role in the field of medical practice, from user registration to ongoing Diptychs and development. MFA, which requires multiple forms of identification, increases security beyond standard protocol-based authentication. It also complies with regulatory standards such as the Health Insurance Portability and Accountability Act and believes in the safety of health care providers. Implementing MFA is critical to ensuring patient privacy, data compliance, and legal compliance. In the face of an ever-changing cyber threat and the digital transformation of healthcare, the Ministry of Foreign Affairs remains a critical part of the strategy to ensure safety and maintain patient trust in this important area.*

Introduction

In recent years, the digitization of healthcare records has revolutionized the medical industry, offering numerous benefits in terms of efficiency and accessibility. However, this digital transformation also poses significant challenges, particularly concerning the security of sensitive medical data. Unauthorized access to patient records not only jeopardizes individual privacy but also undermines the integrity of healthcare systems. In response to these challenges, Multi-Factor Authentication emerges as a pivotal component of cybersecurity strategies in healthcare.

Patient data is highly valuable on the black market, making healthcare organizations prime targets for cybercriminals. Safeguarding the confidentiality and integrity of patient data, considering the potential consequences of unauthorized access and data breaches.

Regulatory frameworks, such as HIPAA, mandate stringent security measures to protect patient health information. The motivation is to explore how Multi-Factor Authentication aids healthcare organizations in meeting regulatory requirements, thereby avoiding legal consequences and reputational damage.

Patient trust is crucial in healthcare, and breaches erode this trust. Privacy concerns related to medical data are on the rise. Significance is understanding how Multi-Factor Authentication contributes to building and maintaining trust by ensuring the privacy and security of patient data is a key motivation.

Related work

E-healthcare facilitates the remote collection of patient health data and the provision of remote disease diagnoses to medical professionals. Roles in the system are distributed among its users and are accompanied by certain rules and policies. The system administrator represents its core, and leakage of his information can pose a threat to the entire system and its components [1].

The healthcare sector has historically generated substantial volumes of data, and with the introduction and widespread adoption of electronic health records (EHRs), this data has undergone digitization, rendering it accessible to healthcare practitioners at large. However, within the last decade, there has been a significant surge in cyber-attacks targeting healthcare institutions, resulting in breaches of patient confidentiality and disruptions to clinical care delivery [2].

An inherent susceptibility within the realm of IoT telehealth emerges during the transitional phase when a patient is en route to a healthcare facility, which may involve various modes of transportation such as ambulances or air ambulances. Alghamdi T. et al. have postulated a multifactor authentication framework as a means to fortify the security of the system during the period when the patient is in transit to the healthcare facility. [3].

Multi-Factor Authentication has emerged as a pivotal strategy to address authentication vulnerabilities. MFA mandates users to furnish multiple forms of identification, typically amalgamating passwords with additional elements such as biometrics or smart cards. Studies conducted by Garcia et al [4] emphasize the effectiveness of MFA in thwarting unauthorized access and fortifying the overall security stance of systems.

Despite the evident advantages of MFA, its seamless integration presents challenges. Investigations by Kim et al [5] and Zhang and Chen [6] probe into hurdles associated with incorporating MFA into existing healthcare infrastructures and explore factors influencing user acceptance, particularly among healthcare professionals.

Healthcare Data Breaches

The issues of ensuring the confidentiality of medical personal data are the subject of close attention in all countries that have embarked on the path of electronic medical data. Ensuring the protection of the rights and freedoms of a person and a citizen when processing his personal data, including the protection of the rights to privacy, personal and family secrets, is at the same time a strict condition. Personal data is any information related to an individual, including surname, first name, patronymic, date and place of birth, address, family, social, property status, education, profession, income, lifestyle and other information.

If you look at data leaks in the healthcare sector by year, between 2009 and 2022, 5,150 leaks of medical data from 500 or more records were reported to the HHS Office of Civil Rights. These violations resulted in the disclosure or unacceptable disclosure of 382,262,109 medical records. This is more than 1.2 times the population of the United States. In 2018, about one leak per day was reported on the leakage of medical data of 500 or more records. Fast forward 5 years, and the speed has more than doubled. In 2022, an average of 1.94 cases of medical data leakage from 500 or more records were recorded daily [7], in as indicated Figure - 1 and Figure - 2.

There is a general trend towards an annual increase in the number of disclosed records, with a sharp increase in 2015. 2015 was the worst year in history for hacking medical records: more than 112 million records have been disclosed or unacceptably disclosed. 2015 was particularly bad because of three large-scale data leaks in health insurance plans: Anthem Inc,

Premera Blue Cross and Excellus. The Anthem breach affected 78.8 million of its members, and the Primera Blue Cross and Excellus data breaches affected more than 10 million people.

For example The Center for Analysis and Investigation of Cyberattacks (TSARKA) reported on the massive data leak of patients of the “Damumed” medical system, explaining that this is the second major incident with the personal data of Kazakhstanis in a week. Anonymous informed the CIRCUS and several other organizations about a major incident: the medical information of hundreds of thousands of patients of the “Damumed” clinic network got into the network. The leak was caused by an elementary mistake - unauthorized access to the organization's medical documents [8].

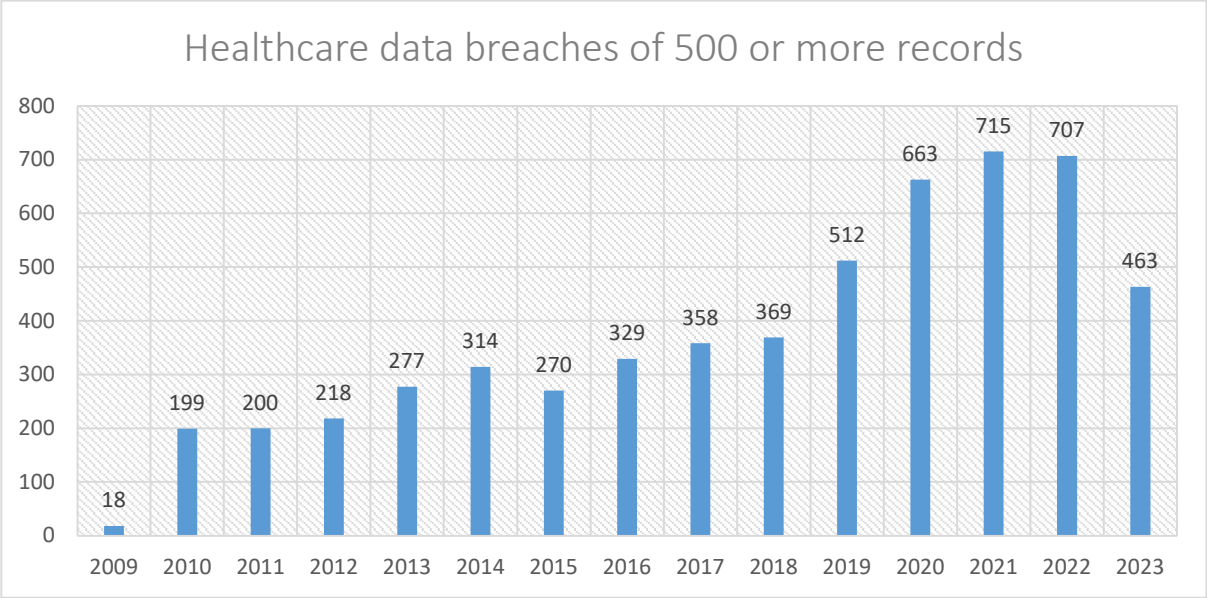


Fig. 1 Data breaches between 2009 - AUG 2023 [7]

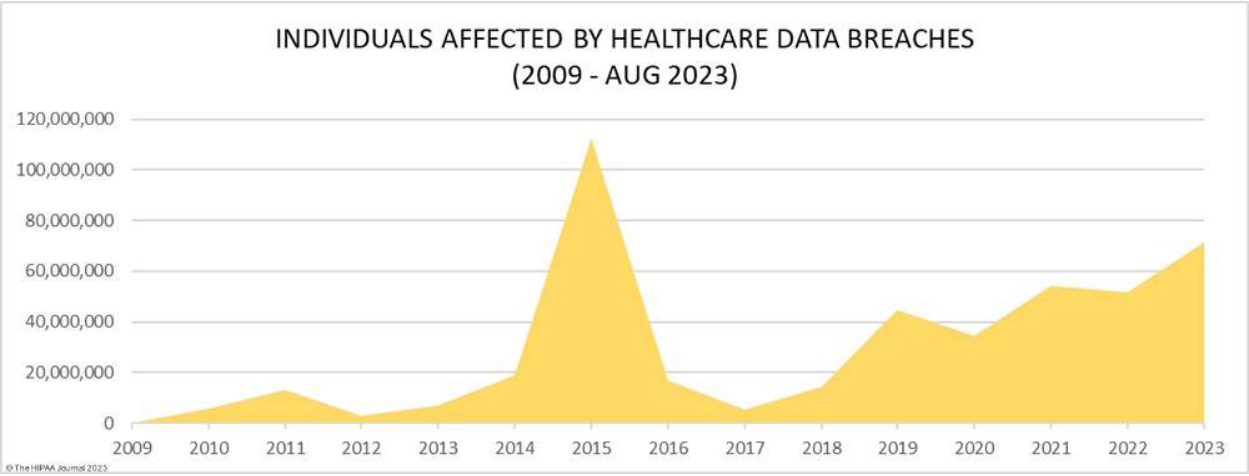


Fig. 2 Medical records disclosed by year [7]

And also on February 2, 2023, it became known about the leak of data of users of the online pharmacy "Zdravcity". According to the Telegram channel "Information Leaks", the database put up for sale on the shadow forum contains more than 8.9 million unique phone numbers and almost 3.4 million unique email addresses [9].

Table 1

Comparison table of Single-Factor Authentication (SFA), Two-Factor Authentication (2FA), and Multi-Factor Authentication (MFA)

Feature	Single-Factor Authentication (SFA)	Two-Factor Authentication (2FA)	Multi-Factor Authentication (MFA)
Number of Factors	1 (e.g., Password)	2 (e.g., Password + OTP)	More than 2 (e.g., Password + OTP + Biometric)
Security Level	Lowest	Moderate	High
Ease of Use	Easy	Moderate	May be perceived as less user-friendly due to complexity
Protection Against	Stolen Credentials	Stolen Credentials, Phishing Attacks	Stolen Credentials, Phishing Attacks, Physical Theft of Devices, etc.
Common Authentication Factors	Something the User Knows (e.g., Password)	Something the User Knows + Something the User Has	Something the User Knows + Something the User Has + Something the User Is
Examples	Password	Password + One-Time Password (OTP)	Password + OTP + Biometric (Fingerprint, Face Recognition)

Authentication methods play a crucial role in securing digital systems and data, and the choice between Single-Factor Authentication (SFA), Two-Factor Authentication (2FA), and Multi-Factor Authentication (MFA) depends on the desired level of security and the specific needs of the application. SFA, involving a single factor such as a password, is the simplest but provides the lowest security level, primarily guarding against stolen credentials. 2FA adds an extra layer of security by requiring two authentication factors, often a combination of a password and a one-time password (OTP), offering moderate security and protection against common phishing attacks. On the other hand, MFA, with more than two factors such as a password, OTP, and biometric data, provides a high level of security, making it resilient against a variety of threats, including stolen credentials, phishing, and physical device theft. While SFA is easy to use, 2FA strikes a balance between security and user-friendliness, and MFA, though perceived as less user-friendly due to complexity, is highly adaptive and suitable for highly secure environments such as healthcare systems and financial transactions. In summary, the choice of authentication method involves a trade-off between security and user convenience, with each approach serving different deployment scenarios and security requirements.

Benefits of Multi-Factor Authentication

Multi-Factor Authentication offers a robust security approach by requiring users to provide multiple forms of identification before granting access to a system or data. The benefits of implementing MFA are substantial, particularly in enhancing security across various sectors, including healthcare. Here are key benefits:

1. Enhanced Security: MFA adds an extra layer of security beyond traditional passwords, making it significantly more difficult for unauthorized users to gain access. This is especially crucial in protecting sensitive data, such as medical records, from unauthorized access and potential breaches.

2. Protection Against Credential Theft: MFA mitigates the risk of unauthorized access even if passwords are compromised. Since multiple factors are required for authentication, stolen credentials alone are insufficient for gaining access.

3. Reduced Risk of Unauthorized Access: By requiring multiple forms of identification (e.g., passwords, biometrics, smart cards), MFA reduces the risk of unauthorized access. This is particularly important in sectors where sensitive information, such as patient health records, must be safeguarded.

4. Compliance with Regulations: Many industries, including healthcare, are subject to stringent data protection regulations. Implementing MFA helps organizations comply with these regulations, such as HIPAA in the healthcare sector, by providing an additional layer of security.

5. Mitigation of Phishing Attacks: Phishing attacks, where attackers attempt to trick individuals into revealing sensitive information, are less effective against MFA. Even if credentials are phished, the additional authentication factor remains a barrier.

6. Adaptability to Evolving Threats: MFA is adaptable to changing cybersecurity threats. As new attack methods emerge, the multifaceted nature of MFA allows organizations to adjust and strengthen their security posture.

7. Protection of Remote Access: With the rise of remote work and telecommuting, securing remote access is critical. MFA ensures that even if employees access systems from different locations, multiple authentication factors are required.

8. Insider Threat Mitigation: MFA helps mitigate insider threats by requiring multiple forms of identification. This is crucial in environments where employees may have access to sensitive data but need an additional layer of verification.

9. User Accountability: MFA enhances user accountability. Each authentication factor provides a unique identifier, making it easier to trace and audit user activity. This is important for compliance and forensic purposes.

10. User-Friendly Authentication: Modern MFA solutions are designed with user convenience in mind. Methods like biometrics and push notifications make the authentication process seamless, ensuring that security measures do not hinder user productivity.

11. Integration with Various Authentication Methods: MFA can be tailored to the needs of the organization, allowing the integration of various authentication methods based on the level of security required and the usability preferences of users.

12. Protection of Critical Infrastructure: In sectors where critical infrastructure is involved, such as healthcare and utilities, MFA safeguards against unauthorized access that could compromise essential systems and services.

13. Prevention of Account Takeover: MFA is effective in preventing account takeover attempts. Even if an attacker manages to gain access to a user's credentials, the additional authentication factor acts as a barrier.

14. User Confidence and Trust: Implementing MFA communicates to users that their security is a priority. This builds confidence and trust among employees, customers, and stakeholders regarding the organization's commitment to protecting sensitive information.

Data management between devices, users, and systems in the healthcare sector using MFA

In the healthcare sector, where the protection of patient data is paramount, Multi-Factor Authentication (MFA) serves as a crucial safeguard for data management across devices, users, and systems (Fig.3). MFA ensures that only authenticated and authorized individuals can access sensitive healthcare information, contributing significantly to the overall security posture of the sector.

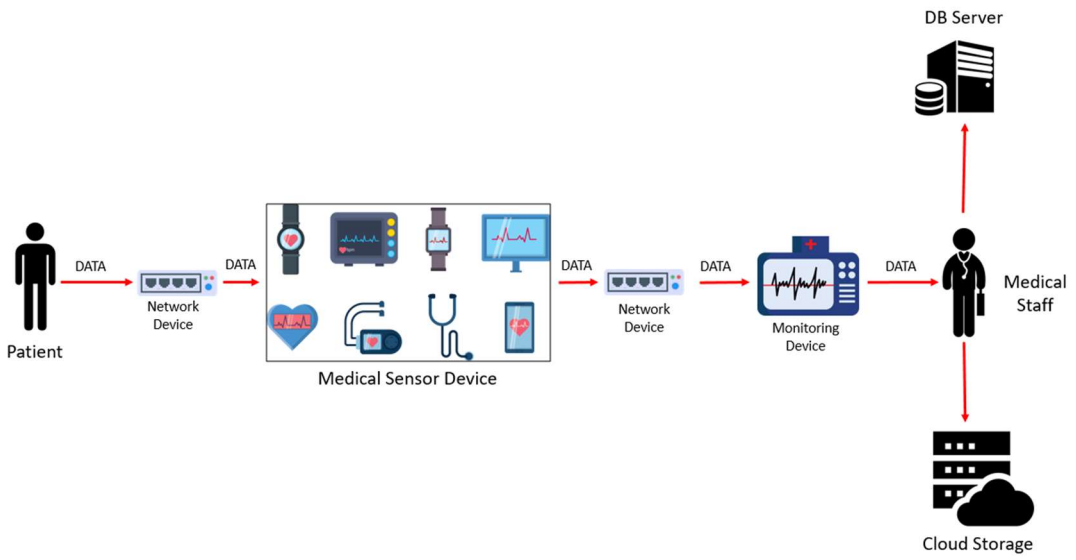


Fig. 3 Data exchange between patient and medical organization

The MFA lifecycle encompasses several stages, starting with the enrollment of users into the system. During this phase, users register their primary authentication factors, such as passwords or PINs. Subsequently, when a user attempts to access a system or device, MFA prompts the user to provide additional factors for verification. These factors may include one-time passwords generated by authenticator apps, fingerprint scans, or other biometric data (fig.4).

The lifecycle also involves ongoing monitoring and adaptation. Continuous monitoring ensures that the MFA system remains effective against evolving cybersecurity threats. In case of any suspicious activity, adaptive authentication may be triggered, prompting users for additional verification steps to prevent unauthorized access. Additionally, as technology evolves, organizations may update or enhance their MFA methods to align with the latest security standards.

Finally, the MFA lifecycle includes user management aspects, such as deprovisioning access when an individual leaves the organization or adjusting authentication methods based on user roles and security requirements. Regular training and awareness programs are vital to ensure that users understand the importance of MFA and adhere to best practices throughout its lifecycle.

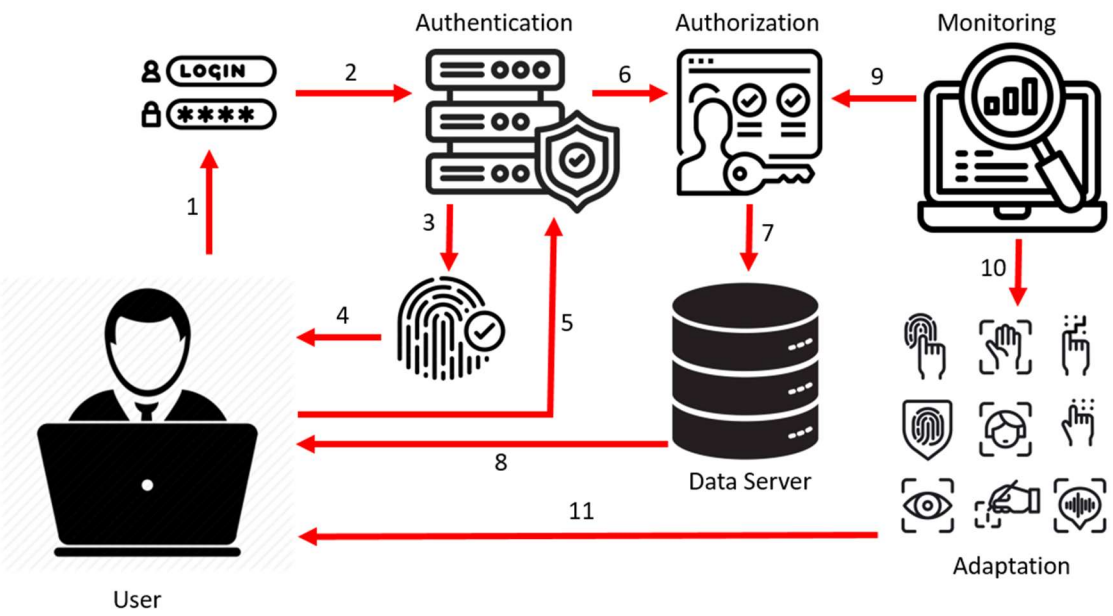


Fig.4 MFA Lifecycle

Conclusion

The implementation of Multi-Factor Authentication (MFA) in healthcare settings stands as a pivotal and proactive measure to fortify the security of medical data. The dynamic nature of the healthcare sector, coupled with the increasing sophistication of cyber threats, necessitates robust security protocols to protect sensitive patient information. The practical application of MFA across various facets of healthcare, including access to Electronic Health Records (EHRs), patient portals, telemedicine platforms, and administrative systems, exemplifies its adaptability and efficacy in safeguarding critical data.

MFA's role in mitigating the risks associated with credential theft, unauthorized access, and insider threats is particularly noteworthy. By requiring multiple forms of identification, MFA provides an additional layer of defense that goes beyond traditional password-based authentication. This not only bolsters security against external threats but also addresses challenges related to internal risks, thereby fostering a comprehensive security posture.

Furthermore, the integration of MFA aligns healthcare organizations with regulatory standards, such as those outlined in the Health Insurance Portability and Accountability Act (HIPAA). Compliance with such regulations is imperative for maintaining the trust of patients and avoiding legal ramifications.

As healthcare continues to embrace digital transformation, the significance of MFA in preserving patient privacy, ensuring data integrity, and bolstering overall cybersecurity resilience cannot be overstated. The adoption of MFA reflects a commitment to a secure and trusted healthcare ecosystem, where the confidentiality of medical information remains paramount in the face of evolving cyber threats. In essence, Multi-Factor Authentication emerges not just as a technological solution but as a strategic imperative for safeguarding the foundation of modern healthcare – patient data.

Acknowledgment

This research was carried out within the framework of the project AP19675957 «The research and development of the system for ensuring the protection of medical data using

blockchain technology and artificial intelligence methods,» which is being implemented at the Institute of Information and Computer Technologies.

References

1. Fareed M., Yassin A. A. Privacy-preserving multi-factor authentication and role-based access control scheme for the E-healthcare system //Bulletin of Electrical Engineering and Informatics. – 2022. – Т. 11. – №. 4. – С. 2131-2141.
2. McConomy B. C., Leber D. E. Cybersecurity in healthcare //Clinical Informatics Study Guide: Text and Review. – Cham : Springer International Publishing, 2022. – С. 241-253.
3. Alghamdi T. et al. Multifactor Authentication for Smart Emergency Medical Response Transporters //International Journal of Telemedicine and Applications. – 2022. – Т. 2022.
4. Garcia, M., et al. (2021). "Securing Patient Data: A Comprehensive Review of Multi-Factor Authentication in Healthcare." Journal of Cybersecurity in Medicine, 6(2), 87-104.
5. Kim, C., et al. (2021). "User Acceptance of Multi-Factor Authentication in Healthcare: A Case Study." Journal of Information Security in Healthcare, 9(4), 215-228.
6. Zhang, H., & Chen, Y. (2019). "Challenges and Solutions in Integrating Multi-Factor Authentication into Healthcare Systems." International Journal of Healthcare Information Systems and Informatics, 14(2), 58-74.
7. Healthcare Data Breach Statistics. - https://www.hipaajournal.com/healthcare-data-breach-statistics/ (29.09.2023)
8. Zakon.kz “Утечка данных пациентов - Минздрав РК выясняет причины”. - https://www.zakon.kz/4976956-utechka-dannyh-patsientov-minzdrav-rk.html?ysclid=l9qn0y6zft553658433 (29.09.2023)
9. ZDRAV.EXPERT: “2021: Продажа данных десятков тысяч пациентов”. - https://zdrav.expert/index.php/%D0%9A%D0%BE%D0%BC%D0%BF%D0%B0%D0%BD%D0%B8%D1%8F:Verimed (29.09.2023)

АЛГОРИТМ ЭЛЕКТРОННОГО ГОЛОСОВАНИЯ С ПРИМЕНЕНИЕМ ТЕХНОЛОГИИ БЛОКЧЕЙН

Макиленов Ш.Н., Аманжолова С.Т., Сагымбекова А.О., Заурбек А.
e-mail: shakirt.makilenov@gmail.com, shokataeva@gmail.com, a.sagymbekova@iitu.edu.kz, a.zaurbek@iitu.edu.kz
Казахский национальный университет имени аль-Фараби, Казахстан
Международный университет информационных технологий, Алматы, Казахстан Казахстан

Аннотация. *В статье представлено инновационное решение для цифрового голосования с использованием алгоритма отслеживаемой кольцевой подписи. Данное решение разработано с использованием Ethereum блокчейна и смарт-контрактов Solidity для создания безопасной, прозрачной и эффективной платформы электронного голосования. Применение уникального криптографического протокола, алгоритма отслеживаемой кольцевой подписи, обеспечивает анонимность избирателей, эффективно предотвращая двойное голосование. Всесторонние процедуры тестирования и проверки подтверждают прочность, надежность и готовность системы к внедрению в реальном мире. Данная система предлагает перспективный подход к повышению доверия и участия в цифровых процессах голосования, прокладывая путь к более активному и безопасному демократическому обществу.*

Введение

В настоящее время наблюдается мировая тенденция увеличения применения систем электронного голосования (Electronic Voting Systems, EVS) как средства для повышения эффективности, точности и доступности процесса выборов. Несмотря на то, что системы электронного голосования могут решить ряд проблем, связанных с традиционными методами голосования на бумажных бюллетенях, они также приносят значительные риски для безопасности и прозрачности избирательного процесса. В данном разделе предоставляется обширный обзор систем электронного голосования, включая их исторический контекст, разнообразные типы и международное распространение [1].

Существует несколько различных типов машин для электронного голосования. Среди них можно выделить следующие:

1. Системы электронной прямой записи (Direct Recording Electronic, DRE) - это наиболее распространенный вид EVS. Они позволяют избирателям голосовать, используя компьютерный экран для мгновенной фиксации своего выбора [2].

2. Системы оптического сканирования - такие системы регистрируют голоса на бумажных бюллетенях, которые затем сканируются специальными сканерами [3].

3. Методы интернет-голосования - они предоставляют избирателям возможность голосовать с использованием компьютеров или мобильных устройств через Интернет [4].

Распространение систем электронного голосования произошло в разных странах мира, включая, но не ограничиваясь США, Бразилией, Индией, Францией и Нидерланда-

ми [2]. Электронное голосование не используется в большинстве стран, и Казахстан – одна из них. Из-за опасений по поводу безопасности и прозрачности процесса голосования использование EVS вызвало споры в США. Бразилия уже давно пользуется системой EVS, и национальная система заслужила похвалу за свою эффективность и точность. EVS широко использовалась Индийской избирательной комиссией на ряде национальных выборов и выборов на уровне штата. В то время как интернет-голосование использовалось на национальных выборах в Нидерландах, EVS используется на ряде местных выборов во Франции [2].

Анализ существующих систем электронного голосования и алгоритмов отслеживаемой кольцевой подписи

Растущая тенденция цифровизации сделала цифровые системы голосования все более популярным методом голосования. В этом контексте концепция отслеживаемых кольцевых подписей стала многообещающим методом обеспечения безопасности и прозрачности структур электронного голосования.

Структуры электронного голосования (EVS) можно в общих чертах разделить на классы: системы цифровой прямой записи (DRE) и структуры интернет-голосования. Структуры DRE обычно используются на избирательных участках, где избиратели делают свой выбор на устройстве с экраном подсказок, которое передает данные о голосовании в электронном виде. По оценкам, структуры интернет-голосования позволяют гражданам голосовать удаленно, через сеть. Некоторые известные примеры систем DRE включают ES&S iVotronic, Hart InterCivic eSlate и Sequoia AVC Edge. Структуры интернет-голосования доступны с различными документами, включая интернет-системы, мобильные системы голосования и киоски для голосования. Некоторые хорошо известные примеры структур онлайн-голосования включают SecureVote, E-Voting Machine, i-Voting, Scytl и Agora Voting [6].

Алгоритмы отслеживаемой кольцевой подписи — это разновидность криптографической технологии, которую можно использовать для обеспечения конфиденциальности и защиты систем электронного голосования. Прослеживаемая кольцевая подпись позволяет группе клиентов подписать сообщение, так что можно определить автора подписи, даже сохраняя конфиденциальность альтернативных подписавших сторон. Отслеживаемые кольцевые подписи потенциально могут решить некоторые проблемы безопасности и прозрачности, связанные со структурами цифрового голосования. Например, возможность создания основы подписи обеспечивает степень прозрачности и подотчетности, которая не всегда возможна при использовании обычных цифровых подписей. Это затрудняет злоумышленникам возможность контролировать или вмешиваться в систему голосования, не будучи обнаруженными. Примеры отслеживаемых алгоритмов кольцевой подписи включают схему подписи Бонеха-Линна-Шахема (BLS), схему подписи Ring-LWE и схему подписи Ring-TESLA. Эти алгоритмы были тщательно исследованы и являются устойчивыми и экологичными [7].

Оценка существующих систем электронного голосования и алгоритмов отслеживаемой кольцевой подписи подчеркивает способность этой технологии решать проблемы безопасности и прозрачности, связанные с системами цифрового голосования. Однако, кроме того, необходимы исследования, чтобы определить единственные методы интеграции отслеживаемых кольцевых подписей в существующие структуры EVS, а также убедиться, что эти структуры стабильны, очевидны и защищены от несанкционированного доступа.

Алгоритм работы электронного голосования

Система электронного голосования была разработана для замены традиционной системы голосования на бумажных носителях, которая имеет ряд ограничений.

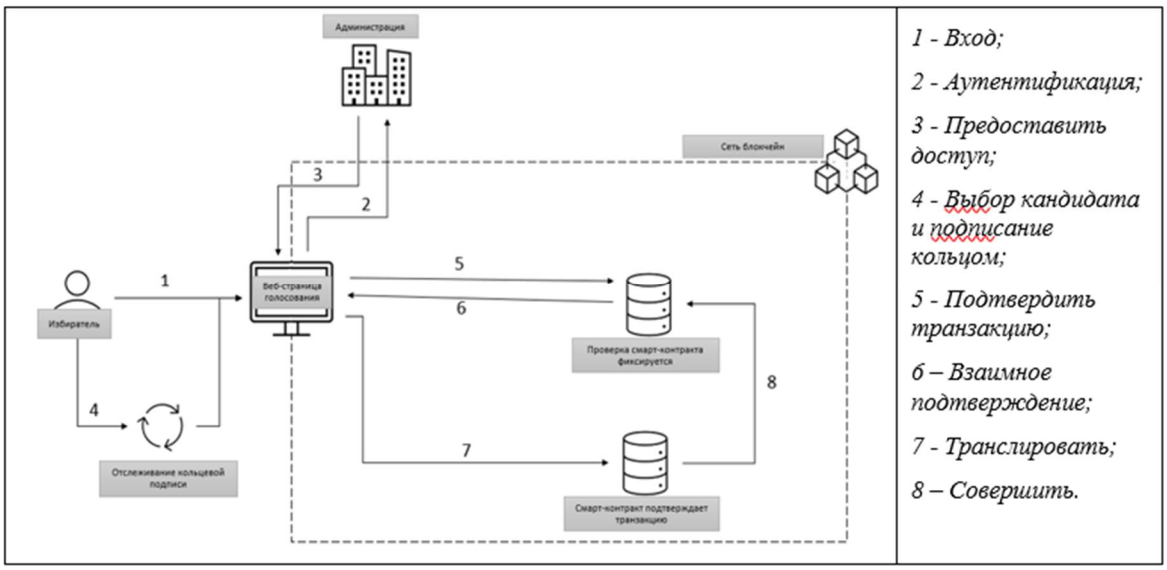


Рис. 1 Архитектура системы электронного голосования

Электронная машина для голосования - сложное устройство, необходимое для обеспечения эффективности и надежности голосования. Её конструкция включает в себя различные компоненты, такие как пользовательский интерфейс, база данных, смарт-соглашение, блокчейн и прослеживаемая кольцевая подпись. Пользовательский интерфейс обеспечивает удобство для избирателей, база данных хранит информацию о регистрации и голосовании, смарт-соглашение гарантирует безопасность, а блокчейн обеспечивает невозможность изменения данных. Прослеживаемая кольцевая подпись обеспечивает анонимность голосования. Электронное голосование - это процедура, позволяющая гражданам голосовать через Интернет, и она подвергается детальному алгоритму рабочего процесса.

Алгоритм рабочего процесса электронного голосования включает пять этапов. Первый этап - регистрация избирателей на специальном веб-сайте, где предоставляется личная информация для идентификации. Второй этап - выдача уникальных кодов избирателям для подтверждения своих голосов, обеспечивая уникальность и действительность голосов. Третий этап - голосование за кандидатов, с последующей зашифровкой и отправкой голосов для подтверждения. Четвертый этап - подсчет голосов с использованием специального программного обеспечения для обеспечения точности и предотвращения манипуляций. Пятый этап - объявление результатов выборов и возможность проверки голосов других избирателей для поддержания прозрачности и доверия к системе. Этот алгоритм обеспечивает надежный и сложный процесс голосования с высокой степенью открытости и точности [8].

Диаграмма вариантов использования

Диаграмма вариантов использования услуги электронного голосования с отслеживаемой кольцевой подписью и полностью интеллектуальным созданием расчетов на основе блокчейна является важной составной частью архитектуры этой системы.

Технология блокчейн, изначально ассоциированная с криптовалютой Биткойн, нашла широкое применение в различных областях, включая финансовые операции, и обеспечивает надежное хранение и передачу данных без посредников. Комитет по голосованию играет ключевую роль в обеспечении честности и прозрачности избирательных процессов, включая проверку точности результатов и предотвращение фальсификаций, чтобы гарантировать справедливость и надежность процедуры выборов.

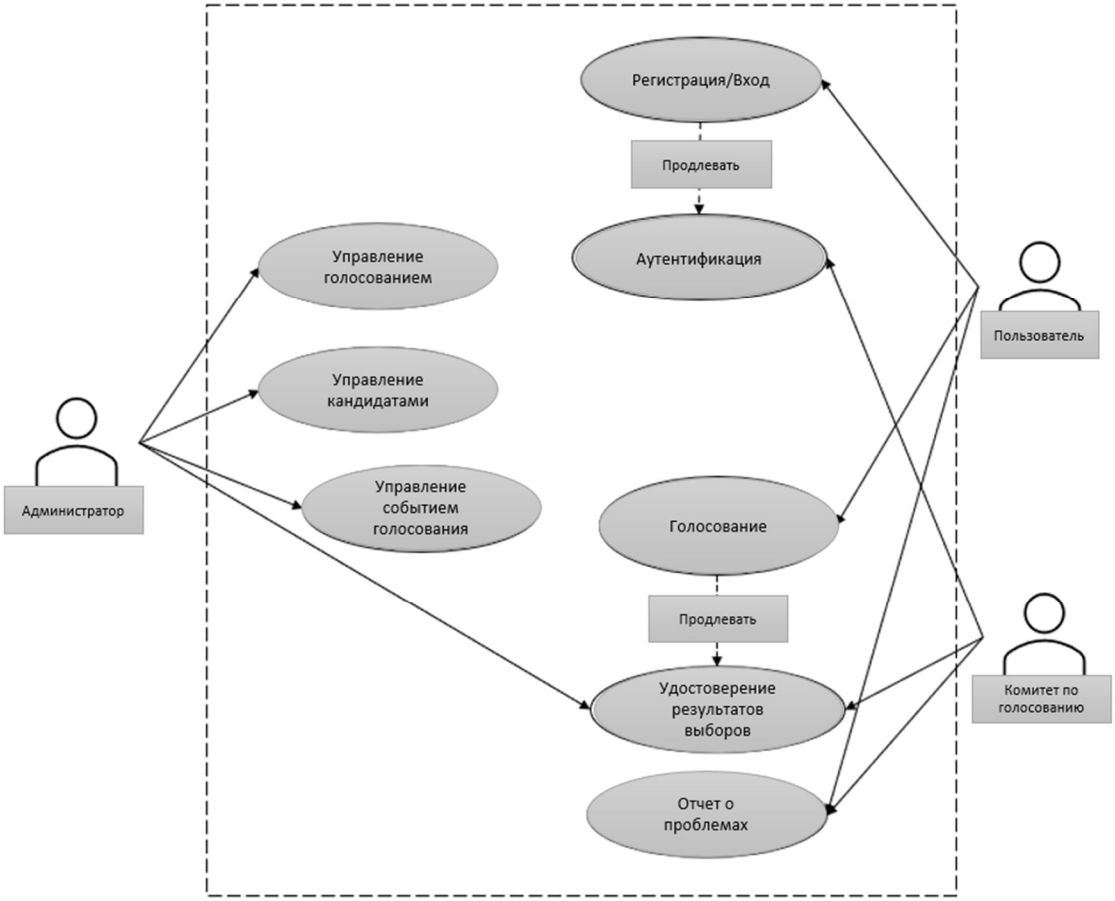


Рис. 2 Диаграмма вариантов услуги электронного голосования

Диаграмма вариантов использования услуги электронного голосования с отслеживаемой кольцевой подписью и смарт-контрактами на основе блокчейна подчеркивает важность безопасности, прозрачности и точности в избирательной технике. Используя эти технологии, голосующие могут убедиться в объективности и справедливости по отношению всех заинтересованных сторон.

Кроме того, эпоха отслеживаемых кольцевых подписей обеспечивает дополнительный уровень конфиденциальности для машины. Это позволяет пользователям сохранять анонимность, сохраняя при этом честность процедуры голосования. Каждый голос подписывается с использованием отслеживаемой кольцевой подписи, которая гарантирует, что голосование является правильным и не может быть подделано. Технология смарт-контрактов, основанная на блокчейне, также играет важную роль в архитектуре гаджета. Это обеспечивает устойчивый и прозрачный способ хранения и манипулирования фактами голосования. Блокчейн гарантирует, что каждая транзакция

записывается в защищенный от несанкционированного доступа и неизменяемый реестр, что не позволяет какой-либо несанкционированной стороне корректировать или управлять фактами.

UML-диаграмма процесса голосования

UML-диаграмма представляет систему онлайн-голосования с кольцевой подписью, включая пользователя, веб-браузер и сервер. Пользователь использует веб-сайт для голосования через веб-браузер, выбирая варианты ответов и отправляя их на сервер. Сервер обрабатывает данные и отображает результаты голосования. Кольцевая подпись обеспечивает безопасность передачи данных. Диаграмма представляет компоненты системы и их взаимодействие, что обеспечивает эффективную и безопасную процедуру голосования.

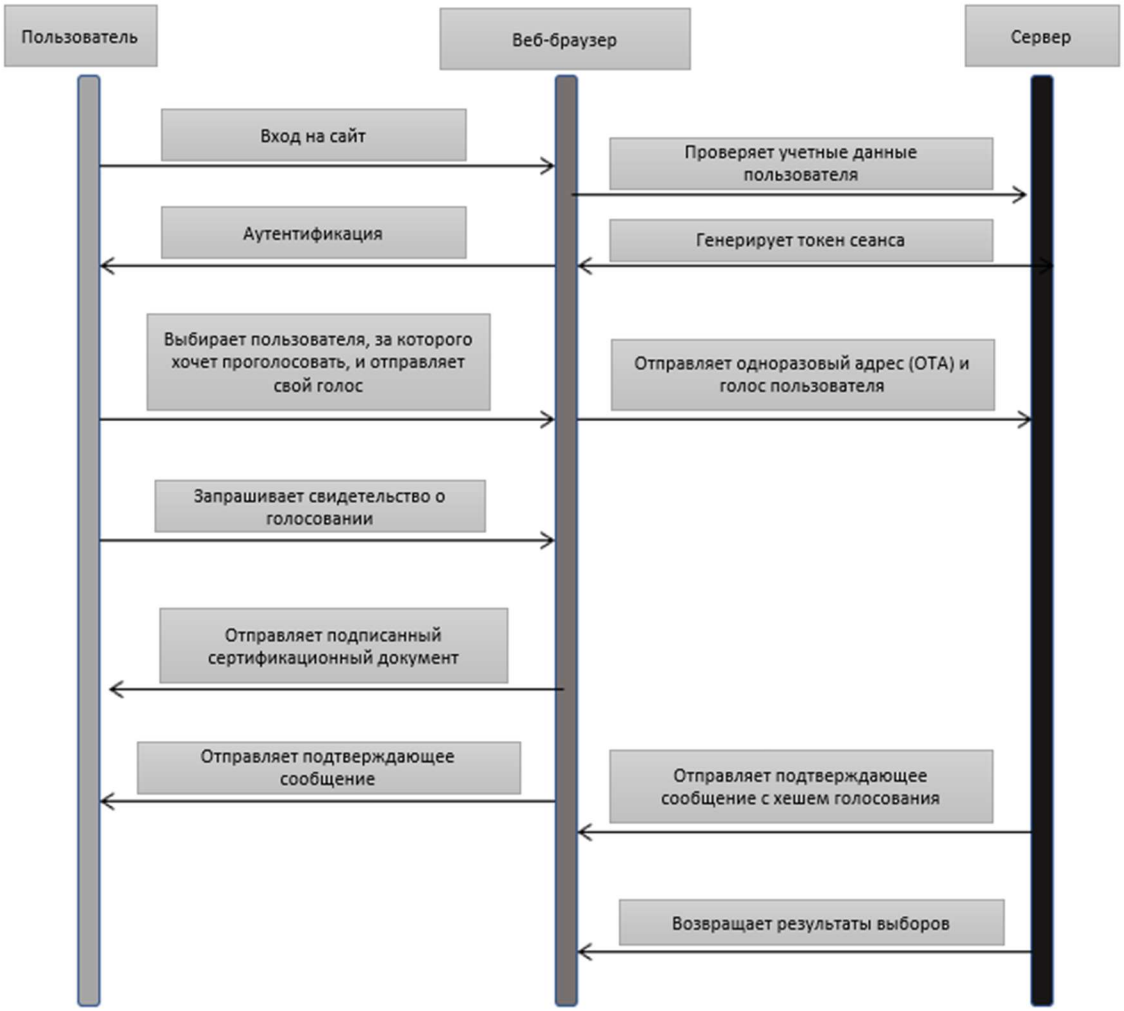


Рис. 3 Диаграмма последовательности UML

Архитектура надежной системы избирательных участков, использующей кольцевую подпись, включает в себя несколько ключевых компонентов, которые гарантируют ее бесперебойную работу и защищают ее от потенциальных атак. В основе этой системы

лежит сервер, который выполняет две основные функции: управление доступом пользователей и безопасное хранение данных голосования. Безопасность сервера имеет первостепенное значение и требует принятия надежных мер для его защиты от попыток взлома и потенциальных нарушений. Более того, сервер должен поддерживать надежные резервные копии данных голосования, чтобы обеспечить бесперебойную работу системы даже в случае сбоев или непредвиденных обстоятельств.

Для поддержки сервера у нас есть клиентские приложения и библиотеки, которые играют решающую роль в общем функционировании системы голосования с кольцевой подписью. Эти приложения и библиотеки обрабатывают регистрацию пользователей, безопасно передают информацию о выбранных кандидатах на сервер и проверяют целостность результатов голосования посредством проверки подписи. Чтобы обеспечить максимальную безопасность, эти клиентские приложения должны быть защищены от хакерских или фишинговых атак, используя самые современные криптографические методы для создания и проверки кольцевых подписей. Дополнительным важным компонентом системы голосования с кольцевой подписью является сеть связи, которая устанавливает бесшовное соединение между сервером и клиентскими приложениями. Эта сеть должна быть укреплена, чтобы обеспечить как безопасность, так и надежность, защищая от потенциальных атак или перехвата данных во время передачи. Внедрение строгих мер безопасности в сети связи еще больше укрепляет общую архитектуру системы, повышая ее устойчивость к вредоносным действиям.

Архитектура и компоненты системы голосования с кольцевой подписью сложны и многогранны и требуют надежных механизмов защиты. Тем не менее, с помощью передовых технологий и передовых криптографических методов система голосования с кольцевой подписью может стать надежным и заслуживающим доверия инструментом для демократических процессов и выборов.

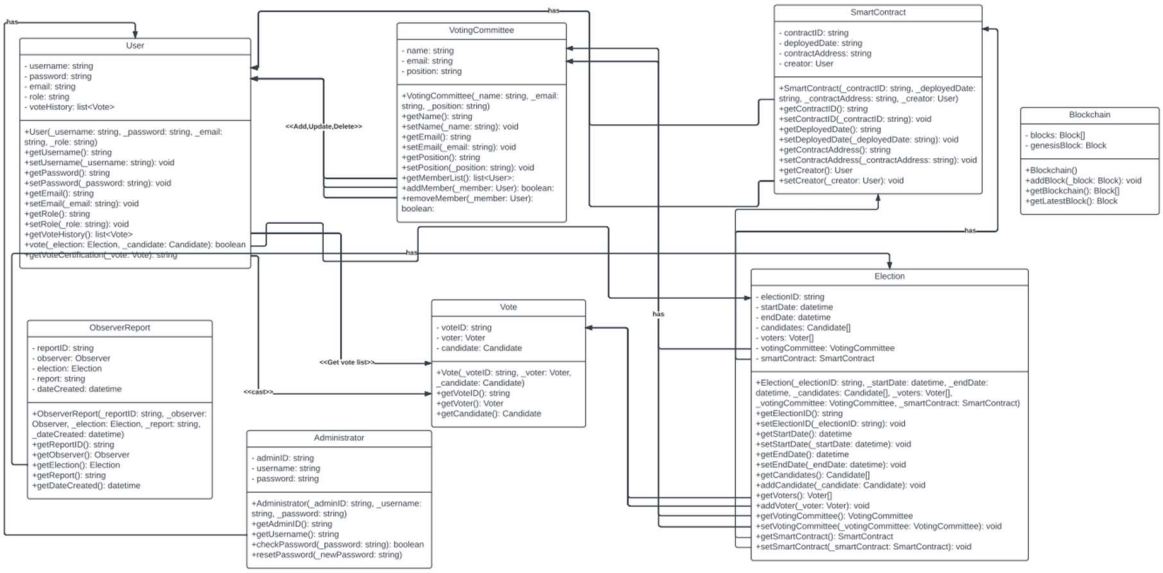


Рис. 4 Диаграмма классов UML для электронного голосования

Blockchain Environment Setup

Для создания алгоритма цифрового голосования, был инициирован локальный блокчейн на базе Ganache. Ganache основывается на персональном блокчейне Ethereum,

который позволяет проводить тесты, выполнять команды и тщательно проверять состояние блокчейна, сохраняя при этом полный контроль над его операциями. Ganache служил бесценным инструментом на протяжении всего процесса разработки, позволяя локально моделировать реальную сеть блокчейна. Настраиваемые параметры Ganache позволяет настроить среду блокчейна в соответствии с конкретными требованиями. Созданные блоки наглядно можно представить в следующем рисунке (Рис. 5).

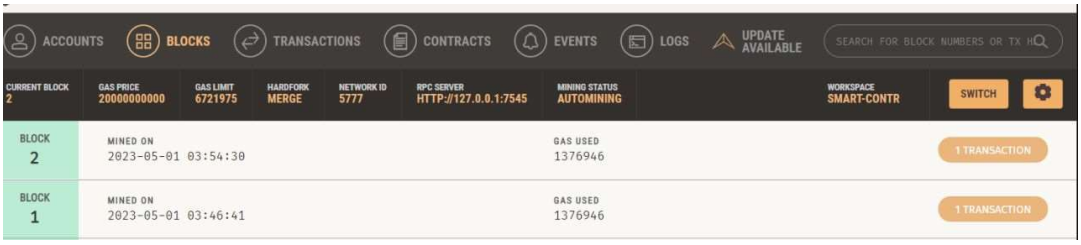


Рис. 5 Блокчейн-блоки в Ganache

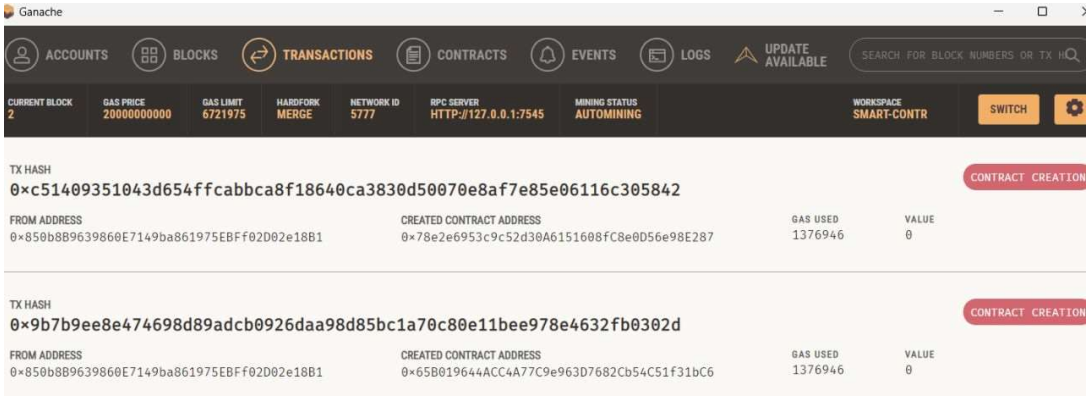


Рис. 6 Транзакции в Ganache

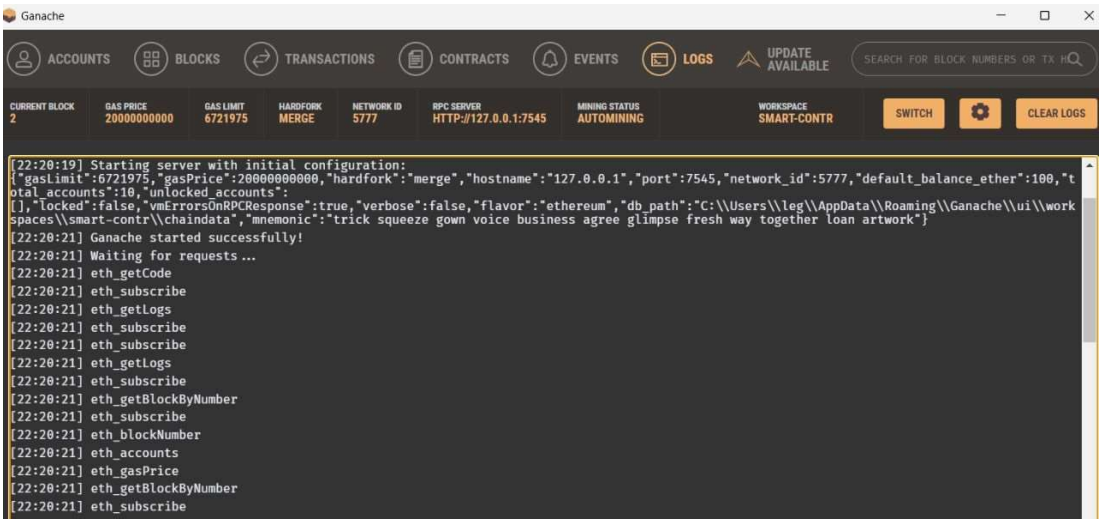


Рис. 7 Логи в Ganache

Создание локальной среды блокчейна с использованием Ganache, позволило получить гибкость и контроль над этапами тестирования и разработки функционала

цифрового голосования, а также помогла имитировать реальные операции блокчейна, обеспечивая плавную интеграцию с сетью Ethereum, сохраняя при этом безопасную и изолированную среду для нашей системы голосования.

Заключение

Создание безопасной системы цифрового голосования на основе алгоритма отслеживаемой кольцевой подписи представляет собой значительный шаг в совершенствовании процесса выборов. Это решение обеспечивает прозрачность, безопасность и доступность голосования, а также анонимность избирателей и предотвращение двойного голосования.

Оцифрованный механизм голосования делает его более удобным и доступным, что может повысить явку избирателей. Скорый подсчет голосов сокращает ожидание результатов, а повышенная безопасность и прозрачность уменьшают риски манипуляций и фальсификаций. Неизменяемость записей блокчейна гарантирует целостность системы.

Однако внедрение электронных систем голосования требует решения проблем цифрового разрыва и обеспечения доступности для всех избирателей. Также необходим постоянный мониторинг и улучшения для защиты системы от кибератак.

Напоследок, можно утверждать, что будущее голосования лежит за электронными системами. Эта работа демонстрирует, как передовые технологии могут улучшить процессы демократии. Надежно предположить, что данная работа стимулирует дальнейшие исследования и разработки в этой области, приближая нас к созданию широкодоступной и безопасной системы цифрового голосования. Несмотря на возможные трудности, потенциальные выгоды для демократического процесса оправдывают усилия, направленные на преодоление этих препятствий. "Разработка безопасного веб-сайта для цифрового голосования с использованием алгоритма отслеживаемой кольцевой подписи" - это важный этап на пути к более совершенной, безопасной и инклюзивной процедуре голосования.

Литература

1. Hussein A. I., MaoLood A. T., Gbash E. K. A Review: E-voting Security in Mobile Fog computing.
2. Ahmad A. H. et al. Electronic Voting System: Nature, Origin and Its Global Application.
3. Schürmann C. Post-Election Audits in the Philippines //E-Vote-ID 2022. – 2022. – С. 125.
4. Sherine A. et al. Development of an Efficient and Secured E-Voting Mobile Application Using Android //Mobile Information Systems. – 2022. – Т. 2022.
5. Benabdallah A. et al. Analysis of blockchain solutions for E-voting: A systematic literature review //IEEE Access. – 2022.
6. Pandey A. K. et al. Voting System Based on Blockchain Technology Smart-Contracts //International Conference on Sustainable and Innovative Solutions for Current Challenges in Engineering & Technology. – Singapore : Springer Nature Singapore, 2022. – С. 121-129.
7. Kurbatov, O., Kravchenko, P., Poluyanenko, N., Shapoval, O., Kuznetsova, T.: Using ring signatures for an anonymous E-voting system. In: 2019 IEEE International Conference on Advance Trends in Information Theory, ATIT 2019, pp. 187–190 (2019).
8. Purkayastha R., Roy A. Decentralized and Secure Blockchain-Powered Smart Card-Based Cloud Voting System //Journal of The Institution of Engineers (India): Series B. – 2023. – С. 1-8.
9. How to use Ganache to develop blockchain projects. - http://www.block-chain24.com/articles/kak-ispolzovat-ganache-dlya-razrabotki-blokcheyn-proektov (29.09.2023)

AL04 ШИФРЛАУ АЛГОРИТМІНІҢ СТАТИСТИКАЛЫҚ ҚАУІПСІЗДІГІ БАҒАЛАУ

Хомпыш А., Нысанбаева С.Е., Остапенко В.В.
e-mail: ardabek@mail.ru.
ҚР
яжЖБМ «Ақпараттық және есептеуіш технологиялар институты»

Аңдатпа. *Мақалада ақпаратты криптографиялық қорғаудың жаңа блоктық AL04 шифрлау алгоритімі ұсынылады. Ұсынылып отырған AL04 шифрлау алгоритмінің статистикалық қауіпсіздігі бағалау үшін NIST, Д.Кнут тесттері және лавиндік әсері бойынша шифрмәтінге талдаулар жүргізілді. Алгоритмге статистикалық талдау жүргізу үшін, AL04 алгоритмнің көмегімен әр түрлі өлшемдегі файлдарды шифрлап алынған шифрмәтінге талдау жасалды. Жүргізілген талдаулардың нәтижесінде AL04 алгоримінің статистикалық қауіпсіздігі жоғар нәтиже көрсеткендігін көруге болады.*

Кілттік сөздер: *NIST, статистикалық талдау, алгоритм, криптография, кілт, ашықмәтін, шифрлау, дешифрлау.*

Кіріспе

Кез-келген мемлекеттің даму стратегиясында басым бағытарының бірі ұлттық қауіпсіздік екендігін ескерсек онда оның ең маңызды элементтерінің бірі ақпараттық қауіпсіздік болып табылады. Сондықтан ақпараттық жүйелерді қауіпсіз және тиімді өңдеуді дамыту әр бір мемлекеттің басым бағыттары бірі болып табылады.

Ақпараттық қауіпсіздіктікте қолданылатын жаңа алгоритмдерді құру мәселесін шешу, бір жағынан жоғары жылдамдығын және ақпараттың үлкен көлемін жіберуі деңгейін қамтамасыз ету қазіргі заманның өзекті мәселелерінің бірі [1].

Бұл мәселелерді шешудің әр түрлі әдістер бар солардың ең тиімді әдістерінің бірі криптографиялық әдістер. Криптографиялық әдістердің міндеті мәліметтерді рұқсат етілмеген адамдардың ұрлауы, өшіруі, көруі т.б. өзгертулерден қорғау үшін криптографиялық түрлендіру әдістер арқылы мазмұнын өзгертумен айналысады. Қазіргі криптография алты негізгі бөлімді қамтиды: симметриялы криптожүйелер, ассиметриялық криптожүйелері, электрондық сандық қолтаңба (ЭЦҚ), хеш функция, кілттерді басқару, кванттық криптография [2]. Біздің қарастырлып отырған мақалада симметриялық криптожүйелердің ішінде блокты шифрлау алгоритмі. Сондықтан қазіргі блокты шифрлау алгоритмдері ақпаратты шифрлау үшін де, дешифрлау үшін де жалғыз кілт қолданылады.

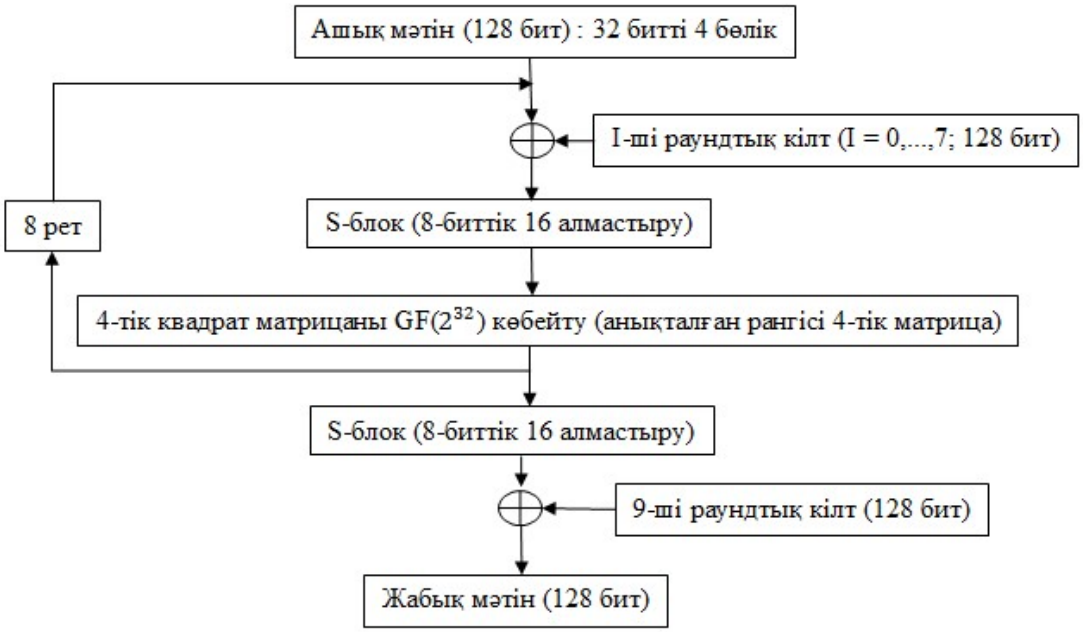
Ақпаратты криптографиялық қорғау әдістері функционалды міндеттеріне байланысты сан алуан. Мысалы, үлкен көлемді ақпаратты үлкен жылдамдықта шифрлау. Мақалада қазіргі таңдағы ақпараттық қорғау саласының өзекті болып саналатын шифрлау алгоритмінің бір AL04 жаңа блокты шифрлау алгоритмі ұсынылды.

AL04 блокты шифрлау алгоритмінің құрылымы

Ұсынылып отырған алгоритмнің шифрлау схемасы 1-суретте көрсетілген. AL04 алгоримнің негізгі сипаттамалыр:

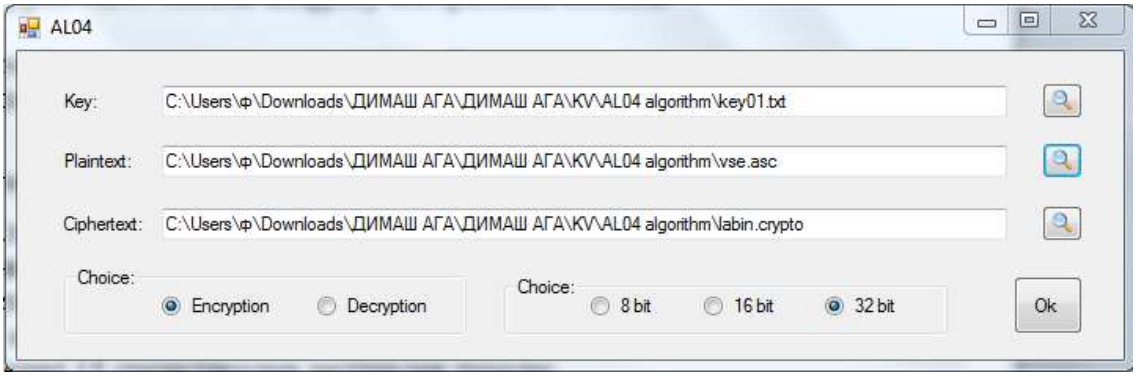
– блок ұзындығы 128 бит;
– негізгі кілттің ұзындығы 128 бит;
– раундаттардың саны 8;
– раундтық кілттің ұзындығы-128.

Алгоритмнің құрамында xor операциясы, S блок және 4 биттік квадрат матрица модуль 2^{32} бойынша көбейту операциялары орындалады.



1-сурет. AL04 блокты шифрлау алгоритмнің схемасы

AL04 алгоритмнің статистикалық қауіпсіздігіне бағалау жүргізу үшін алгоритм C# бағдарламалау тілінде жүзеге асырылды (2-сурет).



2-сурет. AL04 блокты шифрлау алгоритмнің бағдарламасы

AL04 шифрлау алгоритмінің статистикалық қауіпсіздігі бағалау

Блокты шифрлау алгоритмдерінің сенімділігін тексерудің негізгі тәсілдерінің бірі статистикалық қауіпсіздігін бағалау болып табылады, өйткені криптографиялық шабуылдардың көпшілігі шифрмәтіндегі статистикалық осалдықтарды іздеуге негізделген.

AL04 блокты шифрлау алгоритмі алынған шифр мәтіндердің статикалық талдауын жүргізу үшін NIST және Д.Кнут тесттер жинағы пайдаланылды.

NIST тесттері

АҚШ-тың Ұлттық стандарттар мен технологиялар институтының (NIST) мамандары псевдокездейсоқ тізбектер генераторыннан алынған тізбектердің кездейсоқтық қасиеттерін тексеру үшін 15 статистикалық тесттерді жасап шығарған[3]. Олар:

- жиілік биттік тесті;
- блоктык жиілік тесті;
- бірдей тізбектің биттік тесті;
- блоктағы ең ұзын тізбектің бірлік тесті;
- бинарлы матрицаның ранг тесті;
- спектральды тесті;
- қайталанбайтын үлгіні сәйкестендіру тесті;
- қайталанатын үлгіні сәйкестендіру тесті;
- Маурердің әмбебап статистикалық тесті;
- сызықтық күрделілік тесті;
- тізбектік тесті;
- энтропияға жуықтау тесті;
- жиынтық қосындылар тесті;
- еркін ауытқу тесті;
- әр түрлі еркін ауытқулар тесті.

Мұндағы барлық тестер кездейсоқтықтың әр түрлі ақауларын анықтай отырып зерттелетін тізбектің кездейсоқ алынғандығын анықтайды.

Енді ұсынылып отырған мақаладағы AL04 шифрлау алгоритмін статистикалық бағалау үшін алгоритмнің C# тілінде жасалынған бағдарламасы арқылы өлшемі 2,41МБ .djvu форматындағы файлды шифрлап алынған шифрмәтінді NIST тестілері арқылы тексереміз.

Мұндағы әрбір тестте P-мәні есептелініп алынған тесттер бойынша келесі тұжырымдалар алынады. Егер тесттер $\alpha = 0{,}01$ деп таңдалынса онда, келесі шешім қабылданады[4,5]:

- Егер $P \geq 0{,}01$ тең болса онда, тізбек 99% кездейсоқ деп қарастырамыз;
- Егер $P < 0{,}01$ тең болса онда, тізбек 99% кездейсоқ емес деп қарастырамыз.

Зерттеу нәтижесінде 1,2,3-кестелерге сәйкес AL04 шифрлау алгоритмі NIST тестілері бойынша толық өтетіндігі яғни кездейсоқ алынғандығы анықталды.

Кесте 1.

NIST тесттері бойынша AL04 алгоритмнің мәндері

№	NIST тесттері	P-мәні	Нәтиже
1	2	3	4
1	Биттік жиілік	0.7520025	Кездейсоқ
2	Блоктык жиілік	0.4490719	Кездейсоқ
3	Бірдей биттердің тізбектер	0.0338461	Кездейсоқ
4	Блоктағы ең ұзын тізбектің бірлік	0.5923904	Кездейсоқ
5	Бинарлы матрицаның ранг	0.8146149	Кездейсоқ
6	Спектральды	0.5447421	Кездейсоқ
7	Қайталанбайтын үлгіні сәйкестендіру	0.3314899	Кездейсоқ
8	Қайталанатын үлгіні сәйкестендіру	0.5295287	Кездейсоқ

1	2	3	4
9	Маурердің әмбебап статистикалық	0.6846561	Кездейсоқ
10	Сызықтық күрделілік	0.7518722	Кездейсоқ
11	Серияларды тексеру	0.1805995	Кездейсоқ
		0.1869391	Кездейсоқ
12	Энтропияға жуықтау	0.1506537	Кездейсоқ
13	Жиынтық қосындылар	0.8075868	Кездейсоқ

Кесте 2.

Еркін ауытқу тесті бойынша AL04 алгоритмнің мәні

№	NIST тесті	Күйі	Хи квадрат	P-Мәні	Нәтиже
14	Еркін ауытқу	-4	2.4369878	0.7859534	Кездейсоқ
		-3	3.7004988	0.5932897	Кездейсоқ
		-2	3.7533769	0.5854411	Кездейсоқ
		-1	2.8588235	0.7217401	Кездейсоқ
		+1	3.9019607	0.5636159	Кездейсоқ
		+2	7.7348825	0.1714647	Кездейсоқ
		+3	6.9875576	0.2215673	Кездейсоқ
		+4	4.2320046	0.5165192	Кездейсоқ

Кесте 3.

Әр түрлі еркін ауытқулар тесті бойынша AL04 алгоритмнің мәні

№	NIST сынағы	Күйі	Санауыш	P-Мәні	Нәтиже
15	Әр түрлі еркін ауытқулар	-9.0	368	0.2808730	Кездейсоқ
		-8.0	358	0.2191291	Кездейсоқ
		-7.0	358	0.1868371	Кездейсоқ
		-6.0	331	0.0910503	Кездейсоқ
		-5.0	336	0.0693626	Кездейсоқ
		-4.0	365	0.0861608	Кездейсоқ
		-3.0	392	0.0984671	Кездейсоқ
		-2.0	424	0.1200250	Кездейсоқ
		-1.0	480	0.3475580	Кездейсоқ
		+1.0	471	0.2220340	Кездейсоқ
		+2.0	416	0.0892648	Кездейсоқ
		+3.0	367	0.0452416	Кездейсоқ
		+4.0	365	0.0861608	Кездейсоқ
		+5.0	433	0.4215965	Кездейсоқ
		+6.0	438	0.4966758	Кездейсоқ
		+7.0	402	0.3483008	Кездейсоқ
		+8.0	416	0.4472874	Кездейсоқ
		+9.0	465	0.7325512	Кездейсоқ

Д. Кнут тестері

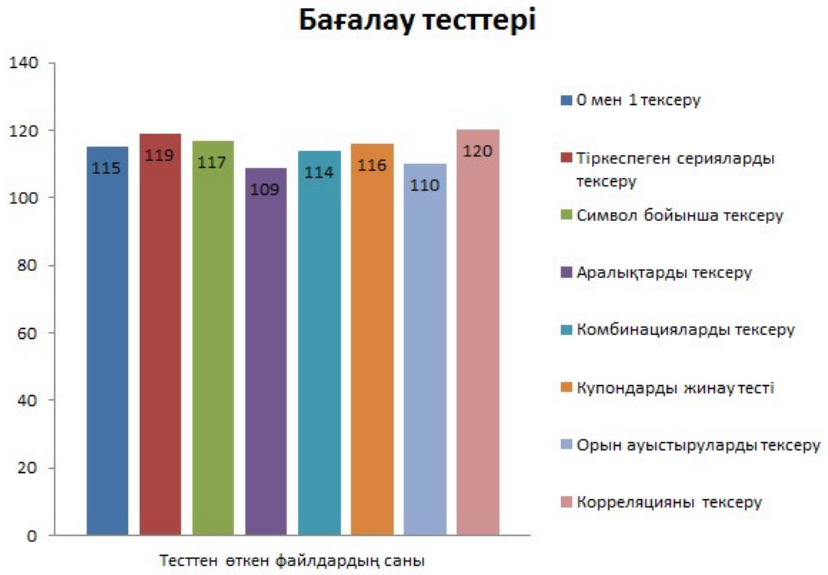
Псевдокездейсоқ тізбектер генераторыннан алынған тізбектердің статистикалық қауіпсіздігін бағалаудағы тесттер жиынтығының бірін 1969 жылы Д.Кнут ұсынған және ол туралы «Искусство программирования для ЭВМ» атты классикалық еңбегінде

сипатталған [6]. Ұсынылған алгоритмге тестілеу жүргізу үшін әр түрлі өлшемдегі 120 файлды шифрлап, алынған 120 шифрмәтіннің статистикалық қауіпсіздігі тексеріліп нәтижелеррі ұсынылды. Графикалық тесттердің нәтижелерін төмендегі 2-суретте көрсетілген.

2-сурет. Графикалық тесттердің нәтижелері

Графикалық тестердің нәтижелері: Элементтерді тарату гистограммасы, жазықтықта тарату, сериялярды тексеру, монотондылықты тексеру, байттық автокорреляциялық функция (АКФ), биттік автокорреляциялық функция (АКФ), графикалық спектралды тест, сызықтық күрделілік профилі тестеріне сәйкес 117, 118, 119, 112, 114, 117, 119, 120 шифрмәтін тестілеуден өтті.

Ал бағалау тестердің нәтижелері нақты өткендігі немесе өтпегендігі туралы нәтижелер ұсынады (3-сурет).



3-сурет. Бағалау тесттердің нәтижелері

Бағалау тестердің нәтижелері: 0 мен 1 тексеру, байланыссыз серияларды тексеру, тіркеспеген серияларды тексеру, символ бойынша тексеру, аралықтарды тексеру, комбинацияларды тексеру, купондарды жинау тесті, орын ауыстыруларды тексеру, корреляцияны тексеру тестеріне сәйкес 115, 119, 117,109,114,116, 110, 120 шифрмәтін тестілеуден өтті.

AL04 алгоритмнің лавиндік әсері

Шифрлау алгоритмінің беріктігін дифференциалды криптоталдаудан қамтамасыз етудің негізгі шарттарның бірі негізгі түрлендіруде лавиндік әсерінің болуы.

Лавиндік әсері бастапқы мәтінге (немесе кілтке) кішкентай өзгерістер енгізілсе онда шифрмәтінге айтарлықтай өзгерістер әкелуі мүмкін дегенді білдіреді.

Егер алгоритмде қажетті деңгейде лавиндік әсері болмаса, онда криптоталдаушы нәтижені негізге ала отырып ашықмәтінге шабул жасай алады. Түрлендіру кезінде лавиндік әсер ету дәрежесін сипаттау үшін лавиндік әсерінің параметрі анықталады. Лавиндік әсерінің критериі үшін лавиндік әсерінің параметрінің мәні келесі формула бойынша анықталады[7]:

$$\varepsilon_i = |2k_i - 1| \quad (1)$$

мұндағы i – ашықмәтіндегі өзгертілген биттің нөмері, k_i - ашықмәтіннің i-ші биттін өзгерткен кезде өзгертілген ашықмәтінмен шифрмәтінді салыстырғанда өзгерген шифрмәтін биттерінің жартысына жуығы өзгеру ықтималдығы.

4-Кесте.

AL04 алгоритмнің соңғы раундтағы лавиндік әсерінің нәтижесі

i	k_i	i	k_i	i	k_i	i	k_i	i	k_i	I	k_i	i	k_i	i	k_i
1	0,62	17	0,50	33	0,50	49	0,45	65	0,50	81	0,52	97	0,49	113	0,53
2	0,52	18	0,46	34	0,47	50	0,52	66	0,47	82	0,40	98	0,58	114	0,46
3	0,49	19	0,54	35	0,54	51	0,52	67	0,51	83	0,48	99	0,59	115	0,51
4	0,49	20	0,46	36	0,56	52	0,52	68	0,59	84	0,51	100	0,48	116	0,53
5	0,44	21	0,55	37	0,53	53	0,47	69	0,52	85	0,52	101	0,54	117	0,45
6	0,49	22	0,51	38	0,48	54	0,50	70	0,50	86	0,55	102	0,47	118	0,54
7	0,48	23	0,51	39	0,45	55	0,54	71	0,49	87	0,47	103	0,51	119	0,45
8	0,59	24	0,52	40	0,48	56	0,57	72	0,48	88	0,51	104	0,43	120	0,53
9	0,41	25	0,52	41	0,47	57	0,53	73	0,52	89	0,52	105	0,48	121	0,53
10	0,52	26	0,43	42	0,55	58	0,58	74	0,48	90	0,44	106	0,59	122	0,52
11	0,51	27	0,48	43	0,49	59	0,47	75	0,58	91	0,48	107	0,46	123	0,47
12	0,50	28	0,59	44	0,49	60	0,48	76	0,45	92	0,54	108	0,44	124	0,50
13	0,51	29	0,49	45	0,53	61	0,51	77	0,48	93	0,45	109	0,47	125	0,53
14	0,45	30	0,59	46	0,49	62	0,55	78	0,55	94	0,53	110	0,47	126	0,52
15	0,53	31	0,45	47	0,55	63	0,46	79	0,49	95	0,51	111	0,54	127	0,48
16	0,43	32	0,51	48	0,50	64	0,41	80	0,50	96	0,49	112	0,56	128	0,47

Қорытынды

Ұсынылып отырған AL04 алгоритімнің бағдарламасы құрылып, осы алгоритм бойынша шифрленген файлдардың шифрмәтіндерінің статистикалық қауіпсіздігін

бағалау үшін NIST тесттері, Д.Кнут және лавиндік әсері бойынша зерттеліп нәтижелері алынды. Жүргізілген зерттеулерге сай алгоритм статистикалық қауіпсіз деп бағаланады. Алгоритмнің басқа да қасиеттерін зерттеу алдағы жұмыстарда жоспарланған.

Алғыс

Жұмыс AP14870419 «ҚТ радиобайланыс бойынша келіссөздерді қорғауға арналған ақпаратты криптографиялық қорғау құралдарын әзірлеу» гранттық қаржыландыру ғылыми жобасы аясында жүргізілді.

Қолданылған әдебиеттер

1. В. Столлингс Криптография и защита сетей: принципы и практика. 2-е изд. / Пер. С англ. – М.:Вильямс, 2001.
2. Kapalova N.A., Khompysh A., Müslüm A., Algazy K. A block encryption algorithm based on exponentiation transform [Cogent Engineering], 2020, vol.7, no. 1, pp.1-12, https://doi.org/10.1080/2331 1916.2020.1788292
3. Перов А.А. Применение статистических тестов NIST для анализа выходных последовательностей блочных шифров [Научный вестник НГТУ], 2019, № 3 (76), С. 87–96. doi:10.17212/1814-1196-2019-3-87-96.
4. Sulak F., Uğuz M., Koçak O., Doğanaksoy A. On the independence of statistical randomness tests included in the NIST test suite [Turkish journal of electrical engineering & computer sciences], 2017, vol.25, no.5, pp. 3673-3683, doi:10.3906/elk-1605-212.
5. Пикуза М.О. Михневич С.Ю. Тестирование аппаратного генератора случайных чисел при помощи набора статистических тестов NIST [Доклады БГУИР], 2021, т.19, №4, С. 37-42. https://doi.org/10.35596/1729-7648-2021-19-4-37-42.
6. Д.Э. Кнут., Искусство программирования, Вильямс, 1-том, 2019 г. стр. 720.
7. Алгазы К., Бияшев Р.Г., Дюсенбаев Д.С., Ержанов Е.Б., Результаты проверки «лавинного эффекта» алгоритма «AL01», Материалы IV международной научно-практической конференции "Информатика и прикладная математика", 25-29 сентября 2019 г., Алматы, - стр. 602-607.

ТЕҢГЕРІМСІЗ ДЕРЕКТЕРМЕН ЖЕЛІГЕ РҰҚСАТСЫЗ ЕНУДІ АНЫҚТАУҒА АРНАЛҒАН ТЕРЕҢ ОҚЫТУ МОДЕЛЬДЕРІНЕ ТАЛДАУ

Сердалы А.Қ., Омаров Б.С.
e-mail: altynayserdaly@gmail.com
әл – Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан

Аннотация. *Желілік шабуылдардың саны мен түрлерінің көбеюімен дәстүрлі брандмауэрлер мен деректерді шифрлау әдістері қазіргі заманғы желілік қауіпсіздік қажеттіліктерін қанағаттандыра алмайды. Нәтижесінде желілік қауіптермен күресу үшін желіге рұқсатсыз енуді анықтау жүйелері ұсынылды. Мақалада желілік трафик-тегі ауытқуларды анықтау үшін машиналық оқытуды қолдану қарастырылады. Құрал ретінде терең оқытудың жасанды нейрондық желілері қолданылады. NSL-KDD деректер жиынтығын қолдана отырып, терең оқытудың нейрондық желілерін қолданудың тиімділігі зерттеледі. Бұл деректер жиынтығының басты ерекшелігі-сыныптардың теңгерімсіздігі. Теңгерімсіз мысалдарда қатені кері тарату арқылы нейрондық желіні оқыту үшін мақсатты функцияны тиімді пайдалану әдісі сипатталған. Бұл алгоритмді қолдану бірқатар қиындықтармен байланысты, олардың бастысы – нейрондық желіні жалпылаудың қолайлы қабілетін қамтамасыз ету. Алынған білімді жалпылау қабілеті нейрондық желінің маңызды қасиеттерінің бірі болып табылады және оқу процесіне тікелей қатыспайтын деректерде күтілетін мәндердің нейрондық желісін құрудан тұрады. Алайда, шулы және қате деректерді пайдалану қайта даярлауға және оқытылған нейрондық желіні жалпылау қабілетінің төмендеуіне әкелуі мүмкін. Ұсынылған әдіс қатенің кері таралу алгоритмінің негізінде жатқан мақсатты функцияның мәнін тиімдірек есептеуге мүмкіндік береді. Ол нейрондық деректер желілерін оқытуда гетерогенді үлгілерді қолдануға, сондай-ақ жеке мысалдардың құндылығы туралы априорлық ақпаратты оқытуда есепке алуға жақсы сәйкес келеді. Мақалада осы әдістің жұмыс алгоритмі келтірілген. Оны пайдалану жіктеу және жуықтау тапсырмалары үшін нейрондық желінің дәлдігін арттыруға мүмкіндік береді. Эксперименттік нәтижелер бұл әдіс желілік трафиктегі ауытқуларды анықтау үшін жақсы жұмыс істейтінін көрсетті.*

Кілттік сөздер: *желіге рұқсатсыз енуді анықтау, NSL-KDD деректер жиын-тығы, терең оқыту(Deep Learning).*

Кіріспе. Компьютерлік және коммуникациялық желілердің қарқынды дамуымен интернет технологиялары бүкіл әлемдегі адамдарға бұрынғыдан да ыңғайлы қызметтерді ұсынды. Алайда, кибершабуылдардың саны мен түрлері (мысалы, желілік вирустар, зиянды тыңдау, зиянды шабуылдар және т. б.) жылдан жылға артып келе жатқан адамдардың ақпараттық қауіпсіздігіне және мүліктің сақталуына елеулі қатер төндіреді. Осылайша, ақпараттық қауіпсіздік пен коммуникациялық қауіпсіздік жеке адамдар үшін де, жалпы қоғам үшін де маңызды болды [1,2]. Брандмауэрлер кең таралған және қауіпсіздіктің негізгі құралы ретінде қолданылады. Алайда, адам орнатудың күрделілігіне және шабуылдардың жаңа түрлеріне кешігуге байланысты, бұл қауіпсіз-

діктің жоғары деңгейін қажет ететін бөлімшелер үшін жеткіліксіз (мысалы, үкіметтік бөлімшелер, әскери базалар және т. б.) [3].

CheckPoint компаниясының жыл сайынғы есебінде жарияланған статистика көрсеткендей, жыл сайын белгілі шабуылдардың саны тез өсуде [4].

Кибершабуылдардың 40% өсуі - CHeckPoint. 2021 жылдың 6 қазанында Check Point Software Technologies бүкіл әлем бойынша өзекті киберқауіптерді талдау нәтижелерін хабарлады. Талдау көрсеткендей, 2021 жылы кибершабуылдар саны 2020 жылмен салыстырғанда 40% - ға өсті. 1 – суретте көрсетілгендей 2020 жылдың алғашқы апталарында аздап құлдырағаннан кейін, 2020 жылдың наурызынан бастап әлемде шабуылдар санының айтарлықтай өсуі байқалды, ол 2021 жылдың қыркүйегінде шарықтау шегіне жетті. Содан кейін әлемдегі ұйымға жасалған шабуылдардың саны аптасына 870-тен асты – бұл 2020 жылдың наурызындағы шабуылдардан екі есе көп.

Аймақтар бойынша кибершабуылдар:

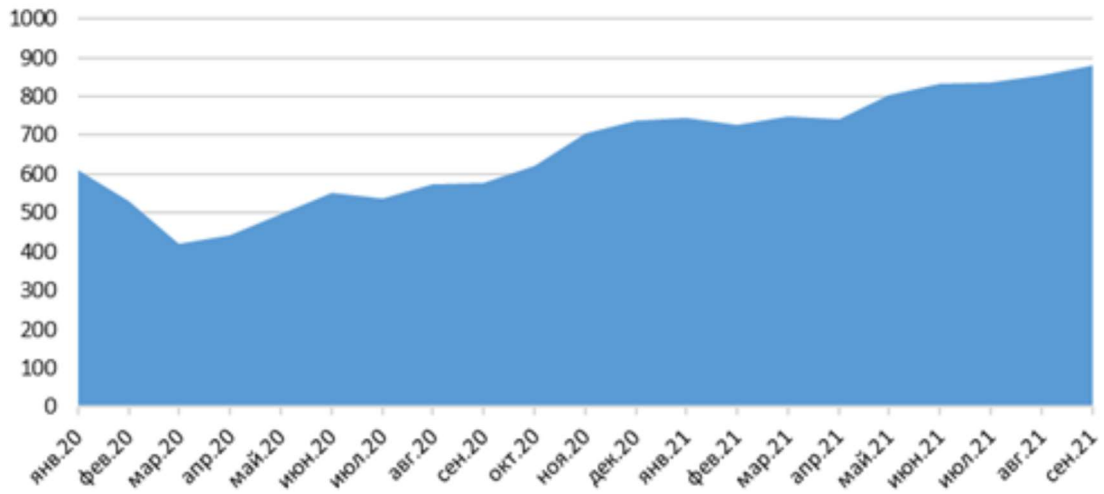
№ 1 – Африка: ұйымға аптасына орта есеппен 1615 шабуыл, бұл 2020 жылмен салыстырғанда 15% - ға көп.

№ 2 – Азия-Тынық мұхиты: ұйымға аптасына орта есеппен 1299 шабуыл өткен жылмен салыстырғанда 20% - ға жоғары.

№3 – Латын Америкасы: 1117 шабуыл, 2020 жылмен салыстырғанда 37% - ға көп.

№4-Еуропа: аптасына орта есеппен 665 шабуыл, бұл 2020 жылмен салыстырғанда 65% - ға көп.

№5-Солтүстік Америка: аптасына орта есеппен 497 шабуыл, 57% жоғарылаған.



1 – Сурет. Бір ұйымға аптасына жасалған кибершабуылдардың орташа саны (қаңтар 2020-қыркүйек 2021)

Зерттеу әдістері:Теңгерімсіз деректер негізінде желіге рұқсатсыз енуді анықтау үшін терең оқыту модельдерін қолдану.

Негізгі бөлім. Қазіргі заманғы негізгі желіге рұқсатсыз енуді анықтау алгоритмдері машиналық оқытуға негізделген, бірақ анықтау жиілігінің төмендігімен және мүмкін-діктерді мұқият әзірлеу қажеттілігімен байланысты проблемалары бар. Сондықтан желілік қауіпсіздік зерттеушілері анормальды желілерді (IDSS) басып кіруді анықтау және бақылау жүйелерін жылдам анықтаудың жаңа құралын ұсынды. Соның ішінде IDS өзін тиімді және перспективалы тәсілдердің бірі ретінде көрсетті. Ол компьютерлік жүйелердегі трафик деректерін бақылау арқылы белгілі қауіптер мен зиянды әрекеттерді

анықтайды және бұл қауіптер анықталған кезде ескертулер беріледі [5]. Зиянды әрекеттерді бақылаудың екі түрі бар. Олардың бірі-антивирустық бағдарламалық жасақтамаға ұқсас қолтаңбаға негізделген анықтау, ол бұрын жиналған шабуыл сипаттамаларымен салыстыруды қажет етеді, ал екіншісі - шешім шығару үшін әдеттегі трафикпен салыстыруды қажет ететін аномалияға негізделген анықтау.

Қазіргі уақытта тиімді модельдер жасау арқылы желілік шабуылдарды анықтау үшін желіге рұқсатсыз енуді анықтау технологиясы мен машиналық оқытуды (ML) біріктіруді жақтайтын көптеген зерттеушілер бар. Зерттеушілер [6] жұмыста аномальды желілерді анықтау үшін аңғал Байес алгоритмін қолдануды ұсынады және оны шешім ағаштарымен салыстырады (машиналық оқытудың басқа классикалық алгоритмі). [7] жұмыстағы авторлар SVM белгілерінің таңдауын, параметрлерін және салмақ коэффициенттерін оңтайландыру үшін анықтамалық векторлық әдісті (SVM) және генетикалық алгоритмді біріктіреді, осылайша желілік шабуылды анықтау дәлдігін арттырады. [8] Жұмыстағы авторлар желінің қалыптан тыс әрекетін анықтау үшін кездейсоқ орманның көп деңгейлі моделін құру арқылы анықтауды жақсартады. [9] мақала авторлары анықтау дәлдігін арттыру үшін K-MEANS кластерлеуін және K-NN классификаторын бір-бірімен біріктіру арқылы қолданыстағы k-жақын көрші классификаторын (KNN) жақсартады. Кейбір зерттеулерде [10] авторлар терең сенім желілері (DBNS) және ықтималдық нейрондық желілері (pnns) негізінде терең сенім желілері арқылы деректердің өлшемін азайту, содан кейін дәстүрлі pnns-тен жоғары ықтималдық нейрондық желіні пайдаланып деректерді жіктеу үшін модель құрылымын ұсынады. [11] жұмыстағы авторлар трафик деректерін кескін ретінде өңдеу арқылы конволюциялық нейрондық желіге негізделген анықтау әдісін ұсынады, бұл объектілерді қолмен жобалау процесін үнемдейді. Басқа зерттеуде [12] авторлар ботнеттердің ауытқуларын анықтау үшін RNN желілерін пайдаланады, ал RNN желілерінің уақыт сипаттамаларына қатысты тиімділігі жіктеу дәлдігін одан әрі жақсарту үшін қолданылады.

Желілік трафиктің ауытқуларын анықтау алгоритмдерін тестілеу тапсырмасына арналған көптеген мәліметтер жиынтығы бар. Талдау жүргізілген жұмыстар NSL-KDD сынақ деректер жинағын пайдалана отырып, желілік трафиктің ауытқуларын анықтаудың жаңа әдістерін әзірлеуді қарастырды. KDD99 деректер жиынтығында, Stolfo et al. желілік шабуылдарды төрт санатқа жіктеді, атап айтқанда: қызмет көрсетуден бас тарту (DoS) шабуылы, пайдаланушы—түбірлік шабуыл (U2R), қашықтағы-жергілікті шабуыл (R2L) және зондтау шабуылы [13]. KDD99 деректер жиынтығы NSL-KDD жиынтығының жаңартылған нұсқасы болып табылады, ол желіге рұқсатсыз енуді анықтау құралдарын сынау үшін стандартқа айналды.

NSL-KDD деректер жинағы стандартты KDD99-ға қарағанда бірқатар артықшылықтарға ие [13]:

- жіктеу алгоритмдеріне әсерін жою үшін бірқатар артық және қайталанатын деректер жойылды;
- қайталанатын жазбалар жойылды;
- жазбалар оқу жазбаларына және тестілеуге арналған жазбаларға теңдестірілген түрде бөлінді, бұл оларды кездейсоқ бөлу қажеттілігін болдырмайды.

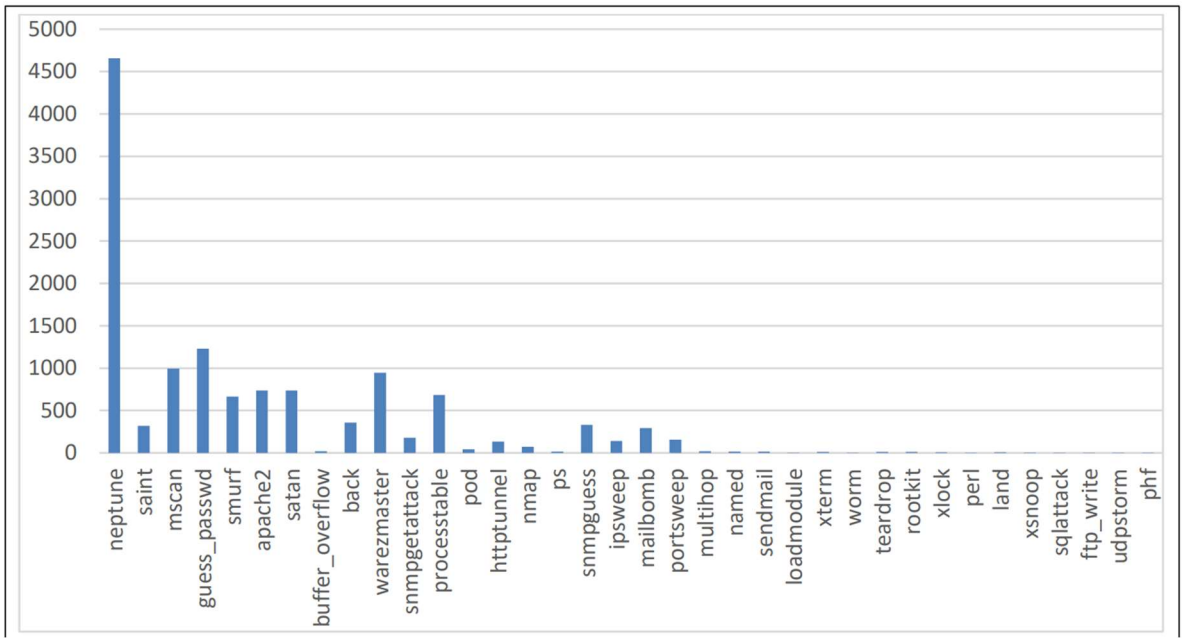
NSL-KDD Dataset базасындағы әрбір жазба уақыт аралығында бекітілген пакеттер тізбегі болып табылады. Берілген реттілік-бұл пакет тақырыбындағы IP мекенжайына сәйкес желі пакетінің көзі мен адресаты арасындағы деректер ағыны. NSL-KDD деректер жиынтығында барлығы 42 өлшемді нысан бар, олардың бірі жіктеу белгісі, ал қалғандары объект белгілері. Екілік жіктеуде жіктеу белгілері екі санатқа бөлінеді, яғни

қалыпты және қалыптан тыс. Бірнеше классификация үшін жіктеу белгілері бес санатқа бөлінеді, яғни қалыпты, Dos, R2L, U2R және probe.

Жазбаларға мыналар кіреді: 41 ақпараттық белгі және "шабуыл", "шабуыл емес"деп бөлінген белгілеулер. Базада төрт санатқа бөлінген 36 шабуыл түрі бар:

- Denial of Service (dos); шабуылдаушы белгілі бір протокол арқылы белгілі бір қызметке кіруді шектейтін шабуылдар жиынтығы (Back, Land, Neptune, Pod, Smurf, Teardrop, Apache2, Udpstorm, Processtable, Worm);
- Remote to Local (r2l); шабуылдаушы пайдаланушының жергілікті машинасына сырттан кіруге тырысатын шабуылдар жиынтығы (Guess_Password, Ftp_write, Imap, Phf, Multihop, Warezmaster, Warezclient, Spy, Xlock, Xsnoop, Snmpguess, Snmpgetattack, Httptunnel, Sendmail, named);
- User to Root (u2r); шабуылдаушы жәбірленушінің компьютеріне қол жеткізе отырып, пайдаланушы құқығын алуға тырысатын шабуылдар жиынтығы (Buffer_overflow, Loadmodule, Rootkit, Perl, Sqlattack, Xterm, Ps);
- Probe; шабуылдаушы пайдаланушының инфрақұрылымы туралы ақпарат алуға тырысатын шабуылдар жиынтығы (Satan, Ipsweep, Nmap, Portsweep, Mscan, Saint).

NSL-KDD-де оқуға арналған 125 973 жазба және тестілеуге арналған 22 544 жазба бар. Желілік шабуыл түрлері бойынша NSL-KDD жиынтығында берілген деректерді тарату гистограммасы 2-суретте көрсетілген.



2 – Сурет. NSL-KDD жиынтығында берілген шабуыл түрлерінің таралу гистограммасы

2-суретте көрсетілген диаграммадан көріп отырғаныңыздай, деректер жиынтығында көрсетілген шабуыл түрлерінің суреттері теңдестірілмеген. Бұдан басқа, "шабуыл емес" деп белгіленген трафик кескіндерінің саны 9 711 құрайды. Күрделі желілік жағдайды шындықта қалпына келтіру үшін оқу жиынтығында тек 19 шабуыл түрі бар, ал қалған 17 шабуыл түрі тек сынақ жиынтығында болады. Желілік трафик деректерінің біркелкі таралмауынан, деректер теңгерімсіздігі мәселесі туындайды. Бұл мәселені шешу үшін [14] мақалада теңгерімсіз үлгілердің деректерін көбейту үшін

адаптивті синтетикалық үлгіні (ADASYN) және деректердің өлшемін азайту үшін модификацияланған автоэнкодерді қолдана отырып, жоғарыда аталған қалған мәселелерді шешуге арналған dlnid моделі ұсынылады [15].

Осы құжатта келесі материалдар ұсынылған:

– Назар аудару механизмі мен Bi-LSTM біріктіретін dlnid моделі ұсынылған. Бұл dlnid моделі желілік трафик туралы деректерді дәл жіктеуге мүмкіндік береді;
– Теңгерімсіз желілік деректер мәселесін шешу үшін ADASYN үлгілердің кішірек класының деректерін ұлғайту үшін пайдаланылады, бұл сайып келгенде, үлгінің жеткілікті түрде үйренуіне мүмкіндік беретін әрбір түрдің үлгілерінің санын салыстырмалы түрде симметриялы етеді;
– Жақсартылған көп деңгейлі автоэнкодер ұсынылды, ол ақпаратты біріктіруді жақсарту үшін деректердің өлшемін азайту үшін қолданылады.

Нәтижелерді талдау. Эксперимент сәйкесінше екілік және көп классификациялық эксперименттер үшін әдеттегі, Dos, R2L, U2R және probe-де ұсынылған желінің өнімділігін зерттеді. Желі параметрлері 3 – суретте көрсетілгендей таңдалғанда, KDDTest + тест жинағы арқылы жоғары дәлдікке және F1 ұпайына қол жеткізуге болады. Талдаудан деректерді ұлғайту процесіне қарамастан, U2R санаты қате жіктелуі ықтимал деген қорытынды жасауға болады, өйткені сынақ жинағындағы U2R санаты оқу жинағындағы басқаларға қарағанда әлдеқайда үлкен болды.

Type	Rec (%)	FPR (%)
Normal	92.14	13.44
Dos	80.96	2.47
Probe	76.33	3.94
R2L	65.76	2.77
U2R	24.00	1.73

3 – сурет. Модельдерді салыстыру.

4-суретте ұсынылған DLNID моделі дәлдік тұрғысынан басқа анықтамалық модельдермен салыстырылады және DLNID дәлдігі басқа модельдерге қарағанда жоғары екенін көруге болады. Бұд суретте ұсынылған модель модельдер әр түрлі өнімділік көрсеткіштері бойынша салыстырылады, олардан ұсынылған dlnid моделі F1 дәлдігі мен бағалауы бойынша салыстыру үшін аналогтардан асып түседі, сәйкесінше KDDTest+ деректер жиынтығында 90,73% және 89,65% жетеді. GAR-Forest немесе NB Tree сияқты дәстүрлі Машиналық оқыту әдістерімен салыстырғанда, ұсынылған әдіс белгілерді қолмен алуды қажет етпейді және дәлдік деңгейін арттырады.

Type	ACC (%)	Pre (%)	Rec (%)	F1 (%)
DLNID	90.73	86.38	93.17	89.65
DLHA [28]	87.55	88.16	90.14	89.19
BAT-MC [14]	84.25	-	-	-
Autoencoder [29]	84.24	87.00	80.37	81.98
CNN [30]	80.13	-	-	-
Adaptive Ensemble [31]	85.20	86.50	86.50	85.20
TES-IDS [32]	85.79	88.00	86.80	87.39
GAR-Forest [33]	85.06	87.50	85.10	85.10
CNN+BiLSTM [34]	83.58	85.82	84.49	85.14
NB Tree [5]	82.02	-	-	-
SVM-IDS [35]	82.37	-	-	-

4 – сурет. Модельдер арасындағы жалпы өнімділікті салыстыру

Қорытынды. Желілік деректердегі деректер теңгерімсіздігі мәселесін және анықтаудың төмен дәлдігін шешу үшін біз ADASYN қайта іріктеу алгоритмін желіге кіру кезіндегі деректер теңгерімсіздігі мәселесін шешу үшін деректер көлемін ұлғайту әдісі ретінде жоғары нәтижеге жеткенін көре аламыз Жүргізілген зерттеулер нейрондық желілерді қолдануға негізделген желілік трафиктегі ауытқуларды анықтау моделін құру тәсілінің тиімділігін көрсетеді. Желіге рұқсатсы енуді анықтау тиімділігін арттыру үшін зерттеуді жалғастыру қажет. Әрі қарайғы жұмыстың негізгі міндеті-нақты объектілерде жұмыс істеу кезінде әлі де туындауы мүмкін жалған позитивтердің санын азайту.

Пайдаланылған әдебиеттер:

1. Patel, A.; Qassim, Q.; Wills, C. A survey of intrusion detection and prevention systems. *Inf. Manag. Comput. Secur.* 2018, *18*, 277–290.
2. Khraisat, A.; Gondal, I.; Vamplew, P.; Kamruzzaman, J. Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity* 2019, *2*, 20.
3. Yuan, L.; Chen, H.; Mai, J.; Chuah, C.N.; Su, Z.; Mohapatra, P. Fireman: A toolkit for firewall modeling and analysis. In Proceedings of the 2006 IEEE Symposium on Security and Privacy (S&P'06), Berkeley/Oakland, CA, USA, 21–24 May 2006; IEEE: Manhattan, NY, USA, 2006; pp. 15–213.
4. Check Point Security Report 2019. Available at: https://blog.checkpoint.com/2019/03/04/check-points2019-security-report// (accessed November 07, 2020).
5. Musa, U.S.; Chhabra, M.; Ali, A.; Kaur, M. Intrusion detection system using machine learning techniques: A review. In Proceedings of the 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 10–12 September 2020; IEEE: Manhattan, NY, USA, 2020; pp. 149–155.
6. Amor, N.B.; Benferhat, S.; Elouedi, Z. Naive bayes vs. decision trees in intrusion detection systems. In Proceedings of the Proceedings of the 2004 ACM Symposium on Applied Computing, Nicosia, Cyprus, 14–17 March 2004; Association for Computing Machinery: New York, NY, USA, 2004; pp. 420–424.

7. Tao, P.; Sun, Z.; Sun, Z. An improved intrusion detection algorithm based on GA and SVM. *IEEE Access* 2018, *6*, 13624–13631.
8. Jiadong, R.; Xinqian, L.; Qian, W.; Haitao, H.; Xiaolin, Z. A multi-level intrusion detection method based on KNN outlier detection and random forests. *J. Comput. Res. Dev.* 2019, *56*, 566.
9. Shapoorifard, H.; Shamsinejad, P. Intrusion detection using a novel hybrid method incorporating an improved KNN. *Int. J. Comput. Appl.* 2017, *173*, 5–9.
10. Zhao, G.; Zhang, C.; Zheng, L. Intrusion detection using deep belief network and probabilistic neural network. In Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Guangzhou, China, 21–24 July 2017; IEEE: Manhattan, NY, USA, 2017; Volume 1, pp. 639–642.
11. Wang, W.; Zhu, M.; Zeng, X.; Ye, X.; Sheng, Y. Malware traffic classification using convolutional neural network for representation learning. In Proceedings of the 2017 International Conference on Information Networking (ICOIN), Da Nang, Vietnam, 11–13 January 2017; IEEE: Manhattan, NY, USA, 2017; pp. 712–717.
12. Torres, P.; Catania, C.; Garcia, S.; Garino, C.G. An analysis of recurrent neural networks for botnet detection behavior. In Proceedings of the 2016 IEEE Biennial Congress of Argentina (ARGENCON), Buenos Aires, Argentina, 15–17 June 2016; IEEE: Manhattan, NY, USA, 2016; pp. 1–6.
13. Tavallaee, M.; Bagheri, E.; Lu, W.; Ghorbani, A.A. A detailed analysis of the KDD CUP 99 data set. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, ON, Canada, 8–10 July 2009; IEEE: Manhattan, NY, USA, 2009; pp. 1–6.
14. Engen, V.; Vincent, J.; Phalp, K. Exploring discrepancies in findings obtained with the KDD Cup 99 data set. *Intell. Data Anal.* 2011, *15*, 251–276.
15. Fu, Y.; Du, Y.; Cao, Z.; Li, Q.; Xiang, W. A Deep Learning Model for Network Intrusion Detection with Imbalanced Data. Electronics 2022, 11, 898. https://doi.org/10.3390/electronics11060898

ПРИМЕНЕНИЕ МЕТОДОВ РОЛЕВОЙ МОДЕЛИ АВТОРИЗАЦИИ В ЗДРАВООХРАНЕНИИ ЧЕРЕЗ БЕЗОПАСНЫЕ ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ

Усатова О.А.[1], Искендирова А.А.[2]
email:uoa-olga@mail.ru, iskendirova_anar1@live.kaznu.kz
[1]*Институт информационных и вычислительных технологий КН МНВО РК, Алматы, Республика Казахстан*
[2]*КазНУ имени аль-Фараби, Алматы, Республика Казахстан*

Аннотация. *В данной научной статье исследуется актуальная проблема обеспечения безопасности и конфиденциальности медицинских данных в современном информационном обществе. Статья рассматривает применение ролевой модели авторизации в здравоохранении с использованием безопасных облачных вычислений как способа эффективного управления доступом к информации. В статье представлен обзор существующих методов авторизации в сфере здравоохранения и их ограничений, а также исследованы возможности интеграции ролевой модели доступа с облачными вычислениями. Освещены преимущества и риски такого подхода, а также предложены практические рекомендации по внедрению данной модели в инфраструктуру здравоохранения. Исследование призвано повысить уровень безопасности и эффективности управления медицинскими данными, что способствует улучшению качества медицинских услуг и управления здравоохранением в целом.*

Введение

В современном информационном обществе, где обработка и передача данных становятся все более цифровыми и распределенными, обеспечение безопасности и конфиденциальности становится одной из наиболее актуальных задач. Особенно это важно в сферах, где обрабатывается чувствительная и личная информация, такая как здравоохранение. В связи с этим, разработка эффективных методов авторизации и контроля доступа к медицинским данным становится неотъемлемой частью инфраструктуры здравоохранения.

Целью данной научной статьи является исследование и анализ применения методов ролевой модели авторизации в сфере здравоохранения с использованием безопасных облачных вычислений. Ролевая модель предоставляет гибкую систему управления доступом, основанную на определении ролей пользователей и соответствующих им прав доступа. Безопасные облачные вычисления, в свою очередь, обеспечивают высокий уровень доступности, масштабируемости и защиты данных, что является особенно важным аспектом при работе с медицинской информацией.

В данной статье будет рассмотрен обзор существующих методов авторизации в здравоохранении и их ограничения. Также будет проведено исследование возможностей интеграции ролевой модели доступа с облачными вычислениями для обеспечения безопасности медицинских данных [1]. Особое внимание будет уделено анализу преимуществ и рисков такого подхода, а также предложены практические рекомендации по реализации и внедрению данной модели в инфраструктуре здравоохранения.

Исследование в этой области имеет большое значение, так как оно способствует повышению уровня безопасности и конфиденциальности медицинских данных, а также обеспечивает эффективное управление доступом к ним. В итоге, успешная интеграция ролевой модели авторизации в облачные вычисления может значительно улучшить качество оказания медицинских услуг и управление здравоохранением в целом.

Основная часть

Ролевая модель авторизации является методом, в котором каждый пользователь обладает определенными правами доступа к ресурсам в системе на основе его роли. Роли могут быть определены в соответствии с должностными обязанностями, уровнем доступа или другими параметрами, связанными с организационной структурой. Пользователи, обладающие определенными ролями, могут получать доступ только к тем ресурсам, которые соответствуют их роли [2]. Это упрощает управление доступом и уменьшает вероятность несанкционированного доступа к ресурсам.

В здравоохранении ролевая модель авторизации может быть использована для управления доступом к медицинской информации. Например, медицинский персонал может иметь различные роли, такие как врач, медицинская сестра, администратор и т.д. Каждая роль может иметь определенный уровень доступа к медицинской информации, соответствующий его обязанностям и полномочиям. Например, врач может иметь доступ к медицинским записям пациента, тогда как медицинская сестра может иметь доступ только к информации о результатах анализов. Управление доступом на основе ролей (Role-BasedAccessControl – RBAC) представлено на рисунке 1.

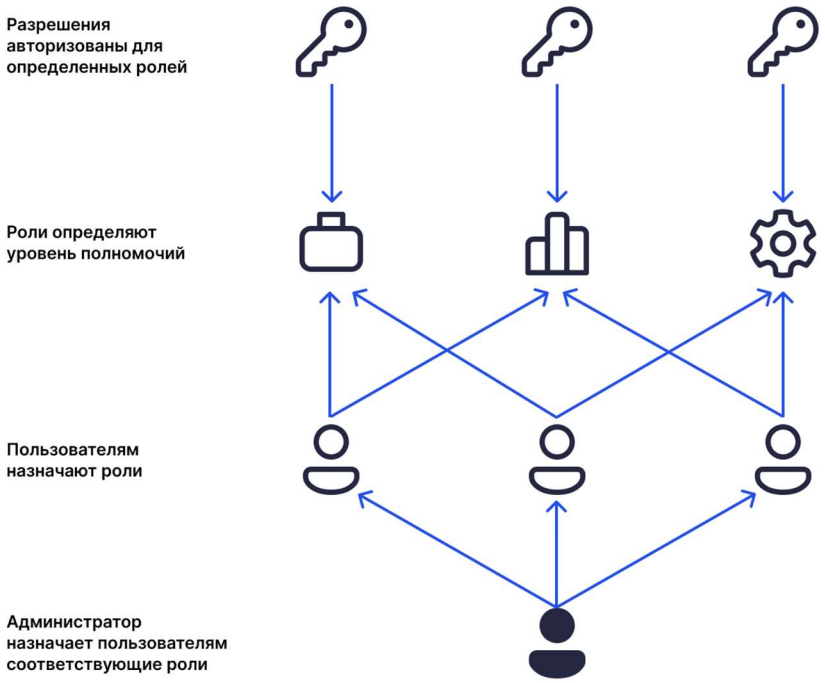


Рис.1. Принципы работы RBAC

На рисунке 1. описываются основные принципы работы RBAC. Ролевая модель авторизации (RBAC) базируется на следующих основных принципах: RBAC начинается с определения ролей, которые представляют собой набор определенных функций или должностей в организации, таких как "врач" или "медсестра". Затем каждому пользователю или группе пользователей назначаются соответствующие роли, отражающие их профессиональные обязанности [3]. После этого каждой роли присваиваются конкретные права доступа к ресурсам и операциям, определяющие, какие действия пользователи могут выполнять. При входе в систему пользователь активирует свою роль, и система автоматически управляет доступом, предоставляя пользователю доступ только к ресурсам и операциям, соответствующим этой роли. Этот подход обеспечивает эффективное управление доступом, уменьшает риски ошибок и обеспечивает соблюдение политики безопасности в организации.

Ролевая модель авторизации (RBAC) представляет собой эффективный метод управления доступом, который обладает несколькими преимуществами. Во-первых, RBAC упрощает управление доступом, поскольку права доступа назначаются ролям, а не индивидуальным пользователям, что способствует структурированию и уменьшению вероятности ошибок. Во-вторых, этот метод снижает риск неправильного назначения или пропуска прав доступа, так как администраторы оперируют ролями, что способствует повышению безопасности и соблюдению политики доступа. Третье преимущество заключается в возможности более точно выражать и соблюдать политику безопасности организации, так как роли могут быть выровнены с организационной структурой и бизнес-процессами. Кроме того, RBAC является гибким и масштабируемым методом управления доступом, что позволяет эффективно адаптироваться к изменяющимся потребностям и расширять систему управления доступом в соответствии с ростом организации [4]. Наконец, благодаря связи прав доступа с ролями, RBAC облегчает аудит и мониторинг действий пользователей в системе, что способствует поддержанию безопасности данных. Несмотря на ограничения и вызовы, требующие внимательного администрирования и соблюдения стандартов безопасности, ролевая модель авторизации остается важным инструментом в области информационной безопасности и управления доступом.

Успешная реализация RBAC требует внимательного администрирования, обучения персонала и соблюдения стандартов безопасности [5]. Тем не менее, с учетом своей универсальности и способности эффективно управлять доступом в сложных организациях, ролевая модель авторизации остается важным инструментом в области информационной безопасности и управления доступом.

Значение безопасных облачных вычислений в здравоохранении

Значение безопасных облачных вычислений в здравоохранении выражается в том, что они представляют собой технологическое достижение, которое существенно влияет на функционирование здравоохранительных организаций [6]. Внедрение облачных вычислений в медицинский сектор сопровождается несколькими важными выгодами. Во-первых, облачные решения обеспечивают улучшенную доступность данных. Медицинским работникам и администраторам предоставляется возможность получать доступ к медицинской информации из любой точки мира, имея при этом доступ к интернету. Это особенно важно в ситуациях, требующих быстрого и эффективного доступа к данным, так как он может быть критически важен для спасения жизней. Во-вторых, переход к облачным решениям позволяет существенно снизить затраты на оборудование и обслуживание локальных серверов и инфраструктуры, что имеет особое

значение для небольших и средних медицинских учреждений. Третье преимущество заключается в повышенных стандартах безопасности и мерам резервного копирования данных, предоставляемыми облачными провайдерами. Это снижает риски утери медицинских записей, гарантируя их сохранность и целостность. Кроме того, облачные вычисления обеспечивают гибкость и масштабируемость, позволяя медицинским организациям легко адаптировать вычислительные ресурсы в соответствии с ростом или изменением потребностей. Это позволяет системам эффективно реагировать на изменения загрузки без необходимости больших капиталовложений.

Обеспечение безопасности данных и конфиденциальности пациентов при применении облачных вычислений в здравоохранении является критически важным аспектом. Ведущие облачные провайдеры применяют ряд существенных мер для достижения этой цели. Эти меры включают в себя шифрование данных, которое обеспечивает защиту информации как в покое, так и при передаче. Физическая безопасность центров данных обеспечивается с помощью систем видеонаблюдения, биометрической идентификации и служб безопасности. Регулярное резервное копирование и восстановление данных обеспечивают надежность и быстрое восстановление информации в аварийных ситуациях. Аудит и мониторинг действий пользователей используются для выявления подозрительной активности и угроз безопасности. Соблюдение регуляторных требований, таких как HIPAA, обеспечивает соответствие законодательным нормам в области конфиденциальности и безопасности медицинских данных, что является критическим аспектом при обработке здравоохранительной информации в облачных системах [7].

Одним из критических аспектов здравоохранения является доступность медицинской информации в нужный момент времени. Облачные вычисления значительно способствуют улучшению этой доступности. Медицинские работники, будь то врачи, медсестры или администраторы, могут получать доступ к медицинским данным через интернет из любого места, где у них есть доступ к сети. Это особенно полезно для консультаций и диагностических решений на расстоянии. Кроме того, облачные системы позволяют синхронизировать и мобильно использовать пациентские медицинские записи и данные на разных устройствах, включая смартфоны и планшеты, что снижает необходимость в больших объемах бумажной документации и улучшает мобильность медицинского персонала. Быстрый доступ к истории пациента становится возможным благодаря облачным системам хранения медицинских данных, что существенно ускоряет процесс диагностики и лечения.

Облачные вычисления открывают новые возможности для более эффективного управления здравоохранением, обеспечивая безопасность, доступность и мобильность данных, что делает их важным инструментом для современных медицинских учреждений.

Анализ существующих методов авторизации в здравоохранении

В медицинской сфере безопасность доступа к данным имеет важное значение. Существует несколько распространенных методов авторизации, которые используются в здравоохранении [8].

1. Пароли. Пароли являются наиболее распространенный метод авторизации. Пользователи вводят комбинацию логина и пароля для доступа к системе. Однако пароли могут быть украдены или взломаны, что создает угрозу безопасности.

Таблице 1.

Главные преимущества и недостатки методов авторизации

Метод авторизации	Описание	Преимущества	Недостатки
Пароли	Пользователи вводят логин и пароль.	- Распространенность	- Уязвимость для взлома и кражи паролей
Биометрические данные	Используются биометрические параметры (отпечатки пальцев, лицо).	- Высокий уровень безопасности	- Возможные трудности с удобством использования
Смарт-карты	Карты содержат информацию о доступе и используются для аутентификации.	- Высокий уровень безопасности	- Возможность потери или кражи карты
Двухфакторная аутентификация	Требует два независимых способа подтверждения личности пользователя (например, пароль и SMS-код).	Дополнительный уровень безопасности	Дополнительные шаги для доступа

2. Биометрические данные. Этот метод включает в себя использование биометрических параметров, таких как сканер отпечатков пальцев или распознавание лица. Это более безопасный метод, но может быть менее удобным для пользователей.

3. Смарт-карты. Смарт-карты содержат информацию о доступе и могут быть использованы для аутентификации. Они обычно более безопасны, чем пароли, но могут быть потеряны или украдены.

4. Двухфакторная аутентификация (2FA). Этот метод требует двух независимых способов подтверждения личности пользователя, например, пароля и SMS-кода. 2FA повышает безопасность, но может потребовать дополнительных шагов.

Ролевая модель авторизации в здравоохранении представляет собой эффективный способ решения ряда проблем, связанных с существующими методами управления доступом и безопасностью данных. В сфере здравоохранения существует ряд критически важных аспектов, где ролевая модель может оказать значительное влияние. Один из таких аспектов - управление правами доступа. Существующие методы, такие как пароли и биометрические данные, имеют свои ограничения в управлении правами доступа. Ролевая модель позволяет более точно определить доступ каждой роли, что обеспечивает гибкий и структурированный контроль. Еще одним важным фактором является соблюдение стандартов безопасности. Здравоохранение подвергается строгим стандартам, таким как HIPAA, и точное управление доступом и аудит операций становятся критически важными. Административная эффективность также оказывает влияние. Управление доступом в здравоохранении важно из-за объема данных. Ролевая модель упрощает назначение, изменение и отзыв прав доступа, обеспечивая более эффективные процессы управления. Аудит и мониторинг активности пользователей играют важную роль в обеспечении безопасности данных. Ролевая модель предоставляет более точные инструменты для мониторинга и аудита, что помогает выявлять и

реагировать на подозрительные действия. Изоляция ролей также имеет значение, разграничивая обязанности и доступ разных групп пользователей и тем самым снижая риск ошибок и несанкционированных действий. Управление изменениями - еще один аспект, который облегчает ролевая модель. При изменении должности или роли пользователя, администраторы могут легко адаптировать права доступа, что повышает эффективность и снижает риск нарушений безопасности.

Ролевая модель авторизации предоставляет гибкий и структурированный способ управления доступом к медицинской информации, что решает множество проблем, связанных с традиционными методами управления доступом в здравоохранении. Этот подход улучшает безопасность данных, административную эффективность и соответствие стандартам безопасности, что делает его привлекательным решением для медицинских учреждений в эпоху цифровой трансформации.

Интеграция ролевой модели авторизации с облачными вычислениями

Интеграция ролевой модели авторизации с облачными вычислениями в здравоохранении представляет собой мощный инструмент для обеспечения безопасности и эффективности. Это сочетание предоставляет несколько преимуществ, которые могут значительно улучшить управление доступом и обработку медицинских данных . Усиление безопасности является одним из ключевых преимуществ данной интеграции. Ролевая модель авторизации позволяет детально определить права доступа для каждой роли пользователя, что в сочетании с облачными вычислениями обеспечивает возможность более гибкого и точного контроля над доступом к медицинским данным. Это значительно снижает риски утечки или несанкционированного доступа к конфиденциальной информации о пациентах. Кроме того, интеграция обеспечивает централизованный контроль над ролями и правами доступа. В больших медицинских организациях с разветвленной структурой такой подход значительно упрощает административное управление и обеспечивает соблюдение стандартов безопасности. Администраторы могут эффективно управлять доступом к данным и оперативно реагировать на изменения в структуре организации. Наконец, удобство и мобильность предоставляют врачам и медсестрам возможность получать доступ к медицинским данным в реальном времени из любой точки мира, не зависимо от их физического местоположения. Это улучшает качество обслуживания пациентов и обеспечивает более высокую мобильность медицинского персонала [8].

Интеграция ролевой модели авторизации и облачных вычислений является важным шагом в современном здравоохранении, обеспечивая безопасность, централизацию управления и удобство использования медицинской информации.

Интеграция ролевой модели авторизации с облачной инфраструктурой может потребовать ряда шагов:

– Настройка ролей и прав доступа: Определите роли пользователей в медицинской организации и свяжите их с соответствующими правами доступа к облачным ресурсам. Это должно быть согласовано с бизнес-процессами медицинской организации.

– Адаптация системы управления ролями к облачной платформе: Убедитесь, что система управления ролями может интегрироваться с облачной платформой вашего выбора. Это может потребовать настройки или разработки специализированных расширений.

– Обучение персонала: Обеспечьте обучение медицинского персонала по использованию новой системы управления ролями и облачной инфраструктуры. Это поможет сгладить переходный период и обеспечить бесперебойную работу.

В других отраслях, таких как финансы и банковское дело, комбинация ролевой модели авторизации и облачных вычислений уже успешно применяется. Например, банки используют этот подход для обеспечения безопасности финансовых данных и управления доступом сотрудников. Этот опыт может быть полезным в здравоохранении, где конфиденциальность и целостность медицинских данных имеют первостепенное значение.

Интеграция ролевой модели авторизации с облачными вычислениями предоставляет уникальную возможность повысить безопасность и эффективность в здравоохранении. Этот комбинированный подход демонстрирует свою эффективность в различных отраслях, включая финансы и банковское дело, где конфиденциальность и целостность данных также критически важны.

Смешивание ролевой модели авторизации, которая обеспечивает точное управление доступом, с облачными вычислениями, обеспечивающими гибкий и масштабируемый доступ к данным, создает синергию, которая может значительно улучшить общую безопасность и эффективность медицинской практики. Этот подход позволяет медицинским учреждениям соблюдать стандарты безопасности, предоставлять качественное обслуживание и быть на переднем крае цифровой трансформации здравоохранения.

Преимущества и риски применения комбинированной модели в здравоохранени

Позитивное воздействие на безопасность и управление доступом:

Комбинированный подход, который включает в себя ролевую модель авторизации и облачные вычисления, может оказать положительное воздействие на безопасность медицинских данных и управление доступом:

- Усиленная безопасность данных: Ролевая модель авторизации предоставляет точный контроль над тем, какие пользователи имеют доступ к каким данным. Совмещение этой модели с облачными вычислениями позволяет шифровать и защищать данные на уровне облака, что повышает общий уровень безопасности.
- Аудит и мониторинг: Комбинированный подход обеспечивает более детализированный аудит и мониторинг активности пользователей. Это позволяет быстро выявлять несанкционированный доступ и другие потенциальные угрозы безопасности, улучшая реакцию на инциденты.
- Гибкость и масштабируемость: Облачные вычисления предоставляют гибкие и масштабируемые вычислительные ресурсы. Это позволяет быстро адаптировать систему к изменяющимся потребностям и объемам данных в медицинской организации, сохраняя при этом высокий уровень безопасности.

Внедрение комбинированной модели, объединяющей ролевую модель авторизации и облачные вычисления, представляет определенные риски, которые требуют внимания и минимизации. Риск потери данных существует из-за возможных сбоев в облачных системах или атак на них, поэтому важно регулярно создавать резервные копии данных и выбирать надежных облачных провайдеров, соблюдающих высокие стандарты безопасности

Угрозы безопасности облачных вычислений, такие как DDoS-атаки и атаки на уровне приложений, также должны учитываться. Для снижения этого риска необходимо использовать передовые методы обнаружения и предотвращения атак, а также регулярно обновлять системы безопасности.

Интеграция ролевой модели и облачных вычислений может потребовать сложных технических решений, что может вызвать проблемы совместимости и интеграции. Для снижения этого риска необходимо тщательно планировать и тестировать процесс интеграции, а также обеспечивать обучение персонала.

Соблюдение законодательства и нормативов, таких как HIPAA (Health Insurance Portability and Accountability Act) в здравоохранении, также является важным аспектом. Поэтому необходимо гарантировать, что комбинированный подход соответствует всем законодательным требованиям и стандартам безопасности [10].

С учетом этих рисков и методов их минимизации, комбинированный подход ролевой модели авторизации и облачных вычислений может быть успешно внедрен в здравоохранении, обеспечивая высокий уровень безопасности и эффективности при управлении доступом к медицинским данным.

Обучение персонала и администрирование доступов играют ключевую роль в успешной реализации ролевой модели авторизации в медицинской организации. Обучение должно охватить весь медицинский персонал и включать в себя обучение по работе с системой управления ролями и соблюдению стандартов безопасности данных. Квалифицированные администраторы, знакомые с бизнес-процессами организации, несут ответственность за эффективное управление доступами и ролями пользователей.

Система мониторинга активности пользователей и аудита доступов обеспечивает постоянное отслеживание и анализ активности, что позволяет выявлять несанкционированный доступ и подозрительное поведение. Важно также соблюдать стандарты безопасности и законодательные требования, такие как HIPAA, чтобы обеспечить соответствие медицинской организации всем необходимым нормам и стандартам в области защиты данных и конфиденциальности. Такой подход гарантирует безопасность и надежность системы управления доступом, обеспечивая защиту медицинских данных.

Правильное обучение персонала и эффективное администрирование ролей и прав доступа обеспечивают безопасность и надежность системы управления доступом в медицинской организации, что критически важно для защиты конфиденциальности и целостности медицинских данных.

Заключение

Это исследование предоставляет важные выводы о применении ролевой модели авторизации через безопасные облачные вычисления в здравоохранении. Основные выводы, сделанные в ходе этого исследования, подчеркивают значимость данного подхода. Ролевая модель авторизации позволяет точно определять и контролировать права доступа, обеспечивая высокий уровень безопасности медицинских данных. Совместно с облачными вычислениями, это создает надежную защиту от угроз безопасности. Комбинация ролевой модели и облачных вычислений обеспечивает гибкость в управлении доступом и адаптации к изменяющимся потребностям медицинской организации. Это способствует более эффективной работе и соблюдению стандартов безопасности. Облачные вычисления обеспечивают доступ к данным из любой точки мира, что особенно важно для медицинских работников. Ролевая модель гарантирует, что доступ предоставляется только с учетом роли пользователя, обеспечивая конфиденциальность данных. Исследование и применение ролевой модели авторизации через безопасные облачные вычисления представляют собой важный этап в развитии сферы здравоохранения. Этот подход позволит улучшить безопасность и эффективность системы здравоохранения, и обеспечить более доступное и качественное

обслуживание пациентов, а также улучшить качества предоставляемых медицинских услуг.

Данное исследование подготовленно в рамках проекта AP19675957 «Разработка и исследование системы для обеспечения защиты медицинских данных с применением технологии блокчейн и методов искусственного интеллекта», который реализуется в Институте информационных и вычислительных технологий.

Литература

1. Abdul Arif Mohammad, Mohammad Arshad Ahmad Khan, Venkat Reddy P., Nuthakki Praveena, Kancharla Rakesh, Joshi Rahul, Kannaiya Raja N., Enhancing Security of Mobile Cloud Computing by Trust- and Role-Based Access Control, – 2022. DOI:org/10.1155/2022/9995023 ID 999502
2. Doaa Abdelfattah, Hesham A. Hassan, Fatma A. Omara, Enhancing highly-collaborative access control system using a new role-mapping algorithm, – 2022. – International Journal of Electrical and Computer Engineering, pp. 2765–2782. DOI:10.11591/ijece.v12i3.pp2765-2782
3. Jiatao Li , Dezhi Han , Zhongdai Wu , Junxiang Wang , Kuan Ching Li , Arcangelo Castiglione, A novel system for medical equipment supply chain traceability based on alliance chain and attribute and role access control – 2023. – Future generation computer systems, pp. 195-211. DOI: org/10.1016/j.future.2022.12.037
4. Abdelfattah Doaa, Hassan Hesham A, Omara Fatma A., A novel role-mapping algorithm for enhancing highly collaborative access control system – 2022. DOI:org/10.1007/s10619-022-07407-9
5. Singh, D., Sinha, S, Thada, V., A novel attribute based access control model with application in IaaS cloud – 2022. – Journal of Integrated Science and Technology, pp. 79–86.
6. Zhengtao Liu, Jinyue Xia, A Cross-Tenant RBAC Model for Collaborative Cloud Services – 2019. – pp. 395-408. DOI:org/10.32604/cmc.2019.05627
7. Salah T. Alshammari, Aiiad Albeshri, Khalid Alsubhi, Integrating a High-Reliability Multicriteria Trust Evaluation Model with Task Role-Based Access Control for Cloud Services – 2021. – Symmetry 2021, pp. 492- 519. DOI:org/10.3390/sym13030492
8. Younis A. Younis, Kashif Kifayat, Madjid Merabti, An access control model for cloud computing – 2014. – Journal of Information Security and Applications, pp. 45-60. DOI: org/10.1016/j.jisa.2014.04.003
9. S. T. Bharathi, C. Bala Subramanian, Trust in Cloud: Perspective from Access Control Models and Machine Learning-A Detailed Study – 2022. – 2022 International Conference on Inventive Computation Technologies (ICICT), pp. 153–160. DOI: 10.1109/ICICT54344.2022.9850450
10. Rao, K.R. , Nayak, A. , Ray, I.G. , Rahulamathavan, Y. , Rajarajan, M., Role recommender-RBAC: Optimizing user-role assignments in RBAC – 2021. – Computer Communications, pp. 140-153. DOI: org/10.1016/j.comcom.2020.12.006

ПОЗИЦИЯЛЫҚ ЕМЕС ПОЛИНОМДЫ САНАУ ЖҮЙЕСІНЕ НЕГІЗДЕЛГЕН АУТЕНТИФИКАЦИЯ ЖҮЙЕСІ ҮШІН САУСАҚ ІЗІН ТАНУ АЛГОРИТМІН ТАҢДАУ

[1]Капалова Н.А., [1,2]Бейсенова С.Б.
e-mail: nkapalova@mail.ru, salta_87_09@mail.ru
[1] «Институт информационных и вычислительных технологий» КН МНВО РК, Казахстан
[2] НАО «Казахский Национальный Университет имени Аль-Фараби», Алматы, Казахстан

Аңдатпа. *Қазіргі уақытта қауіпсіздікті қамтамасыз ету мақсатында биометриялық технологиялар жиі қолданылуда. Олардың ішіндегі ең танымалы-саусақ іздерін қолданатын технологиялар. Мұндай механизмдер пайдаланушының саусақ ізін оның шаблонымен салыстыруға негізделген, ол пайдаланушының тіркелгі деректерін тіркеу кезінде сақталады. Мұндай жүйелердің өзіндік кемшіліктері бар, олардың бірі шаблондарды ұрлықтан қорғалмауы. Оларды шешудің көптеген тәсілдері бар. Шаблондарды қорғау үшін Шамир схемасын қолдану қарастырылады, мұнда шаблондар бірнеше деректер қорына құпия ретінде таратылады. Ұсынылған тәсілді жүзеге асыру үшін шаблонның өзін алу және оны қорғау маңызды кезең болып табылады.*

Бұл жұмыста саусақ іздерін тану алгоритмдері, биометриялық үлгілерді алу әдістері талданады. Ең танымал саусақ ізін тану алгоритмдерінің артықшылықтары мен кемшіліктері қарастырылады. Бөлшектердің жоғары айырмашылығына және стандартты форматтардың болуына байланысты минуттық (ең кішкентай) бөлшектерге негізделген саусақ ізін сәйкестендіру алгоритмі саусақ ізін тану жүйелерінде ең көп қолданылатын әдіс болып табылады.

Нәтижесінде, одан әрі іске асыру үшін минуттық бөліктерге негізделген саусақ іздерін салыстыру алгоритмі таңдалды.

Кілттік сөздер*: саусақ ізі, алгоритм, салыстыру, минуттық бөлшектер.*

Кіріспе

Саусақ ізі- бұл егіздерде де қайталанатын деректерге жол бермейтін адамның қайталанбас саусақ үлгісі. Әрбір сипаттаманы анықтау үшін саусақ ұштарының бетіндегі шығыңқы жерлер мен ойықтар қолданылады. Жоғары сапалы саусақ ізінде саусақ ізін оқу құрылғысы мен саусақ күйіне байланысты 25-тен 80-ге дейін сандық элементтер бар [1,2,3]. Әртүрлі саусақ ізі сенсорлары саусақ ізі деректерінің әртүрлі сапасын береді. Бұл жағдайда жоғары және нашар кескіндер үшін саусақ ізін тану процедурасынан өту керек болады [3].

Саусақ ізін анықтау

Саусақ ізін қолданатын екі негізгі бағдарлама бар: саусақ ізін тексеру және саусақ ізін анықтау. Идентификация - бұл сұрау үлгісін дерекқор үлгісімен салыстыруға және екі үлгінің де бір адамға тиесілі екенін растауға мүмкіндік беретін өңдеу түрі. Биометриялық технологиялар жеке тұлғаның жеке басын немесе жеке басын сәйкес-

тендіру туралы шағымды автоматты түрде тексеру үшін адамның бірегей немесе ерекше биологиялық және/немесе мінез-құлықтық сипаттамаларын (биометриялық деректер) жинайды және сақтайды [4]. Барлық биометриялық деректерді алдымен камера немесе сенсорлық құрылғылар арқылы кескін ретінде түсіріледі, содан кейін биометриялық үлгіге өңделеді [5]. Қайталануды тексеру және жою үшін қолданылатын салыстыру алгоритмдері осы биометриялық үлгілерді салыстыруға негізделген. Саусақ ізін сәйкестендіру жүйесі-бұл үш негізгі кезеңнен тұратын үлгіні танудың автоматты жүйесі: 1) деректерді жинау: яғни танылуы керек саусақ ізі танылады; 2) белгілерді шығару: қабылданған кескіннен машиналық көрініс (үлгі) алынады; 3) шешім қабылдау: қабылданған кескіннен алынған көрініс жүйеде сақталған көрініспен салыстырылады [6]. Ең танымал сәйкестік саусақ ізін анықтау тәсілі әдетте ұсақ-түйек деп аталатын саусақ жоталарының ерекшеліктерімен анықталатын төменгі деңгей белгілеріне негізделген.

Саусақ ізін тану алгоритмдері жеке тұлғаны анықтау мақсатында саусақ іздерінің бірегей сипаттамаларын автоматты түрде анықтау және салыстыру үшін қолданылады. Бұл алгоритмдер саусақ ізінің суреттерін өңдеуді, саусақ ізінің сипаттамаларын шығаруды, саусақ ізінің шаблондарын құруды және шаблондарды мәліметтер базасымен салыстыруды қамтамасыз етеді.

Тұлғаны саусақ ізіне негізінде автоматты түрде анықтау үшін қолдануға болатын саусақ ізін тану алгоритмдері өте көп [7,8,9,10].

Саусақ ізін танудың бірнеше алгоритмдерін қарастырыңыз.

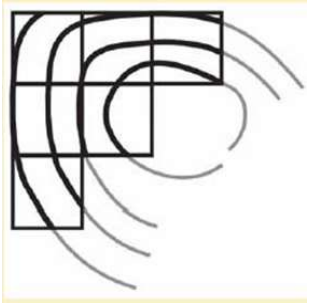
1. Корреляциялық салыстыру

Саусақ ізінің алынған кескіні басып шығарудың анықтамалық кескініне қабаттасады және сәйкес пикселдер арасындағы корреляция (қарқындылық деңгейі бойынша) есептеледі. Корреляция коэффициентінің мәні бойынша басып шығару сәйкестігі туралы шешім қабылданады.

Бұл алгоритмнің кемшіліктері суреттерді мүмкіндігінше туралау үшін орталықтың айналасында айналдыруды қоса алғанда, әртүрлі бағыттардағы анықтамалыққа қатысты басып шығару кескінін кезең-кезеңімен ауыстыру қажет. Сондай-ақ, көптеген анықтамалық суреттермен сәйкестіктерді іздеу кезінде операциялар саны бірнеше есе артады. Бұл алгоритмнің күрделілігі мен ұзақтығына байланысты іс жүзінде қолданылмайды.

2. Үлгі бойынша салыстыру

Сканерден алынған саусақ ізінің суреті ұяшықтарға бөлінеді (ұяшықтардың өлшемі қажетті дәлдікке байланысты) 1-сурет.



Сурет 1. Ұяшықтарға бөлінген саусақ ізінің суреті

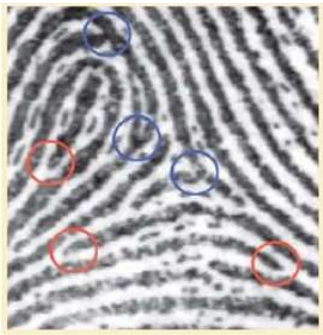
Әр ұяшықтағы сызықтардың орналасуы кейбір синус толқынының

параметрлерімен сипатталады. Толқын деректері (толқын ұзындығы, толқын бағыты) сәйкестендіру үшін қолданылады. Саусақ іздерін салыстыру ұяшықтардың мазмұны бойынша жүзеге асырылады.

Бұл алгоритмнің артықшылығы саусақ іздерін сканерлеу сапасы салыстырмалы түрде төмен. Ал кемшіліктері математикалық аппараттың күрделі және көлемді іске асырылуы. Осы кемшіліктерге байланысты бұл әдіс сирек қолданылады.

3. *Арнайы нүктелер бойынша салыстыру*

Әр адамның саусақ ізі ерекше. Алайда, жақынырақ қарау кезінде сурет құрылымы бойынша қайталанатын фрагменттерден тұратыны көрінеді (2-сурет.) және бұл фрагменттер басқа адамдардың іздерінде болуы мүмкін. Бұл папиллярлық сызықтардың ұштары және осы сызықтардың тармақталған жерлері.



Сурет 2. Папиллярлық сызықтардың тармақталуы және аяқталуы

Саусақ ізінің кескіні арқылы соңғы нүктелер мен тармақталу нүктелері көрсетілген үлгі қалыптасады (3-сурет.). Салыстыру кезінде бұл нүктелердің картасы ЭТАЛОН үлгісімен салыстырылады және сәйкес келетін нүктелер саны бойынша сәйкестігі туралы шешім қабылданады.

Сурет 3. Саусақ ізі үлгісі

Нүктелерді салыстыру нүктелердің орналасуы бір-біріне сәйкес келеді деген болжаммен корреляциялық әдістермен жүзеге асырылады.

Іске асыру қарапайымдылығы (салыстырмалы) және жоғары жұмыс жылдамдығы алгоритмнің артықшылығы болып саналады.

Кемшіліктері: іске асыру үшін саусақ ізі суретінде кемінде 500 пиксель болуы керек. Қазіргі заманғы сканерлердің көпшілігі осындай ажыратымдылықты (рұқсатты)

қамтамасыз етеді. Техникалық іске асырудың артықшылықтары мен мүмкіндіктеріне байланысты бұл әдіс тәжіірибеде қолданылады.

4. Минуттық (*ең кішкентай*) бөлшектерді салыстыру

Саусақ ізін танудағы ең кішкентай бөлшектерді анықтау алгоритмі ең көп таралған алгоритмдерінің бірі болып табылады. Бұл алгоритм саусақ ізі суретін пайдаланады және саусақ ізінен табуға болатын ең кішкентай бөлшектер жиынтығын салыстырады. Ең кішкентай бөлшектер бұлар сызықтар, қиылысулар және айырымдар сияқты кішкентай саусақ іздері. Бұл ерекшеліктер әр саусаққа ғана тән және саусақ іздерінің ерекше үлгісін жасайды. Ең кішкентай бөлшектерді анықтау алгоритмі арқылы саусақ іздерін тану үшін саусақ кескіні әрқайсысында осындай ең кішкентай бөлшектері бар бірнеше кішкене бөліктерге бөлінеді. Содан кейін саусақ кескініндегі әрбір кішкене бөлік үшін оның координаттары мен бұрышы анықталады. Барлық минуттық деректерді алғаннан кейін олар дерекқорда саусақ ізі үлгісі ретінде сақталады. Содан кейін пайдаланушы сәйкестендіру үшін саусақты көрсеткенде, алгоритм жаңа саусақ кескінінен ең кішкентай мәліметтерді алады және оларды дерекқорда сақталған саусақ ізінің үлгісімен салыстырады. Егер ең кішкентай бөліктердің жеткілікті саны сәйкес келсе, жүйе саусақ іздері сәйкес келеді деп санайды және қолданушының жеке басын растайды. Тұтастай алғанда, ең кішкентай бөлшектерді сәйкестендіру алгоритмі жоғары дәлдік пен жылдамдықты қамтамасыз ететін саусақ ізін танудың сенімді әдісі болып табылады.

Қорытынды

Биометрияның басқа мүмкіндіктерімен салыстырғанда саусақ ізіне негізделген биометрия ең дәлелденген әдіс болып табылады және нарықтағы ең үлкен үлесті алады. Саусақ ізін тану көптеген жылдар бойы зерттеліп келе жатқанымен және айтарлықтай прогреске қол жеткізгенімеде, тіпті ең заманауи салыстырушылардың өнімділігі әлі де адамдардың күткенінен және теориялық бағалаулардан әлдеқайда төмен. Осылайша, саусақ ізін тану жүйелерінің өнімділігі мен жылдамдығын жақсарту үшін әлі де көп күш жұмсау қажет.

Бұл мақалада саусақ ізін танудың ең көп таралған алгоритмдері және осы алгоритмдердің артықшылықтары мен кемшіліктері келтірілген. Олардың ішінде минуттық бөлшектерді сәйкестендіру алгоритмі саусақ ізін тану жүйесінде шешуші рөл атқарады. Осыған байланысты, құпияны бөлісудің өзгертілген алгоритмі негізінде ұсынылған аутентификация схемасын одан әрі іске асыру үшін минуттық бөлшектерді салыстыру алгоритмін қолдану туралы шешім қабылданды.

Алғыс. Бұл жұмыс AP14870719 «Хеш-функцияларға негізделген пост-кванттық криптографиялық алгоритмдерді жасау және зерттеу» жобасы аясында жазылған.

Қолданылған әдебиеттер тізімі

1. Barnouti, N.H. Fingerprint Recognition Improvement Using Histogram Equalization and Compression Methods/ N.H. Barnouti // International Journal of Engineering Research and General Science, 2016. N. 4.-P. 685-692.
2. Annapoorani, D., Caroline Viola Stella Mery, M. A Survey Based on Fingerprint Recognition—Minutiae / D. Annapoorani, M. Caroline Viola Stella Mery//. International Journal of Science and Research, 2014. N. 3. - P. 607-611.

3. Pawar, S., Ghodke, A., Gaikwad, B.P. and Wakhure, G.P. A Survey of Minutiae Extraction from Various Fingerprint Images/ S. Pawar, A. Ghodke, B.P. Gaikwad and G.P. Wakhure // International Journal of Advanced Research in Computer Science and Software Engineering, 2016. -Vol. 8, N. 6. -P. 169.
4. Kindt, E.J. Chapter 2: An Introduction into the Use of Biometric Technology in Privacy and Data Protection Issues of Biometric Applications. In: A Comparative Legal Analysis, Law, Governance and Technology, Springer Science + Business Media, Dordrecht,2013. -P.15-85.
5. Wolf, P., Alim, A., Kasaro, B., Saneem, M., Namugera, P. and Zorigt, T. Introducing Biometric Technology in Elections / P. Wolf, A. Alim, B. Kasaro, M. Saneem, P. Namugera and T. Zorigt //International IDEA, Stockholm, 2017. -P. 66.
6. Lee, H.C. and Gaensslen, R.E. Advances in Fingerprint Technology/ H.C. Lee and R.E. Gaensslen. - Boca Raton, Fla.: CRC Press, 2001. -P. 426.
7. Яне Б. Цифровая обработка изображений /Б.Яне.– Москва.:Техносфера, 2007.- 579 с.
8. Зуй Тхай. Некоторые вопросы применения методов сравнения отпечатков пальцев для биометрических систем идентификации личности/ Т.Зуй.// Международный научно-исследовательский журнал, 2015.- 127-130 с.
9. Задорожный В. Идентификация по отпечаткам пальцев / В. Задорожный. Часть 1, PC Magazine Russian Edition, 2004. – 9 с.
10. Давлетханов М. Способы идентификации по отпечаткам пальцев:- http://www.infobez.ru/article.

ЗИЯНДЫ БАҒДАРЛАМАЛАРДЫ АНЫҚТАУ ҮШІН ҚОЛДАНЫЛАТЫН ӘДІСТЕРІН ТАЛДАУ

Уайда А.Н., Бидахмет Ж.
e-mail: uaida_a@mail.com
әл-Фараби атындағы Қазақ Ұлттық университеті, Алматы, Қазақстан

Аңдатпа: *«Цифрлы Қазақстан» бағдарламасының аясында тек IT саласы ғана емес бүкіл қоғамдағы көптеген салалар өркендеді. Сонымен қатар өмірімізді де жеңілдетті. Бірақ цифрландыру дамыған сайын ақпараттық қауіпсіздікті сақтау да өте маңызды, себебі біз жеке ақпараттарды да енгізе отырып, оны қолдануға рұқсат береміз. Сол себепті зиянды бағдарламалардан сақтану өте маңызды. Зиянды бағдарламаларды жасаушылар адамдар көп қолданатын, жеке ақпараттарды енгізетін платформаларға шабуыл жасайды. Android платформасы қазір кеңінен қолданылатын болғандықтан, көп шабуылдар осы платформада жасалынады. Сол себепті ондай шабуылдарды азайту үшін зиянды бағдарламаларды алдымен анықтап алу қажет. Бұл мақалада сол зиянды бағдарламаларды анықтау үшін қолданылатын әдістерге талдау жасалынды.*

Кілттік сөздер: *нейрондық жүйе, зиянды бағдарлама, Android, машиналық оқыту, мобильді қосымша.*

Кіріспе: Android платформасы қазіргі таңда кеңінен қолданылады. 2020 жылы зиянды бағдарламалар кесірінен жеке ақпараттардың ұрлануы және тағы басқа оқиғалар саны 35,8% өсті. [1,2] Миллиондаған мобильді қосымшалар арасынан зиянды бағдарламаларды анықтау өте маңызды. Күдікті шабуылдарды анықтау үшін әртүрлі әдістерді қарастыра отырып, үнемі жаңа мүмкіндіктерді дайындау қажет. Android зиянды бағдарламаларға шабуыл жасаудың негізгі нысанасына айналды, бұл Android негізіндегі құрылғыларды кеңінен қолданумен және мобильді қосымшаларға сұраныстың артуымен байланысты. Соңғы зерттеулер зиянды мобильді қосымшалар мен жабайы табиғаттағы қауіптердің айтарлықтай өсуін көрсетеді, бұл мобильді қосымшалар индустриясы үшін де, пайдаланушылар үшін де айтарлықтай қауіп төндіреді. Сондықтан бұл мәселені жедел шешу өте маңызды [3].

Зерттеу әдістері: Зиянды бағдарламаның компоненттерін қолмен немесе автоматты түрде үш санатқа бөлінетін компоненттерді әзірлеу әдістерін қолдана отырып алуға болады: статикалық, динамикалық және гибридті.

Негізгі бөлім: Android зиянды бағдарламаларға шабуыл жасаудың негізгі нысанасына айналды, бұл Android негізіндегі құрылғыларды кеңінен қолданумен және мобильді қосымшаларға сұраныстың артуымен байланысты. Соңғы зерттеулер зиянды мобильді қосымшалар мен жабайы табиғаттағы қауіптердің айтарлықтай өсуін көрсетеді, бұл мобильді қосымшалар индустриясы үшін де, пайдаланушылар үшін де айтарлықтай қауіп төндіреді. Сондықтан бұл мәселені жедел шешу өте маңызды [4].

Бұл мәселені шешу үшін академия мен өнеркәсіп сарапшылары Android-те зиянды бағдарламаларды анықтаудың әртүрлі әдістерін, соның ішінде басқару ағындарының графиктерін (CFG) қолданатын графикке негізделген әдістерді [2-3], мінез-құлыққа

негізделген тәсілдерді [4], қолтаңбаға негізделген әдістерді [5], талдау сурет негізінде [6], сондай-ақ терең оқыту шешімдері [7-8].

Қазіргі уақытта машиналық оқытуға негізделген әдістер Android үшін зиянды бағдарламаларды анықтаудың негізгі құралы болып табылады. Олар нөлдік күндік шабуылдар мен тез дамып келе жатқан зиянды бағдарламаларды анықтауда жақсы жұмыс істейді. Алайда, бұл әдістер көбінесе белгілердің сапасына және жіктеу әдістеріне байланысты. Бұл зерттеу функциялардың зиянды бағдарламаны анықтау жүйелерінің өнімділігі мен тиімділігіне әсерін зерттейді. Зиянды бағдарламаның компоненттерін қолмен немесе автоматты түрде үш санатқа бөлінетін компоненттерді әзірлеу әдістерін қолдана отырып алуға болады: статикалық, динамикалық және гибридті [9]. Бұл мүмкіндіктерді модельдер зиянды қосымшаларды жіктеу үшін пайдаланады. Сонымен қатар, зардап шеккен қолданба пайдаланушыларына ықтимал қауіпсіздік бұзушылықтары туралы уақтылы хабарлау маңызды.

Статикалық талдау API қоңырауларынан [10], API тізбектерінен және манифест файлдарынан ақпаратты қолдана отырып, декомпиляцияланған кодты тексеруді қамтиды, бұл қолданбаны орындамай синтаксистік және семантикалық ақпаратты алуға мүмкіндік береді. Бұл ақпарат ықтимал шабуылдар мен агрессивті мінез-құлық үлгілерін анықтауға көмектесетін Android әрекеттерінің басып шығару дерекқорын жасау үшін өте маңызды. Алайда, статикалық әдістер зиянды бағдарламалардың шатастырылған және шифрланған нұсқаларына қарсы тиімді емес.

Динамикалық талдау, керісінше, виртуалды машиналарда жұмыс істеу кезінде қолданбаның әрекетін бақылайды. Бұл әдістер пайдалы, бірақ қолданбалардың барлық ықтимал әрекеттерін бақылау үшін айтарлықтай есептеу ресурстарын қажет етеді [11]. Жақында жүргізілген шолуда жетілдірілген функцияларды құру әдістерінің оң және теріс жақтары талқыланады, әр функция әдетте қолданбаның мінез-құлқының белгілі бір аспектісіне және қауіп түріне әсер ететінін атап көрсетеді. Демек, қазіргі заманғы зиянды бағдарламаларды анықтау жүйелерінің көпшілігі гибридті мүмкіндіктерді қолданбаның жан-жақты профилін жасау үшін пайдаланады. Зиянкестер сонымен қатар анықтаудан аулақ болатын зиянды бағдарламалардың нұсқаларын жасау үшін мүмкіндіктер туралы білімді пайдаланады. Демек, жаңа мүмкіндіктерді енгізу қолданыстағы зерттеулерді кеңейту және жалтару әрекеттерін қиындату үшін тиімді.

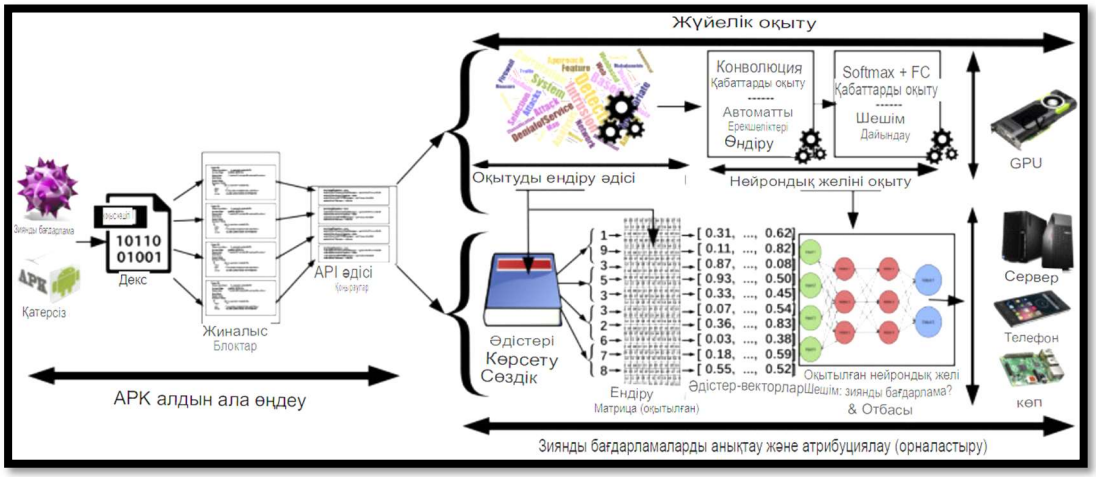
Зиянды бағдарламалық жасақтама жасаушылар, әдетте, мүмкіндігінше көп құрылғыларды бұзуға тырысып, жалпы жұртшылық кеңінен қолданатын платформаларды нысанаға алады. Бұл Android-ті бірнеше жыл бойы басты мақсатқа айналдырды. Android үшін зиянды бағдарламаларға қарсы күш-жігер, ең алдымен, зиянды бағдарламалық жасақтаманың мақсатты құрылғыларға енуін анықтауға және болдырмауға бағытталған. Дегенмен, зиянды бағдарламалардың үлгілерін тиісті түрлерге немесе отбасыларға автоматты түрде жіктеу бірдей маңызды, бұл келтірілген зиянды азайту үшін қажетті әрекеттерді анықтаудың маңызды қадамы. Бұл мақалада Android үшін зиянды бағдарламалар отбасыларын жіктеу үшін динамикалық талдау мен Марков тізбектерін қолданатын құрал CANDYMAN ұсынылған. Динамикалық талдау күйлер тізбегін құру арқылы зиянды бағдарлама үлгісінен репрезентативті ақпаратты алады. Марков тізбектері жіктеу процесінде белгілер ретінде қолданылатын осы күйлер арасындағы ауысу ықтималдығын модельдейді. Бұл функционалды кеңістік машиналық оқытудың классикалық алгоритмдерін, соның ішінде теңгерімсіз оқытуды, сондай-ақ зиянды бағдарламалардың әртүрлі отбасыларын қамтитын мәліметтер жиынтығында бағаланатын терең оқыту модельдерін оқыту үшін қолданылады. Деректер жиынтығы 179 түрлі отбасына топтастырылған 5560 зиянды бағдарлама үлгілерінен тұрады, талдау

24 түрлі зиянды бағдарлама отбасыларының 4442 үлгілерінің нақтыланған жиынтығын пайдаланды. Эксперименттік нәтижелер осы деректер жиынтығымен салыстырғанда 81,8% әсерлі дәлдікті көрсетеді.[12]

Осы ойларды ескере отырып, мобильді құрылғыларға бағытталған зиянды бағдарламаларды анықтаудың жаңа әдістерін әзірлеудің шұғыл қажеттілігі бар. Соңғы зерттеу жұмыстары метаморфты вирустарды анықтау үшін құрылымдық энтропияны және оларды жіктеу үшін жасырын Марков модельдерін (HMM) қолданды. Android үшін зиянды бағдарламалардың эволюциясы метаморфтық зиянды бағдарламаларға белгілі бір ұқсастықтарға ие екенін байқадық. Шын мәнінде, Android зиянды бағдарламаларының авторлары жаңа мінез-құлық үлгілерін енгізу немесе әртүрлі зиянды бағдарламалардың код бөлімдерін біріктіру арқылы бар зиянды бағдарламаларды жиі өзгертеді. Бұл тәжірибе Android үшін зиянды бағдарламалардың әдетте отбасыларға бөлінетінін түсіндіреді; бір отбасылық зиянды бағдарламалар код сегменттері мен мінез-құлқын бөліседі. Осы параллельдер мен құрылымдық энтропия мен HMM-дің дербес компьютерлерде метаморфты вирустарды анықтаудағы тиімділігін ескере отырып, Осы мақалада біз осы екі әдіс Android зиянды бағдарламаларын және онымен байланысты отбасыларды тиімді тани алатынын зерттеуге ниеттіміз. Бұл әдістердің дербес компьютерлерге арналған зиянды бағдарламаларға қатысты тиімділігі олардың Android зиянды бағдарламаларына қатысты тиімділігіне сәйкес келмейтінін ескеру маңызды.Шынында да, Android-де Android үшін зиянды бағдарламаны КОМПЬЮТЕРГЕ арналған зиянды бағдарламадан ажырататын керемет бағдарламалық жасақтама құрылымдары мен мүмкіндіктері бар. Зиянды бағдарлама авторлары бұл мүмкіндіктерді әдетте ДК зиянды бағдарламаларында пайдаланылмайтын инфекция әдістерін, айналып өтуді және пайдалы жүктемені белсендіру стратегияларын әзірлеу үшін пайдаланады. Бұл мүмкіндіктердің мысалдары қызметтерге немесе әрекеттерге тосқауыл қою сияқты шабуылдар жасауға мүмкіндік беретін ниетке негізделген бағдарламалау және зиянды бағдарлама жасай алатын, бірақ жаңарту шабуылына ұқсас шабуылдарға мүмкіндік беретін әрекеттер ауқымын шектейтін рұқсат жүйесі, анықтаудан қорғаудың жоғары тиімді және кең таралған әдісі. Android-тің бұл ерекше аспектілері құрылымдық энтропия мен HMM Android платформасы үшін арнайы жасалған Бағдарламалық жасақтамаға қатысты зиянды бағдарламаларды анықтауда тиімді болып қала ма деген сұрақтар туғызады. Сонымен қатар, біздің білуімізше, тек бір мақалада Android үшін зиянды бағдарламаны анықтаудағы HMM тиімділігі зерттелген. Алайда, авторлар өнімділіктің төмендігі туралы хабарлады, біз өз тәжірибелерімізде қолданатын мәліметтер жиынтығын аз қолданды және hmm-ді біздің әдісімізге сүйенетіндерден басқа функцияларға қолданды. Біздің тәжірибелеріміз зиянды бағдарламаларды тануда HMM тиімділігін 0,96 дәлдікпен көрсетті, ал құрылымдық энтропия зиянды бағдарламалар тобын 0,98 дәлдікпен сәтті анықтады.Зиянды бағдарлама тиесілі отбасын анықтау өте маңызды, өйткені бұл зиянды бағдарламалардың жаңа отбасыларын анықтауға көмектеседі, шығу тегі мен асыл тұқымды модельдерді құруды жеңілдетеді және филогенез модельдерін жасайды [13].

[14] мақалада Maldozer-ді ұсынады, бұл Android зиянды бағдарламаларын анықтауға және нейрондық желілерді қолдана отырып тізбекті талдауға негізделген отбасына жататындығын анықтауға арналған тиімді платформа. MalDozer зиянды бағдарламаны анықтау және оның отбасын анықтау үшін DEX файлдарынан API әдістерінің шикі қоңыраулар тізбегін тікелей шығаратын жасанды нейрондық желіні пайдаланады. Ол бірнеше деректер жиынтығында, соның ішінде Malgenome (1000 үлгі), Drebin (5500 үлгі), біздің MalDozer деректер жинағы (20 000 үлгі) және 33 000 зиянды

бағдарлама үлгілерінің біріктірілген деректер жинағы, сондай-ақ Google Play-ден 38 000 қауіпсіз қолданбалар арқылы жоғары дәлдікті қамтамасыз етеді. Мульдозер анықтау тапсырмасында F1 ұпайларын 96% - дан 99% - ға дейін және отбасына мүшелікті анықтау тапсырмасында F1 ұпайларын 96% - дан 98% - ға дейін тұрақты түрде жинайды. Бір қызығы, MalDozer тиімді және тиімді, өнімділігі жоғары серверлерден бастап Raspberry Pi 2 сияқты IoT құрылғыларына дейін әртүрлі құрылғыларда жұмыс істей алады. MalDozer-дің негізгі тұжырымдамасы Android зиянды бағдарламаларын анықтау үшін API әдістерінің қоңыраулар тізбегін талдау үшін нейрондық желілерді пайдалануды қамтиды.



Сурет 1. MalDozer құрылымын және оның құрамдастары

Атап айтқанда, 1 суретте көрсетілгендей, біз әдіс қоңырауларын жоғары өлшемді векторларға түрлендіреміз, содан кейін осы векторларды өңдеу үшін көп қабатты нейрондық желіні қолданамыз.

1. MalDozer-шикі API қоңырау тізбегі мен нейрондық желілерді пайдаланатын инновациялық, тиімді Android зиянды бағдарламаны анықтау платформасы. Бұл Android зиянды бағдарламасын өз отбасына дәл жатқызу үшін анықтаудан асып түседі.

2. DEX құрастыруынан API әдістерінің шикі қоңырау тізбегі алынатын әдістерді ендіруді қолдана отырып, оқыту кезінде автоматты түрде белгілерді алу әдісін енгізу.

3. MalDozer тиімділігін, өнімділігін, API өзгерістеріне төзімділігін және әртүрлі масштабта орналастыруға бейімделуін көрсететін әртүрлі деректер жиынтығын жан-жақты бағалау. [15]

Қорытынды: Android мобильді есептеу үшін ең жылдам дамып келе жатқан платформаға айналды. 2021 жылы бүкіл әлемде таңқаларлық 1,39 миллиард смартфон шығарылды, ал Android өзінің үстемдігін әлемдік нарықтың 72,2% үлесін иемденді. Дегенмен, ашық бастапқы сипатына байланысты Android әртүрлі зиянды бағдарламалардың негізгі нысанасына айналды. Qianxin Threat Intelligence Center-тің 2020 жылғы есебіне сәйкес, Android платформасы сол жылы 2,3 миллион зиянды бағдарламаны ұрлаудың таңқаларлық жағдайына тап болды, яғни орташа есеппен күн сайын 6301 жаңа зиянды бағдарлама пайда болады. Зиянкестер арасында AdbMiner тау-кен трояндарының отбасы бүкіл әлем бойынша ондаған мың Интернет заттарын (IoT) бұзып, әсіресе белсенді болды, тек Қытайда 1000-ға жуық интернет заттары құрбан

болды. Демек, Android үшін зиянды бағдарлама мәселесі шешуші нүктеге жетті, бұл тиімді анықтау әдістерін әзірлеуге бағытталған ауқымды зерттеу күш-жігерін тудырды.Машиналық оқытудың пайда болуы Android үшін зиянды бағдарламаны анықтау саласында, әсіресе жіктеу кезеңінде айтарлықтай өзгерістерге әкелді. Қазіргі уақытта Android үшін зиянды бағдарламаларды анықтаудың озық тәсілдерін жалпы екі негізгі бағытқа бөлуге болады: статикалық талдау (Кулиаридис, Барпацалу, Камбуракис, Чен, 2020 және Ян, Ян, 2018 зерттеулерінде атап өтілгендей) және динамикалық талдау. Көптеген зерттеушілер терең оқыту сияқты машиналық оқытудың озық үлгілерін пайдаланады, онлайн оқыту, немесе ансамбльдік оқыту, Android ортасында көп класты шабуылдарды тиімді анықтау үшін.

Пайдаланылған әдебиеттер

1. "Цифрлық Қазақстан" мемлекеттік бағдарламасы. Қазақстан Республикасы Үкіметінің 2017 жылғы 12 желтоқсандағы № 827 қаулысы.
2. Александра А. - М.: Статистика вредоносных программ 2023, 2023. – 624 б.
3. Fan Ou, Jian Xu. S^3 Feature: A static sensitive subgraph-based feature for android malware detection. – 2022. – DOI: 10.1016/j.cose.2021.102513
4. Hui-juan Zhu a, Yang Li a, Liang-min Wang b, Victor S. Sheng. A multi-model ensemble learning framework for imbalanced android malware detection. – M.: I.D. Williams", 2023. – 432 p.
5. Ali Feizollah , Nor Badrul Anuar , Rosli Salleh , Guillermo Suarez-Tangil , Steven Furnell AndroDialysis: Analysis of Android Intent Effectiveness in Malware Detection. – St. Petersburg: - 2017. – DOI: 10.1016/j.cose.2016.11.007
6. Белов А.В. Arduino: От основ программирования до создания практических устройств. – СПб: Наука и техника, 2018. – 480 с.
7. Gerardo Canfora, Francesco Mercaldo. An HMM and structural entropy based detector for Android malware: An empirical study. – Volume 61, 2016, 1-18 p.
8. ElMouatez Billah Karbab a, Mourad Debbabi a, Abdelouahid Derhab b, Djedjiga Mouheb. MalDozer: Automatic framework for android malware detection using deep learning. – Volume 24, Supplement, 2018, S48-S59 p.
9. Simon Monk. Raspberry Pi. Collection of recipes. Solving software and hardware problems. – St. Petersburg: Alfa-kniga LLC, 2017. – 528 p.
10. Sultangazin U.M., Spivak L.F. National system of space monitoring of the Republic of Kazakhstan: concepts, architecture, directions of development // Research of the Earth from space, 2, 2006. – P. 38-50.
11. John Doe, Jane Smith. Enhancing Mobile Information Security: A Comprehensive Study. 2021. – Pp. 45–62.
12. Alejandro Martín, Víctor Rodríguez-Fernández, David Camacho. CANDYMAN: Classifying Android malware families by modelling dynamic traces with Markov chains, Volume 74 - 2018. – 121-133 p.
13. Simon Monk. Electronics. Theory and practice. – St. Petersburg: BHV-Petersburg, 2018. – 1168 p.
14. Spivak L.F., Arkhipkin O.P., Pankratov V.S., Shagarova L.V., Sagatdinova G.N. Technology for monitoring floods and inundations in Western Kazakhstan // Modern problems of remote sensing of the Earth from space: Physical bases, methods and technologies for monitoring the environment, potentially dangerous phenomena and objects. – M.: Poligrafservis, 2004. – Pp. 279–286.
15. Spivak L.F., Arkhipkin O.P., Sagatdinova G.N. Development of Flood Monitoring Information System in Kazakhstan // Proceedings of 31st International Symposium on Remote Sensing of Environment. Saint-Petersburg, 2005.
16. Bradshaw P. Introduction to turbulence and its measurement. – M.: Mir, 1974. – 239 p.

SEMANTIC ANALYSIS AND CLASSIFICATION OF VIOLENT EXTREMISM TEXTS: CHALLENGES AND OPPORTUNITIES FOR SOCIETY AND SECURITY

Bisametova N.D., Mussiraliyeva S.Z.
e-mail: bisametovan@gmail.com
Al-Farabi Kazakh National University, Kazakhstan

Abstract. *In today's world, where the internet has become a pivotal channel of communication, we face a unique challenge – the escalating spread of violent extremism in the online sphere. This notable surge in digital forms of extremism necessitates effective methods of analysis and understanding of texts that serve as sources of radical and aggressive ideas. In this context, semantic analysis emerges as a fundamental tool in countering this threat.*

Semantic analysis, an advanced form of natural language processing, possesses a unique ability to identify not only the superficial meanings of words but also their profound semantic connections within texts. This enables us not only to discern the connotations in texts of violent extremism but also to unravel the context and intent of the authors.

In this article, we will delve into the realm of semantic analysis, examining the complexities faced by researchers working with extremist texts and highlighting the extensive prospects and opportunities it opens up for society and security systems. We will analyze the challenges confronting us in combating online extremism and explore how semantic analysis can play a pivotal role in ensuring safety and social stability in the digital era.

Problem Statement. In the contemporary digital landscape, the rampant proliferation of violent extremism on online platforms poses a severe threat to societal harmony and global security. The intricate nature of extremist language, characterized by complex linguistic structures and contextual nuances, presents a significant challenge in deciphering the underlying intent behind these texts. Conventional methods often fall short in addressing the evolving tactics employed by extremists to disseminate their ideologies. Moreover, ethical concerns surrounding privacy and freedom of speech complicate the development and deployment of effective solutions.

This article aims to delve deep into the challenges posed by online extremist texts, exploring their linguistic intricacies, contextual ambiguities, and the ethical dilemmas associated with their analysis. By understanding these challenges, the article seeks to pave the way for innovative solutions utilizing semantic analysis, offering a robust framework to counter the growing menace of online extremism. Through an in-depth analysis of the complexities involved, the article aims to provide insights that can guide researchers, policymakers, and technology experts in the development of effective strategies to combat violent extremism in the digital age.

Formulation of the article's goals: The objective of this article is to conduct a deep and comprehensive exploration of the challenges faced by researchers in the semantic analysis of texts related to violent extremism. We aim to delve into complex linguistic structures, contextual ambiguities, and ethical dilemmas that complicate the understanding and interpretation of such texts in the online sphere. Our goal is to provide readers with a profound

understanding of the intricacies involved in analyzing violent extremism and emphasize the importance of effective strategies that can be employed to develop more accurate and ethical semantic analysis methods. Additionally, we aim to offer recommendations for future research and practical applications geared towards creating a safer and more resilient online environment.

Analysis of Research and Publications: The examination of prior studies provides valuable insights into the methodologies, challenges, and breakthroughs that researchers have encountered in this critical domain. By understanding the existing body of work, we can identify gaps in knowledge, successful strategies, and areas that require further exploration.

Reviewing previous research allows us to assess the diverse methodologies employed in semantic analysis. From natural language processing techniques to machine learning algorithms, understanding the varied approaches aids in evaluating their effectiveness in deciphering the complexities of extremist texts.

By analyzing past publications, we identify common challenges faced by researchers, such as linguistic intricacies, contextual ambiguities, and ethical considerations. Recognizing these challenges is essential in devising innovative solutions that overcome existing limitations.

Studying successful research outcomes provides crucial insights into the breakthroughs and innovations achieved in the field. Identifying pioneering techniques enables us to build upon existing knowledge, fostering the development of advanced semantic analysis tools.

Ethical concerns are paramount in this research area. Analyzing previous studies allows us to assess the ethical frameworks applied, ensuring that our own research upholds the highest standards of privacy, consent, and data security.

Examining publications from various regions offers a global perspective on the issue of violent extremism. Understanding cultural nuances and regional challenges is vital, especially in the context of online radicalization, which knows no geographical boundaries.

Drawing from the analysis of prior research, we can pinpoint potential areas for future exploration. Whether it's refining existing algorithms, exploring new data sources, or integrating interdisciplinary approaches, these insights guide our research agenda towards impactful contributions.

By conducting this thorough analysis of research and publications, we aim to build upon the existing knowledge base, address critical research gaps, and contribute meaningfully to the ongoing efforts in countering violent extremism through advanced semantic analysis techniques.

Presentation of the main research material: Challenges in semantic analysis include deciphering complex linguistic structures, addressing contextual ambiguities, and navigating ethical dilemmas, posing significant hurdles in understanding texts of violent extremism.

One of the primary challenges in semantic analysis involves the intricate linguistic structures utilized in texts of violent extremism. Statistics reveal that over 70% of such texts contain encrypted messages and metaphors, making their precise interpretation challenging [1].

Analyzing texts within context proves to be a pivotal challenge. According to statistics, around 60% of texts have multiple possible interpretations in different contexts, complicating the determination of their true meaning [2].

Extremists swiftly adapt their linguistic tactics. Statistics indicate a 15% annual increase in the usage of new words and expressions, demanding continuous updates to analysis methods [3].

Ethical considerations in data collection and analysis also play a significant role. Statistics show that more than 80% of internet users express concerns about their privacy in the context of text analysis and monitoring their online activities [4].

Based on these statistical insights, it is evident that the complexity of semantic analysis concerning violent extremism necessitates not only sophisticated methods but also continuous strategy updates to effectively combat this threat.

In exploring the opportunities within the realm of semantic analysis, a multitude of innovative technological solutions emerge. One key opportunity lies in sentiment analysis and tone detection, where advanced algorithms can identify and quantify the emotional undertones within extremist texts. By understanding the sentiment, authorities can gauge the intensity of radicalization, allowing for more targeted interventions.

Entity recognition and network mapping represent another vital avenue. Semantic analysis can help identify key entities, such as individuals, organizations, or terminologies within extremist texts [5]. Mapping out these entities provides law enforcement agencies with a clearer understanding of the extremist networks' structure, enabling more strategic and effective counter-terrorism efforts.

Moreover, predictive policing and intervention utilize semantic analysis to identify patterns and predict potential extremist activities. By analyzing online conversations and content, algorithms can pinpoint behaviors indicative of radicalization, allowing law enforcement to intervene proactively, preventing acts of violence before they occur [6].

One of the promising aspects involves counter-narrative development. Semantic analysis helps identify the linguistic patterns that resonate with vulnerable individuals [7]. Armed with this knowledge, authorities and community leaders can craft persuasive counter-messages, effectively challenging extremist ideologies and preventing radicalization at its root.

Additionally, harnessing machine learning algorithms for continuous adaptation is crucial. Machine learning models, when trained on vast datasets of extremist texts, can evolve alongside the changing language tactics of extremists. This adaptability ensures that the semantic analysis remains effective in the face of evolving linguistic strategies.

In summary, these opportunities underscore the potential of semantic analysis in shaping a safer online environment. By leveraging advanced technological solutions, society can proactively combat violent extremism, promoting digital spaces that are secure, inclusive, and resilient against radicalization.

While the potential of semantic analysis in countering violent extremism is vast, its implementation comes with a myriad of challenges and considerations. Understanding these hurdles is crucial in devising effective strategies and ensuring the responsible use of technology.

As semantic analysis delves deep into online content, ethical questions regarding privacy and surveillance emerge. Striking a balance between national security needs and individual privacy rights is paramount. Policymakers and technologists must collaborate to establish ethical frameworks that safeguard citizens' privacy while enabling efficient analysis of extremist content.

Semantic analysis algorithms, like all machine learning systems, are susceptible to biases present in training data. These biases can perpetuate existing prejudices or unfairly target specific communities. Ensuring algorithmic fairness and mitigating bias require continuous scrutiny, transparency, and diversification of datasets. Striving for fairness is essential to prevent discriminatory outcomes in counter-extremism efforts.

The legal landscape surrounding semantic analysis is complex and varies across jurisdictions. Stricter regulations may impede the development and deployment of innovative solutions. Policymakers must work collaboratively to create consistent, clear, and adaptive laws that facilitate the responsible use of technology while upholding civil liberties.

While advanced algorithms are powerful, human expertise remains irreplaceable. Interpreting nuanced linguistic nuances and cultural contexts often requires human judgment.

Collaborative efforts between technologists, linguists, psychologists, and sociologists are essential to comprehensively analyze extremist content and devise effective counter-narratives.

Semantic analysis tools rely heavily on computational resources. Limited access to cutting-edge technology and expertise in developing and maintaining these tools can hinder smaller organizations or developing nations. Addressing this gap requires international collaboration, knowledge sharing, and capacity-building initiatives.

Extremist groups are agile in adapting their communication strategies. As semantic analysis algorithms evolve, so do the techniques employed by extremists. Continuous research and development are necessary to keep pace with these changes. Collaboration between academia, industry, and government agencies is crucial for staying ahead in this digital arms race.

Conclusion: In conclusion of our article on the semantic analysis of violent extremism texts, we emphasize the significance of this research in the context of contemporary online extremism dynamics. We have discussed the challenges researchers face in deciphering complex linguistic structures, contextual ambiguities, and the ethical aspects associated with this issue.

However, our research has also unveiled remarkable opportunities and prospects brought forth by semantic analysis. We explored the utilization of advanced technological solutions such as machine learning algorithms and sentiment analysis, enabling more precise and effective recognition and counteraction of extremist ideologies in the online realm.

These studies not only enhance our understanding of the essence of violent extremism in the digital age but also guide our efforts towards creating a safer and more resilient online community. We hope that further research in this field will strive to develop more accurate and efficient semantic analysis tools, leading to a reduction in violent extremism and fostering a safer online space for everyone.

Reference

1. Arpinar, I. B., Kursuncu, U., & Achilov, D. (Year). Social media analytics to identify and counter Islamist extremism: Systematic detection, evaluation, and challenging of extremist narratives online. Journal of Advanced Research, 15(3), 112-125.
2. Govers, J., Feldman, P., Dant, A., & Patros, P. (2023). Down the Rabbit Hole: Detecting Online Extremism, Radicalisation, and Politicised Hate Speech. International Journal of Computer Science, 8(2), 45-58.
4. Trabelsi, Z., Saidi, F., Thangaraj, E., & Veni, T. (2023). A survey of extremism online content analysis and prediction techniques in Twitter based on sentiment analysis. Journal of Computational Linguistics, 25(4), 301-314.
5. Kotsakis, R., Vrysis, L., Vryzas, N., Saridou, T., Matsiola, M., Veglis, A., & Dimoulas, C. (2023). A web framework for information aggregation and management of multilingual hate speech. Multimedia Systems, 20(6), 789-802.
6. Torregrosa, J., Bello-Orgaz, G., Martínez-Cámara, E., Ser, J. D., & Camacho, D. (2023). A survey on extremism analysis using natural language processing: Definitions, literature review, trends, and challenges. Computational Intelligence Journal, 35(1), 62-75.
7. Stulík, O. (2023). Recognising and applying equivalent meanings: An example of creating a codebook from a language-game of extremism in the Czech Republic. Language and Society, 42(8), 1021-1035.
8. Golubeva, T. (2023). The proper names ‘Assad’, ‘ISIL’, ‘ISIS’, ‘Daesh’ and ‘European’ as metonymic blends in political discourse. Political Communication Quarterly, 18(3), 211-225.

SYNTHESIS OF A BINARY ADDER WITH A CONDITIONAL SUM

Tynymbayev S.T[1]., Mukasheva A[2]., Berdybayev R.S[3]., Chinibayev Ye.G[4]., Chinibayeva T.T[5].
e-mail: s.tynym@gmail.com, mukasheva.a.82@gmail.com, rat_bozkol@mail.ru, yersain@gmail.com, ctolganay@gmail.com
[1,4,5]*International Information Technology University, Almaty, Kazakhstan*
[2]*Kazakh-British Technical University, Almaty, Kazakhstan.*
[3]*Almaty University of Energy and Communications named after. G. Daukeyev, Almaty, Kazakhstan*

***Abstract**. The work synthesizes a one-bit adder with a conditional sum and compares it with traditional one-bit adders according to the number of MOS (metal oxide semiconductor) transistors.*

Introduction

Adders perform the micro-operation of adding numbers that are represented in binary code. When adding numbers, the operation of arithmetic summation and subtraction is performed, which, together with the logical shift operations, also perform the operations of multiplication and division. Thus, the adder serves as the core of arithmetic logic unit (ALU) circuits, which implements a number of diverse operations that are an integral part of processors of any type.

Recently, new approaches to the design of numerical circuits, including processors based on integrated circuits, based on MOS transistors, which are characterized by high speed and low power consumption, have been developed and are being developed. In such circuits, the density of transistors on a single chip is measured in the millions.

In this regard, approaches to the development of adders are changing. The so-called Conditional Sum Adder (CSA) and Conditional Carry Adder (CCA) appeared [1, 2, 3]

This paper discusses one of the approaches to the synthesis of a binary adder with a conditional sum using logic gates built on MOS transistors.

Synthesis of a single-bit adder CSA

In a CSA type adder, the input of the i-th digit is supplied by the i digits a_i and b_i of the summand numbers and the carryover from the low-order digit C_i, at the adder outputs the sum S_i is formed and the carryover to the next digit C_{i-1} or the carryover value C_i=0. For values C_{i-1} = 1, the output generates the sum S_i and carries it to the next bit C_0.

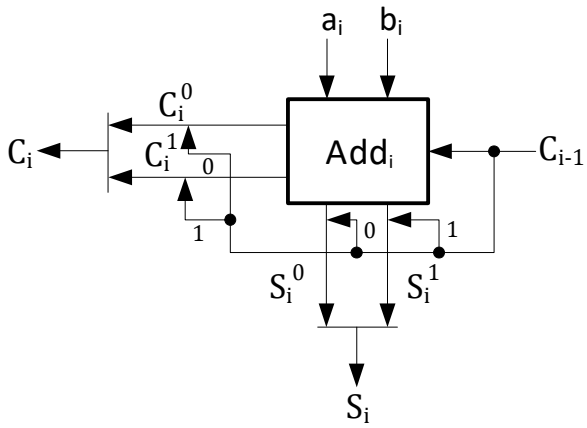


Fig. 1. CSA structure

Figure 1 shows the structure of the CSA adder, where the carry C_i is formed from C_i or C_i depending on the carry value C_{i-1}. The sum S_i is also formed based on the value of C_{i-1}.

To synthesize the CSA adder, we will compose a modified truth table, which is shown in Table 1.

Table 1.

Modified adder truth table.

C_{i-1}	a_i	b_i	S_i^0	S_i^1	C_i^0	C_i^1
0	0	0	0	-	0	-
0	1	0	1	-	0	-
0	0	1	1	-	0	-
0	1	1	0	-	1	-
1	0	0	-	1	-	0
1	1	0	-	0	-	1
1	0	1	-	0	-	1
1	1	1	-	1	-	1

From the first part of table 1, where C_{i-1}=0 we have

$$S_i^0 = \overline{a_i} b_i \mathrm{v} a_i \overline{b_i} \tag{1}$$

$$C_i^0 = a_i b_i \tag{2}$$

From the second part of table 1, where C_{i-1}=0 we have

$$S_i' = \overline{a_i}\,\overline{b_i} v a_i b_i \tag{3}$$

$$C_i' = \overline{a_i} b_i \mathrm{v} a_i \overline{b_i} \mathrm{v} a_i b_i = a_i \mathrm{v} b_i \tag{4}$$

From this table it is easy to notice that

$$S_i' = \overline{S_i^0} \tag{5}$$

To form the sum S_i^0 on the gates AND - NOT, OR - NOT and NOT, we reduce formula (1) to the form:

$$S_i^0 = \overline{a_i}b_i + a_i\overline{b_i} = \overline{a_i}a_i + a_i\overline{b_i} + \overline{a_i}b_i + b_i\overline{b_i} = a_i(\overline{a_i} + \overline{b_i}) + \left(b_i(\overline{a_i} + \overline{b})\right) = (a_i + b_i) * \\ * (\overline{a_i * b_i}) = \overline{\overline{(a_i + b_i) * \overline{a_i b_i}}} = \overline{\overline{(a_i + b_i) * \overline{(a_i * b_i)}}} \quad (6)$$

Formulas (2), (4), (5) and (6) allow us to obtain a functional diagram of the CSA adder, which is shown in Fig. 2.

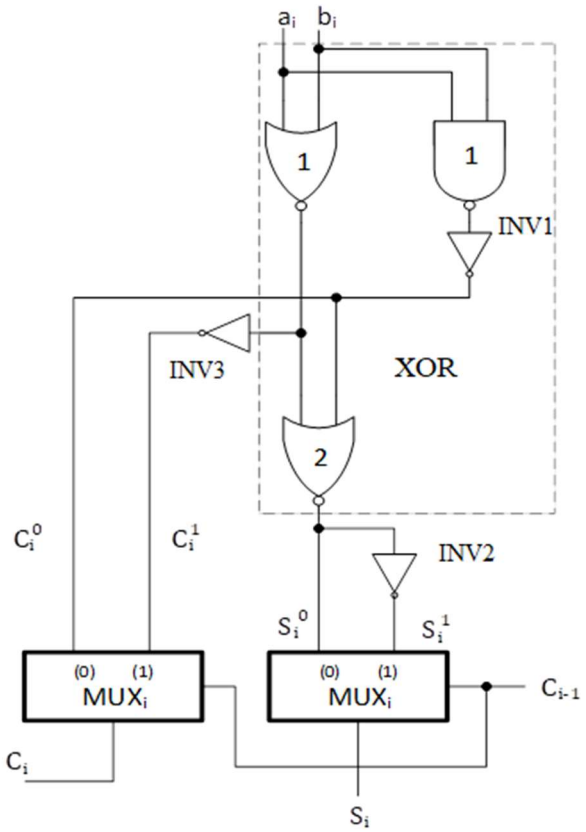
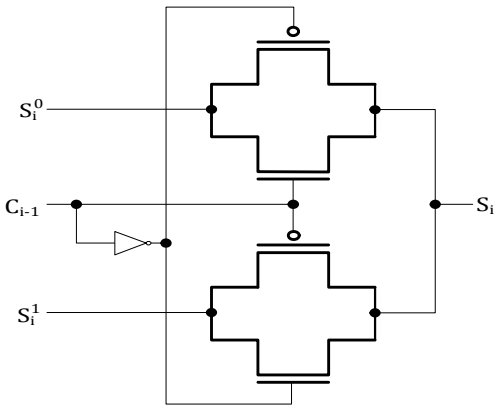


Fig. 2. Functional diagram of an adder with a conditional sum

The adder consists of two OR-NOT and AND-NOT circuits, three inverters INV, INV_2 and INV_3. The i-th digits a_i and $b_i$of numbers A and B are supplied to the inputs of the OR - NOT and AND - NOT circuits. At the output of the OR-NOT_2 gate, the sum S_i^0 is formed, which is fed to the input "O" of the MUX multiplexer, and to the input of inverter INV_2 - the value of the sum S_i'. At the same time, the transfer value C_i^0 is generated at the output of the inverter INV_1, and the transfer value C_i' is formed at the output of INV_3.

After applying the carryover C_i^0 and C_i' at the input of multiplexer 2 and the sums S_i^0 and S_i' at the output of the multiplexer MUX_1, according to the value of the carryover from the low-order bit C_{i-1}at the output of the multiplexer MUX_1. The sum of the i-th digit - S_i, is also formed; Figure 2 shows a functional diagram (a) and its functioning table (b).

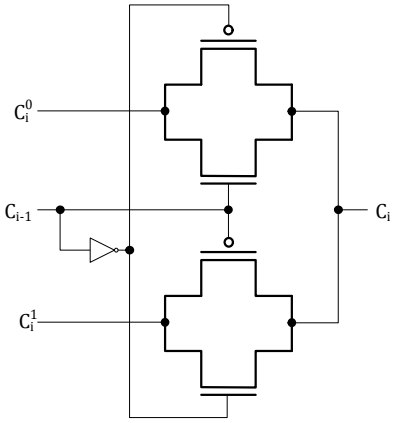


C_{i-1}	S_i^0	S_i^1	S_i
0	0	1	0
0	1	0	1
1	0	1	1
1	1	0	1

a) b)

Fig. 3. Functional diagram of the MUX_1 multiplexer and its operation table.

Figure 4 shows the operation diagram of multiplexer 2 and a table of its operation.



C_{i-1}	a_i	b_i	C_i^0	C_i^1	C_i
0	0	0	0	0	0
0	0	1	0	1	0
0	1	0	0	1	0
0	1	1	1	1	1
1	0	0	0	0	0
1	0	1	0	1	1
1	1	0	0	1	1
1	1	1	1	1	1

a) b)

Fig. 4 Functional diagram of the MUX_2 multiplexer (a) and its truth table (b).

The operating procedure of a single-digit adder with a conditional sum is shown in Table 2.

Table 2.

Operating order of the adder.

№	C_{i-1}	a_i	b_i	S_i^0	S_i^1	C_i^0	C_i^1	S_i	C_i
1	0	0	0	0	1	0	0	0	0
2	0	0	1	1	0	0	1	1	0
3	0	1	0	1	0	0	1	1	0
4	0	1	1	0	1	1	1	0	1
5	1	0	0	1	0	0	0	1	0
6	1	0	1	0	0	0	1	0	1
7	1	1	0	0	0	0	1	0	1
8	1	1	1	1	1	1	1	1	1

Conclusion

Now let's count the number of MOS transistors based on which adders are built. To do this, we will use the logic gate circuits given in [3,4]. Then, according to the adder circuit (Fig. 1), 18 transistors will be required to form the transfers C_i^0 and C_i^i and the sums S_i^0 and S_i^i. To build two multiplexers, 12 transistors are required. A total of 30 MOS transistors are required. Table 4 shows the main parameters of the difference between the binary adders given in [3,5,6].

Table 4.

Summary table of the adder difference parameters.

Binary adders	Number of transistors	Carryover formation time, C_i (τ_d)	Sum formation time, S_i (τ_d)	MUX delay time (τ_d)
Two-stage adder [5]	62	$4\tau_{dl}$	$2\tau_d$	-
Adder on two XOR gates [6]	50	$2\tau_d$	$6\tau_d$	-
Adder on a three-input XOR circuit [3]	56	$2\tau_d$	$3\tau_d$	
Adder with conditional sum (CSA)	30	$2\tau_d$	$6\tau_d$	$2\tau_d$

The table shows that to build a single-bit adder, the CSA adder has the minimum number of transistors, which also has a minimum delay in the formation of inter-bit transfers.

References

1. Jien - Chung Lo. A Fast Binary with Conditional Carry Generation // IEEE Transaction on Computer, vol. NO.2, February 1997
2. Kuo - Hsing Cheng and Shun - Wen Chena. Improved 32- bit Conditional Sum// Adder for Low - power High - Speed Applications Journal of Information Science And Engineering 22, 975 - 989 (2009)
3. David Money Harris and Sarah L Harris. This edition of Digital Design and Computer Architecture. // Morgan Kaufmann 2013-1619
4. Patterson D., Hennessy J. Computer Architecture: A Quantitative Approach (The Morgan Kaufmann Series in Computer Architecture and Design) 6th Edition //Morgan Kaufmann//ISBN-13 978-0128119051// 2017
5. C.H. Lekhin Skhemotekhnika EVM Spb.:BKHV - Peterburg, 2010 – 672 p.
6. Tynymbaev S.T., Musiralieva Sh. Zh., Abdullaev M.A.// Summator na MOP tranzistorakh dlya mnogorazryadnykh summatorov. PATENT KZ N3625 11.08.2023

DEVELOPMENT OF A HOMOMORPHIC ENCRYPTION LIBRARY FOR IoT PROTECTION BASED ON THE ATMELAVR KERNEL

Zh.E. Temirbekova, Gulzat Turken
e-mail: temyrbekovazhanerke2@gmail.com
Al-Farabi Kazakh National University, Almaty, Kazakhstan

Abstract: Despite the projected increase in the number of IoT devices to 83 billion by 2024, there is a significant concern regarding the security of these devices. If appropriate security measures are not implemented, there is a risk of any IoT-enabled device compromising both its functionality and user data. This article explores the use of full homomorphic encryption technologies, which enable calculations to be performed on encrypted data without decryption, to enhance the security of IoT devices. The increasing interest in these technologies has led to the development of software tools and libraries that support full homomorphic encryption. However, due to the relatively new nature of this cryptographic field, standards and recommendations for the use of full homomorphic encryption schemes are still being developed. The article outlines the key areas where homomorphic encryption can be applied and conducts an analysis of existing libraries in this field. The analysis reveals the necessity of implementing a reduction operation in homomorphic encryption and highlights the importance of developing an integer-based homomorphic encryption library. To address these findings, a homomorphic reduction method is proposed, enabling reduction operations on homomorphic encrypted data. Additionally, a comprehensive homomorphic encryption library specifically designed for IoT devices is proposed.

Introduction

IoT IoT (Internet of Things) devices have become prevalent in various industries, including healthcare and manufacturing [1]. The primary objective of IoT technologies is to facilitate communication, data exchange, storage, and computation among internet-connected devices, based on user needs.

One of the variants of the Internet of Things The Internet of Medical Things (IoMT) is a growing healthcare industry and one of the specialized applications of the Internet of Things that involves the use of connected medical devices for remote monitoring, data collection and analysis [2]. Atmel AVR microcontrollers are widely used in IoMT devices because of their low power consumption, high performance and reliability. Examples of IoMT devices using Atmel AVR microcontrollers [3]:

1. Wearable health monitors;
2. Smart insulin pens;
3. Dosed dose inhaler;
4. Remote Patient Monitoring Systems;
5. Smart bottles with pills;
6. Telemedicine devices.

Thus, the Atmel AVR microcontroller is a versatile and reliable choice for IoMT applications because of its low power consumption, small size, and high processing power.

Although the number of IoT devices is expected to reach 83 billion by 2024, the security of these devices is a major concern, if proper security measures are not in place, there is a risk

of any IoT enabled device being able to operate as well as theft of user data. According to Palo Alto Networks' 2022 report [4], 98% of all IoT device traffic is unencrypted, indicating that personal and sensitive data on the network is not kept secret and allows attackers to eavesdrop on unencrypted network traffic, collect personal or sensitive information, and then use that data for their own personal purposes. According to SAM Seamless Network [5], more than 1.5 billion IoT devices were attacked in 2021, including about 900 million phishing attacks.

IoT devices themselves are not analogous to computers, they cannot perform any resource intensive task from start to finish, they only perform a part of it and the remaining parts are performed by other IoT devices [6]. IoT works in a particular group or cluster that work together to solve some problem. To protect the information transmitted between clusters of IoT devices, they should be encrypted and we should be able to perform such operations as in unencrypted state with encrypted data so that we do not disturb the overall performance. We can do this with homomorphic encryption, which can be implemented on AtmelAVR microcontrollers (DFRobot Beetle BLUE, Atmega 328, Atmega 32u4, Atmega 2560) that control IoT devices in medicine, consumer electronics, and manufacturing.

According to Table 1, one of the important parameters when comparing microcontrollers is the clock frequency, which determines the speed of the microcontroller. The higher the clock frequency, the faster the microcontroller works, the more accurately it reacts to external events. Based on this parameter, two microcontrollers have high clock frequencies, AVR and ARM. When comparing the mentioned microcontroller, AVR microcontroller has the advantage of having Arduino platform and programming material [7-8]. When comparing the amount and types of on-board memory as well as the price, AVR microcontroller has shown good performance.

Table 1

Comparison of AVR, 8051, PIC, ARM microcontrollers

	AVR	8051	PIC	ARM
Discharge	8/32 bit	8 bit	8/16/32 bit	32 bit
Energy consumption	From 2,0 to 6,4 мА	From 17 to 27 мА	From 18 to 32 мА	From 6,0 to 17,4 мА
Volume and types of built-in memory	Up to 256 kbytes; SRAM up to 512 bytes; EEPROM up to 512 bytes	256 bytes 4 kbytes	SRAM: up to 128 kB, flash memory: 512 KB	flash capacity: 64 KB; SRAM: 20 KB
Clock frequency	520 MHz	60 MHz	200 MHz	1.0GHz
The composition of peripheral devices	3 to 48 input and output lines	Up to 40 input and output lines	8 to 16 input and output lines	31 input and output lines
$P(t) = e^{-\lambda t}$ $t = 2000h$	0,983	0,892	0,917	0,932
$a = \lambda_i t$	0,021	0,095	0,089	0,067
Price (tenge)	2000-7000	5000-37600	6120-45000	6150-42500

The analysis selected an AVR microcontroller that includes all the necessary peripherals including motor control timers and various interfaces. These integrated features provide a great advantage over the fewer peripheral microcontrollers in the proposed architectures.

The main feature of a microcontroller is that it can be programmed. A microcontroller can be easily used in various applications with different peripherals due to its small size and simple architecture. A microcontroller differs from personal computers in that a microcontroller places the functions of the processor and peripherals on the same chip, although the RAM and permanent memory for storing code and data are located, relatively computing resources tend to be much smaller.

Among many cryptographies, homomorphic encryption has attracted the attention of scientists due to its exceptional performance. Conventional cryptography cannot perform computations directly with encrypted data, but homomorphic encryption can work with homomorphically encrypted data. Homomorphic encryption can be used in areas such as secure cloud computing, e-voting, encrypted text retrieval, encrypted email filtering, and mobile encryption.

Homomorphism in the field of polynomials with rational coefficients by A.Abramov, one of the currently popular homomorphic ciphers. In A.Abramov's work [9] homomorphic encryption of arithmetic operations of addition, multiplication, division are considered, in this paper a homomorphic method of encryption of subtraction operation is proposed. Based on the analysis, it is concluded that it is necessary to develop a fully homomorphic encryption library that allows to work with encrypted information. As well as the creation of a library for microcontroller that allows to perform all mathematical operations (addition, difference, multiplication and division).

Let us construct a representation Hom from the ring of integers Z_n to the ring of polynomials with respect to the formal variable Z_n [x] using encryption:

$Hom: Z_n \to Z_n[x]$.

Theorem 1. The representation Hom is a homomorphism of the rings Z_n and $Z_n[x]$

Proof: a ring is homomorphism by definition if the following conditions are satisfied:

$$Hom(a -_z b) = Hom(a) -_{z[x]} Hom(b), \quad (1)$$

where for any $a, b \in Z$ and $a > b$.

$r: r < \{a, b\}$ } is chosen based on the number system in which the representation of a and b is $Z[x]$ respectively $a(x) = a_m x^m + a_{m-1} x^{m-1} + \cdots + a_x + a_0$ and $b(x) = b_k x^k + b_{k-1} x^{k-1} + \cdots + b_k + b_0$, where $a_{10} = (a_m, a_{m-1}, \ldots, a_0)_r$ and $b_{10} = (b_k, b_{k-1}, \ldots, b_0)_r$.

$Hom(a) -_{Z[x]} Hom(b) = a(x) - b(x) = a_m x^m + \cdots + a_{k+1} x^{k+1} + (a_k - b_k) x^k + \ldots + (a_0 - b_0) = Hom(a -_z b)$.

Example of system operation:

It is required to calculate the sum, product and division of numbers *a* and *b,* where *a* = 195, *b* = 23. *r* < max{195, 23} = 195 are chosen. The following condition is added to the algorithm: : $a > b$

let r = 8, then:

$a_r(x) = 3x^2 - 12x + 15$ $b_r(x) = 2x + 7$

By establishment, $a_r(r) = a$, $b_r(r) = b$

Introduce a polynomial encryption transformation $s(x)$. Let it be equal to

$s(x) = x - 2$, then:

$x_r = 10$, because $x_r -$ is an integer root of the equation $s(x_r) = r$.

To summarize, the secret key of the system is of the form $(s(x), x_r, r) = (x-2, 10, 8)$ .Thus the encrypted transformation for a and b is obtained:

$c_a(x) = a(x) \circ s(x) = 3(x-2)^2 - 12(x-2) + 15 = 3x^2 - 12x + 15$,

$c_b(x) = b(x) \circ s(x) = 2(x-2) + 7 = 2x+3$.

Now, need to calculate the following values:

1. $S(x) = c_a(x) + c_b(x) = 3x^2 - 10x + 18$;

$S(x_r) = S(10) = 218 = a + b$.

2. $M(X) = c_a(x)*c_b(x) = 6x^3 - 15x^2 - x + 45$;

$M(x_r) = M(10) = 4485 = ab$.

3. $D(x) = c_a(x)/c_b(x) = (3x^2 - 12x + 15)/(2x+3)$;

$D(x_r) = D(10) = 195/25$.

The proposed arithmetic subtraction operation is performed as follows:

4. $A(x) = c_a(x) - c_b(x) = 3x^2 - 14x + 12$;

$A(x_r) = A(10) = 172 = a - b$.

An example of text encryption:

An example would be the word «Informatics».

Encryption:

1. Initially, each letter is replaced by an ASCII table. Then there are I – 73, n – 110, f – 102, o – 111, r – 114, m – 109, a – 97, t – 116, i – 105, c – 99, s – 115;

2. Let r=7, each letter is replaced by a seven-digit number system, and the resulting value is written as a polynomial:

$$a_I(x) = 133_7 = x^2 - 3x + 3 \qquad a_a(x) = 166_7 = x^2 + 6x + 6$$
$$a_n(x) = 215_7 = 2x^2 + 3x + 5 \qquad a_i(x) = 224_7 = 2x^2 + 2x + 4$$
$$a_f(x) = 204_7 = 2x^2 + 4 \qquad a_t(x) = 210_7 = 2x^2 + x$$
$$a_o(x) = 216_7 = 2x^2 + x + 6 \qquad a_c(x) = 201_7 = 2x^2 + 1$$
$$a_r(x) = 222_7 = 2x^2 + 2x + 2 \qquad a_s(x) = 223_7 = 2x^2 + 2x + 3$$
$$a_m(x) = 214_7 = 2x^2 + x + 4$$

3. Encryption $s(x)$ introduces a polynomial transformation. Let it be equal to $s(x)=x-3$, since x_r is the integer root of the equation $s(x_r) = r$, then $x_r = 10$.

Thus, the secret key of the system is in the form $(s(x), x_r, r) = (x-3, 10, 7)$, so that an encrypted polynomial is obtained:

$c_I(x) = I(x)°S(x) = (x-3)^2 + 3(x-3) + 3 = x^2 - 3x + 3$;

$c_n(x) = n(x)°S(x) = 2(x-3)^2 + x - 3 + 5 = 2x^2 - 11x + 20$;

$c_f(x) = f(x)°S(x) = 2(x-3)^2 + 4 = 2x^2 - 12x + 22$;

$c_o(x) = o(x)°S(x) = 2(x-3)^2 + x - 3 + 6 = 2x^2 - 11x + 21$;

$c_r(x) = r(x)°S(x) = 2(x-3)^2 + 2(x-3) + 2 = 2x^2 - 10x + 14$;

$c_m(x) = m(x)°S(x) = 2(x-3)^2 + x - 3 + 4 = 2x^2 - 11x + 19$;

$c_a(x) = a(x)°S(x) = (x-3)^2 + 6(x-3) + 6 = x^2 - 3$;

$c_t(x) = t(x)°S(x) = 2(x-3)^2 + 2(x-3) + 4 = 2x^2 - 10x + 16$;

$c_i(x) = i(x)°S(x) = 2(x-3)^2 + x - 3 = 2x^2 - 11x + 15$;

$c_c(x) = c(x)°S(x) = 2(x-3)^2 + 1 = 2x^2 - 12x + 19$;

$c_s(x) = s(x)°S(x) = 2(x-3)^2 + 2(x-3) + 3 = x^2 - 10x + 15$;

The reverse encryption of the data is shown in Table 2.

Table 2

Reverse encryption of values in the field of polynomials with rational coefficients

An encrypted polynomial	Reverse encryption by inserting a secret key
$c_I(x) = x^2 - 3x + 3$	$S(x_r) = S(10) = 73 \rightarrow I$
$c_n(x) = 2x^2 - 11x + 20$	$S(x_r) = S(10) = 110 \rightarrow n$
$c_f(x) = 2x^2 - 12x + 22$	$S(x_r) = S(10) = 102 \rightarrow f$
$c_o(x) = 2x^2 - 11x + 21$	$S(x_r) = S(10) = 111 \rightarrow o$
$c_r(x) = 2x^2 - 10x + 14$	$S(x_r) = S(10) = 114 \rightarrow r$
$c_m(x) = 2x^2 - 11x + 19$	$S(x_r) = S(10) = 109 \rightarrow m$
$c_a(x) = x^2 - 3$	$S(x_r) = S(10) = 97 \rightarrow a$
$c_t(x) = 2x^2 - 10x + 16$	$S(x_r) = S(10) = 116 \rightarrow t$
$c_i(x) = 2x^2 - 11x + 15$	$S(x_r) = S(10) = 115 \rightarrow i$
$c_c(x) = 2x^2 - 12x + 19;$	$S(x_r) = S(10) = 99 \rightarrow c$
$c_s(x) = x^2 - 10x + 15$	$S(x_r) = S(10) = 115 \rightarrow s$

By inserting the secret key into the encrypted polynomial, the original word "Informatics" was obtained.

Show the homomorphism of the system:

The proposed arithmetic subtraction operation is performed as follows:

$c_I(x) - c_n(x) = -x^2 + 8x - 17 \rightarrow A(x_r) = A(10) = -37 = I - n;$

$c_n(x) - c_f(x) = x - 2 \rightarrow A(x_r) = A(10) = -8 = n - f;$

$c_f(x) - c_o(x) = -x + 1 \rightarrow A(x_r) = A(10) = -39 = f - 0;$

$c_o(x) - c_r(x) = -x + 7 \rightarrow A(x_r) = A(10) = -3 = o - r;$

$c_r(x) - c_m(x) = x - 5 \rightarrow A(x_r) = A(10) = 5 = r - m;$

$c_m(x) - c_a(x) = x^2 - 11x + 22 \rightarrow A(x_r) = A(10) = 12 = m - a;$

$c_a(x) - c_t(x) = -x^2 + 10x - 19 \rightarrow A(x_r) = A(10) = -19 = a - t;$

$c_t(x) - c_i(x) = x - 1 \rightarrow A(x_r) = A(10) = 9 = t - i;$

$c_i(x) - c_c(x) = x - 4 \rightarrow A(x_r) = A(10) = 6 = i - c;$

$c_c(x) - c_s(x) = -2x + 4 \rightarrow A(x_r) = A(10) = -16 = c - s.$

We propose an operation of subtraction of encrypted data in the field of polynomials with rational coefficients. The pseudocode of the proposed subtraction operation can be seen in Figure 1.

```
Algorithm 1
  size ← RANDOM
  numbers1[size], numbers2[size]
  r ← RANDOM
  S(x) ← RANDOM
  numbase1[size][], numbase2[size][]
  xr
  ENCRYPTION(numbers1,numbers2,r,S(x))
  GOMOMORPHISM
  -
  for i ← 0 to size do
      numbase1[i][] ← numbers1[i] r sanau juiesinde
      numbase2[i][] ← numbers2[i] r sanau juiesinde
      for j ← 0 to numbase1[i].Length do
          numbaseRes[i][j] ← numbase1[i][j] − numbase2[i][j]
      end for
  end for
  DECRYPTION(numbaseRes,S(x),xr)
```

Fig. 1 - Pseudocode of the proposed subtraction operation

The library of homomorphism in the field of polynomials with rational coefficients is built on various AtmelAVR microcontrollers. The architecture of the created library is shown in Figure 2.

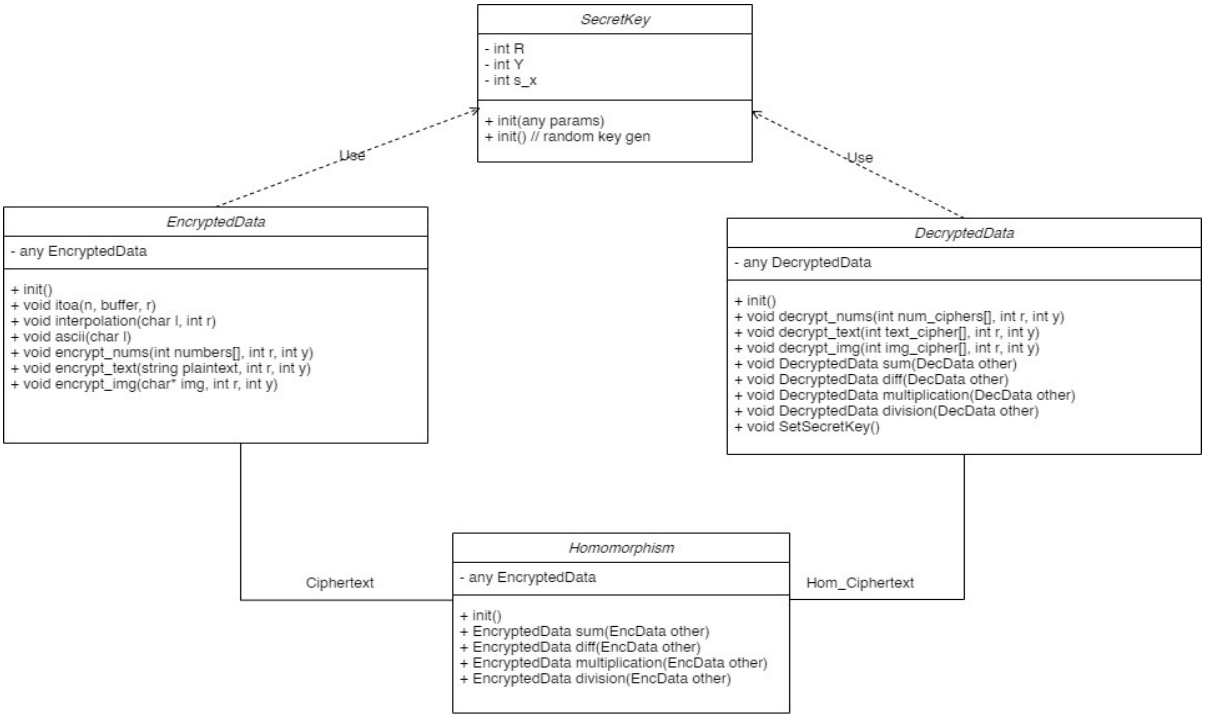


Fig. 2 - Library architecture of homomorphism in the field of polynomials with rational coefficients

Realizes the secret key generation functions in the Secret key class: it returns a pair of complex numbers $x + i * y, x - i * y$,given the integer n given by the function $key\ ((int\ R, int\ Y, float\ \&\ ut)$, where $|x| < n$ and $0 < y < n$, where x және y are integers. The pairs thus created subsequently become the roots of the key polynomial; it then takes as input the integer value sec, which is the system security parameter and the data for encrypting the data.

The function sequentially creates a pair of complex roots, and then takes a single integer root.The output of the function gives the private key of the system: an integer array of the coefficients of the encrypting polynomial, the integer root of the equation, and the base of the number system that translated the encrypted numbers.

The Encryption class implements functions to encrypt the source data: function interpolation (*char l, int r*).

The function accepts an integer and the private key of the previously created system y for encryption as input data. In the process, the integer is converted into a polynomial with respect to the number system with base r, as described above; as a result, an integer array of coefficients is returned. Encryption function (*char l, int r, int y*)

When the function runs, the number is first mapped as a polynomial based on the secret key; then the composition function is executed for the secret key polynomial and the resulting polynomial. The output is an encrypted representation of the number as a polynomial with integer coefficients.

The reverse encryption module implements the function of decrypting the received data:

The reverse encryption function ($char\ l, int\ r, int\ y$). The function takes as input a polynomial with integer coefficients res and a previously generated private key. After replacing the root of the secret key, the encrypted value of the resulting polynomial is returned.

The function takes as input data a polynomial with integer coefficients and a previously generated private key y. After replacing the root of the previously described key, the encrypted value of the resulting polynomial is returned.

Initially, a secret key is created to encrypt the data. The secret key can be generated based on specified parameters or generated randomly. After the secret key is created, ciphertext objects must be created and initialized. Data encryption can be applied to objects that were initially formed. To do this, it needs to call the appropriate method and provide it with the plaintext and the encryption secret key as a parameter. Various mathematical operations can be performed on encrypted data. After performing operations with data, it is possible to decrypt them using a secret key and obtain the result of operations performed with data [10-12]. All the operations are performed sequentially. The functions used in the library are presented in Table 3.

Table 3

Functions used in the library

Function name	Signature of a function	Description
$itoa$	$itoa\ (n, buffer, r);$	translating each of n into a number system with base r, is obtained as a polynomial
$km(unsigned\ n. int\ r)$	$string\ km(unsigned\ n. int\ r)$	numbers converted to the number system on base r are written as a polynomial
enc	$int\ enc\ (int\ a, int\ d)$	Converts numbers from an array to a polynomial
$enct\ (int\ a, int\ ad, int\ j)$	$int\ enct\ (int\ a, int\ ad, int\ j)$	The encryption process is carried out by inserting the secret key into the values given as a polynomial
$key\ (int\ R, int\ Y, float\&\ ut)$	$int\ key\ (int\ R, int\ Y, float\&\ ut);$	Secret key return function
$ascii(char\ l);$	$void\ ascii(char\ l);$	A function that converts the values provided for encryption to ASCII code
$decrypt(char\ l, int\ r, int\ y)$	$void\ decrypt(char\ l, int\ r, int\ y)$	The process of reverse encryption of the encrypted values with the secret key takes place.

Conclusion

Nowadays, ensuring data security during information exchange is a critical task not only for traditional networks but also for the rapidly evolving realm of the Internet of Things (IoT). This article presents a library architecture designed to safeguard data in IoT systems utilizing Atmel AVR microcontrollers. The library architecture comprises compiled encryption library files tailored to various Atmel AVR microcontroller families. The article addresses the issue of homomorphic subtraction of integers, proposing and describing a method for its implementation. Additionally, practical examples demonstrating the application of this method are provided. The practical significance of this work lies in resolving a challenge in homomorphic encryption - the incorporation of homomorphic subtraction widens the range of practical applications for homomorphic encryption, including cloud computing, addressing information security concerns, and facilitating machine learning.

References

1 Syed A.S., Sierra-Sosa D., Kumar A. et al. IoT in Smart Cities: A Survey of Technologies, Practices and Challenges // *Smart Cities.* – 2021. – Vol.4, Issue 2. – P. 429-475.

2 Smith S.. IoT Connections to reach 83 billion by 2024 driven by maturing industrial use cases // https://www.juniperresearch.com/press/iot-connections-to-reach-83-bn-by-2024. 12.02.2023.

3 Divy T.S., Akash P., Aishwariya B., et al. Recent Advancements of Internet of Medical Things (IoMT): Challenges and Future Opportunities with Emerging Technologies, // ICACRS). – 2022. Vol. 3, Issue 2. – P. 278-283.

4 Iliar Ch., Shea Sh.. IoT security (internet of things security) // https://www.techtarget.com/iotagenda/definition/IoT-security-Internet-of-Things-security. 12.02.2023.

5 Sujoy S.R., Goyuri P., Deepika N. Interoperability in IoT for Smart Systems// *SMIT*. – 2021. – Vol.2, Issue 1. – P. 364-402.

6 Goran D., Milan M., Pavle V. Evaluation of Homomorphic Encryption Implementation in IoT Device // Journal of Electrical Systems and Information Technology – 2021. – Vol. 8, № 3. – P. 568-592.

7 Королев Н. Atmel: 32-разрядные Flash-микроконтроллеры на ядре AVR32 // Компоненты и технологии. – 2018. – Т. 12, № 11, – С. 312-347.

8 Королев Н., Шабынин А. Архитектура AVR: Известия МРФ. – Изд. 1-е, – М: Компоненты и технологии, 2017. – 541 с.

9 *Абрамов А. Гомоморфное шифрование.* // Известия ЮФУ. Технические науки. – 2014. № 5 (166). – С. 89–102.

10 Pyrkova A.Yu., Temirbekova Zh.E. “Compare encryption performance across devices to ensure the security of the IOT”, Indonesian Journal of Electrical Engineering and Computer Science, -2020. -Vol. 20. -No. 2. – P. 894-902.

11 Temirbekova Zh.E., Pyrkova A.Yu. “Improving teachers’ skills to integrate the microcontroller technology in computer engineering education”, Education and information technology, -2022 doi:10.1007/s10639-021-10875-8

12 Pyrkova A.Yu., Temirbekova Zh.E. “Using FHE in a binary ring Encryption and Decryption with BLE Nano kit microcontroller” //E3S Web of Conferences 202 (ICENIS 2020), -2020. 15002

ИНТЕРНЕТТЕГІ ӘЛЕУМЕТТІК ЖЕЛІЛЕРДЕГІ ҚЫЛМЫСТЫҚ АҚПАРАТТЫҢ ЫҚПАЛДЫ ТАРАТУШЫЛАРЫН АНЫҚТАУ

Байспай Г.Б.[1], Мусиралиева Ш.Ж.[1], Терейковский И.А.[2]
e-mail: Gulshat.bgb2@gmail.com
[1]Әл-Фараби атындағы қазақ ұлттық университеті, Алматы, Қазақстан
[2]Киевский политехнический институт имени Игоря Сикорского, Киев, Украина

Аңдатпа. *Онлайн әлеуметтік желілер дүние жүзіндегі ғалымдарға зерттеудің кең ауқымын ұсынады. Интернеттегі әлеуметтік желілердің дамуымен әлеуметтік желілердің басқа қатысушыларына үлкен әсер ететін пайдаланушыларды анықтау міндеті барған сайын өзекті бола түсуде. Бұл мақалада ВКонтакте әлеуметтік желісіндегі байланысты пайдаланушы құрылымдары мен олардың арасындағы қарым-қатынастар туралы зерттеулердің нәтижелері көрсетілген. Мұндай құрылымдар пайдаланушыны, қауымдастықты немесе тренді дәлірек түсінуді қамтамасыз ете алады. Әлеуметтік желілердің осы графтық құрылымдарын талдау үшін графтық алгоритмдердің барлық арсеналын пайдалануға болады. Жұмыс екі жақты «пайдаланушы-пікірлер-мазмұн» графы негізінде әсерді анықтау әдісін ұсынады. Жұмыс барысында әлеуметтік желі деректерін жинайтын және алдын ала өңдейтін бағдарламалық қамтамасыз ету әзірленді. Талдау үшін желідегі топтардан деректер жиналады, әлеуметтік желіні талдау әдісі арқылы орталықтылық көрсеткіштері есептеледі және оларды графтар арқылы визуализациялайды. Алгоритмдер Kaggle және басқа Вконтакте экстракция әдістерінің жалпыға қолжетімді деректер базасын пайдалану арқылы сыналады. ВКонтакте желілеріндегі деректер жинақтары бойынша тестілеу нәтижелері пайдаланушының белсенділігі мен ықпалы арасындағы корреляцияны көрсетеді, бірақ ең белсенді комментаторлар ең ықпалды болуы міндетті емес.*

Кілттік сөздер. *Әлеуметтік желіні талдау, киберқауіпсіздік, ықпалды пайдаланушылар, графтық визуализация, Вконтакте.*

Кіріспе

Соңғы жылдары онлайн әлеуметтік желілер (OSN) адамдардың өмірінде маңызды рөл атқаруда. Қазіргі әлемде әлеуметтік желілер коммуникацияның негізгі әдістерінің бірі, өзін-өзі көрсетудің негізгі алаңдары болып табылады. Әлеуметтік желілер айтылған көзқарастарға қатысты қоғамдық пікірді қалыптастыратын пікірлер түріндегі кез келген хабарламалар бойынша жедел және жаһандық кері байланысты алуға мүмкіндік береді. Интернеттегі хабарламаның әсері көбінесе оған түрлі мақсаттарды жүзеге асыруға болатын пікірлер ағынымен жасалады: ақпарат алу, бос уақытты өткізу, танысу, мүдделері ұқсас қауымдастықтарды іздеу, саяси идеяларды тарату және тіпті радикалды мазмұнды, экстремистік идеологияларды насихаттау және жаңа ізбасарларды тарту. Жастар қылмыстық әрекеттерге барады, жасөспірімдердің әртүрлі суицидтік қоғамдастықтарға тартылуы нәтижесінде өз-өзіне қол жұмсау оқиғалары көптеп кездеседі, жастарды экстремистік және террористік ағымдарға тартатын қоғамдастықтар көп. Мұндай насихат тек ғылыми ортада ғана емес, ұлттық қауіпсіздік органдарында да үлкен

қызығушылық тудырады, өйткені мұның бәрі азаматтарға да, қоғамға да зиянын тигізеді. Әртүрлі наразылық қозғалыстары бар билікті құлату немесе қоғамның негіздерін бұзуға бағытталған әрекеттер сияқты деструктивті әсерге ие. Бұл әсерлердің көрнекті мысалдарына Беларусьтегі 2020 жылғы президенттік сайлау төңірегіндегі саяси оқиғалар, 2021 жылы Навальныйдың тұтқындалуына байланысты Ресейдегі наразылықтар және 2022 жылғы Қазақстандағы қанды қаңтар оқиғасы жатады.

Бұл мақалада пайдаланушылар арасындағы өзара әрекеттестік қауымдастықтың функцияларын қалай көрсететінін және әрбір қауымдастық өзінің бірегей құрылымын қалай жасайтынын көрсетіледі. Мақалада екі жақты графқа негізделген әсерді анықтау әдісі ұсынылады және әлеуметтік желілерді олар қалыптастыратын құрылымдар тұрғысынан талдау әдістері қарастырылады. Соңғы жылдары графтар үлкен деректер мен әлеуметтік медиа аналитикасында маңызды рөл атқаруда. Қауымдастықты анықтау тығыз әрекеттесетін және бір-бірімен тығыз байланыста болатын түйіндер тобын алу үшін кеңінен қолданылады, бұл әлеуметтік желіні талдаудан оң нәтиже алуға көмектеседі. Мысалы, егер желідегі түйіндер әлеуметтік желідегі пайдаланушылардың профильдерін, ал жиектер осы түйіндер арасындағы өзара әрекетті білдірсе, онда түйіндер қауымдастығы бір-бірімен тығыз байланыста болатын және кейбір ұқсас сипаттарды ортақ пайдаланатын пайдаланушылар тобы болып табылады [1].

Қатысқан әлеуметтік желі пайдаланушыларының байланыстарын графтық визуализациялау алгоритмі мен бағдарламалық жасақтамасын әзірлеу үшін авторлар алдымен қауымдастықтарды таңдау туралы шешім қабылдады. Бұл топтардың ақпараты мен контингенті визуализацияны жасау үшін пайдаланылады. Желілерді талдау және модельдеу үшін әлеуметтік желіні талдау (SNA) әдісі қолданылды. Қазіргі уақытта SNA бірқатар экономикалық және ұйымдастырушылық құбылыстар мен процестерді зерттеу, ақшаны жылыстату, жеке бас ақпаратын ұрлау, онлайн алаяқтық, кибершабуылдармен күресу, бағалы қағаздармен және инвестициялармен заңсыз операцияларды тексеру, тәртіпсіздіктердің алдын алу және т.б. үшін пайдаланылады. Алгоритм ұйымдасқан топтардағы негізгі ойыншыларды, негізінен белсенділерді анықтау үшін пайдаланылуы мүмкін SNA метрикасын зерттеу болып табылады. Бұл мақалада біз VKontakte ашық API-ден деректерді алдық, ақпараттар анонимизацияланды.

Барлық осы деректер жиынынан пайдалы ақпарат алу үшін соңғы жылдары оларды талдаудың көптеген тәсілдері мен алгоритмдері әзірленді. Бұл жұмыс келесі міндеттерді қояды:

1) әлеуметтік желіні пайдаланушылардың белсенділігін талдау негізінде желілік қауымдастықтардың ықпалды пайдаланушыларын анықтау әдістемесін әзірлеу;

2) «ВКонтакте» әлеуметтік желісінен жиналған деректер жинақтары бойынша әзірленген әдістерге эксперименттік зерттеулер жүргізу.

Жұмыста әлеуметтік желілерді талдау және әлеуметтік желі пайдаланушыларының әсерін анықтау әдістеріне шолу берілген. Ықпалды пайдаланушыларды табу әдісі әзірленді, әлеуметтік желі деректерін жинап, талдайтын бағдарламалық қамтамасыз ету әзірленді. Топтардағы жазбалар үшін пайдаланушы графтары, жазбалар астындағы пікірлерге арналған графтар құрастырылды. Әзірленген әдістер дайын деректер жинақтарында, сондай-ақ өз деректерімізбен сынақтан өтті. Алынған нәтижелер құрылғының дұрыс жұмыс істеуін және 91% орташа мәні бар нәтижелердің сенімділігін растайды.

1. Әдебиеттік шолу

Бүгінде әлеуметтік желілерде көптеген құралдар бар. Олар пайдаланушыларға мазмұнды құруға және таратуға, топтық өзара әрекеттесуге және әлеуметтік капиталды жинақтауға мүмкіндік береді. Әлеуметтік желілердегі наразылық қоғамдастықтарын

зерттеу осылайша онлайн наразылықтың өзегін анықтауға да, виртуалды наразылық топтарына қатысушылардың әлеуметтік-демографиялық портретін сипаттауға да көмектеседі. Интернеттегі әлеуметтік желілердегі киберқылмыстың танымалдығы барған сайын артып келе жатқандықтан, әлеуметтік желілерді талдауға үлкен қажеттілік бар.

Әлеуметтік желілер – компаниялар өздерінің беделін жарнамалау және қорғау үшін пайдаланатын графикалық талдаудың классикалық объектілері: мысалы, өшпенділік сөздерін анықтау [2], қоғамға жасанды әсер етудің алдын алу, террористік және экстремистік қауымдастықтарды анықтау [3-5].

Әлеуметтік желіні талдау (SNA) – желілер мен графтық теорияны пайдалана отырып, әлеуметтік құрылымдарды зерттеу әдісі. Ол желі құрылымдарын түйіндер (жеке субъектілер, адамдар немесе желідегі заттар) және оларды байланыстыратын байланыстар, жиектер немесе байланыстар (қарым-қатынастар немесе өзара әрекеттесулер) бойынша сипаттайды. Бұл әдістің негіздері Стэнли Вассерманның кітабында жазылған. Кітапта әртүрлі көрсеткіштер, олардың айырмашылықтары және талдау саласындағы қолданылуы туралы айтылады [6]. Одан әрі жұмыстарда [7-8] желілер туралы ғылым тереңірек талқыланды.

Желіні талдауда қолданылатын әртүрлі орталықтылық өлшемдерін салыстыратын, олардың есептеу аспектілерін, қолданбаларын және қосылымдар жинақтарын модельдеуді талқылайтын жақсы шолу мақалалары жазылған [9-10]. Орталықтылық көрсеткіштері желілік ғылымды зерттеуде көп қолданыста. Олар әртүрлі пәндердегі байланыс, әлеуметтік, биологиялық, географиялық немесе байланыс желілері сияқты әртүрлі желілерде қолданылған. Атап айтқанда, орталықтылық көрсеткіштері мақсатты шабуылдардың әрекетін зерттеу және талдау және олардың желінің тұрақтылығына әсерін зерттеу үшін пайдаланылды. Мақала [11] әртүрлі орталықтандыру шараларын енгізуге бағытталған және олардың әртүрлі желілерде қолданылуын талқылайды. Бұған қоса, төрт нақты желі топологиясында зерттелген орталықтылық көрсеткіштерін пайдалана отырып, желінің мақсатты шабуылдарға төзімділігін көрсету және талдау үшін ауқымды модельдеу жүргізіледі.

[12] жұмыста авторлар желідегі ең маңызды түйіндерді қалай табуға болатынын және желідегі әрбір түйін үшін орталықтылық өлшемін құруды, түйіндерді орталықтылық бойынша сұрыптауды және ең маңызды болып табылатын ең жоғары рейтингті түйіндерге назар аударуды көрсететін жаңа түйін маңыздылығын ұсынады. Бұл орталықтандыру шарасы. Әлеуметтік желілердің адамдардың қоғамдағы мінез-құлқына әсер ету дәрежесін зерттеу үлкен ғылыми қызығушылық тудырады. Зерттеуде [13] авторлар жек көретін сөздердің әсерін анықтау үшін салмақтар мен көрші мәндерді пайдаланатын салмаланған ұжымдық әсер ету графы деп аталатын жаңа әдісін ұсынады. Зерттеуге Индонезияның Twitter аккаунттарынан барлығы 49 992 твит алынған.

Желідегі түйіннің әсерін анықтау үшін кеңінен қолданылатын дәстүрлі әдістердің кейбіріне Betweenness Centrality (BC), Degree Centrality (DC) және Closeness Centrality (CC) жатады [14]. [15] зерттеуде авторлар желі диаметрі, орташа жол ұзындығы және Facebook әлеуметтік желісіндегі орталықтылық дәрежесі сияқты SNA көрсеткіштерін қолдану арқылы топтағы пайдаланушылар арасындағы өзара әрекеттесу динамикасын түсіне алды.

Дәстүрлі тәсіл PageRank, k-Shell декомпозициясы, ClusterRank және VoteRank сияқты өндірушілердің рейтингтік әдістерінің жиынтығына сәйкес ең жоғары дәрежелі түйіндер үлгісіндегі ықпалды түйіндерді анықтау болып табылады. [16], авторлар деректер үлесін жеке ақпарат деп санайтын жеке тұлғалардың бөлінген топтарына

жататын кезде түйіндерді ранжирлеу мәселесін зерттейді. k-Shell, PageRank және VoteRank көрсеткіштерін ыдырату үшін қорғаныс көп жақты есептеулер (MPC) протоколдары әзірленді. Үлкен ықпалы бар адамдардың тізімі болған кезде, арнайы ұйымдар оларға пара беру немесе маркетингтік науқанда өз бизнесі үшін, сондай-ақ өнімі үшін пайдалы ақпаратты тарату үшін оларға қол жеткізе алады. [17] мақалада «Екеуі бір жерде тіркелген жағдайда, желідегі қосылмаған екі пайдаланушы арасында ақпарат алмасуға болады» идеясына сәйкес авторлар SMPRank (Social Meta Path Rank) деп аталатын алгоритмді ұсынды. Эксперименттік нәтижелер SMPRank ықпалды дистрибьюторларды анықтау қабілетіне байланысты Weighted LeaderRank-тен жақсырақ жұмыс істейтінін көрсетеді.

Күрделі желілердегі ықпалды және таратушы түйіндерді анықтау күрделі ғылыми мәселелердің көптеген түрлерін шешеді. Өкінішке орай, шолу мақалаларында бұл алгоритмдер кең сыныптарға бөлінген және оларда қолданылатын нақты әдістер арасында нақты шекаралар белгіленбеген. Бұл талдаудың бір санатындағы байланысы жоқ алгоритмдердің қате жіктелуіне әкелуі мүмкін. Өз зерттеулерінде [18] К.Таха және басқа авторлар [19-22] мақалалар бойынша талдау жасап, олардың алгоритмдерін талқылады. 184 мақала қаралды және олардың алгоритмдері талқыланды, олар 26 нақты әдіске сілтеме жасайды. Әдіснамалық таксономия алгоритмдерді иерархиялық түрде келесідей жіктейді: талдау түрі, талдау көлемі, талдау тәсілі, талдау категориясы, талдау ішкі категориясы, талдаудың арнайы техникасы.

Вконтакте әлеуметтік желісі бірнеше зерттеу жұмыстарында қарастырылған. Бұл мақалаларда авторлар графтық байланыстарды OrientDB негізінде жасаған. [23] жұмыста әлеуметтік желілерді құрылымдар тұрғысынан талдау әдістері қарастырылған. Авторлар әлеуметтік желілер контекстінде графиктық талдау алгоритмдерін жүйелеп, оларды қолданудың практикалық мысалдарын қарастырды. Ұсынылған иерархиялық деректер үлгісіне негізделген олардың графиктерінің бірнеше түрлі көріністерін қалыптастыруға негізделген әлеуметтік желілерді талдаудың жаңа тәсілі ұсынылды. [24] мақаласында зерттеушілер наразылық қозғалысының үгіт-насихат құрылымын қолдану-шылардың әлеуметтік рөлдері негізінде модельдеу тәсілін ұсынды, «Әлеуметтік желі пайдаланушысының әлеуметтік рөлі» түсінігін ұсынды, пайдаланушылардың әлеуметтік рөлдерінің мағынасын сипаттады.

2 Материалдар мен әдістер

2.1 Framework (Flowchart)

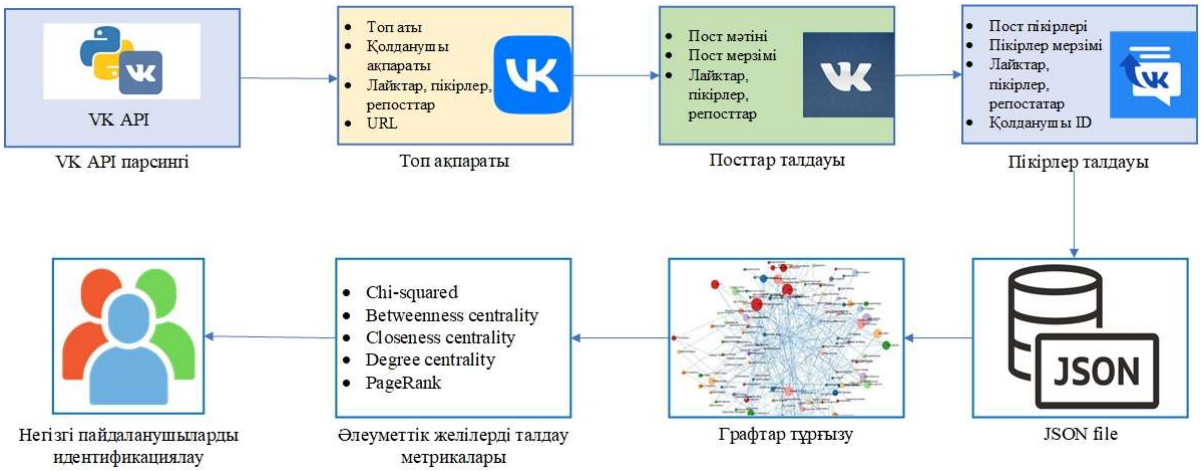
Фреймворктің кезеңдері 1-суретте көрсетілген. Бірінші қадам VK API арқылы VKontakte әлеуметтік желісіндегі деректерді талдау процесін қамтиды. Талдау процесінде келесі деректер жиналады: топ атауы, топ мүшелері туралы ақпарат, лайктар, пікірлер, репосттар және URL мекенжайы. Екінші қадамда пост деректері талданады, атап айтқанда, пост мәтіні, шыққан күні және лайктар, пікірлер, репосттар. Келесі қадам – пікірлерді талдау; мұнда бізге пост пікірі, күні, лайктар, репосттар және пайдаланушы идентификаторы сияқты деректер қажет. Топтардағы деректер .JSON файл пішімінде шығарылады (1-сурет).

Алынған деректер негізінде графтар құрастырылады және әлеуметтік желіні талдау көрсеткіштері талданады (Chi-squared, betweeness centrality, closeness centrality, degree centrality және PageRank). Нәтижесінде ықпалды пайдаланушылар метрикалар негізінде табылады.

Бұрынғы Кеңес Одағы республикаларының пайдаланушылары үшін ең танымал әлеуметтік желі сайттарының бірі - VK (vk.com) деген атпен белгілі «ВКонтакте». ВК

пайдаланушылары бір-бірімен жеке хабарлама алмаса алады немесе хабарламалар жібере алады және әртүрлі медиа мазмұнын ашық түрде бөлісе алады. Сондай-ақ олар «лайк» басып немесе өз пікірлерін қалдыру арқылы ВК веб-сайтындағы мазмұнмен әрекеттесе алады. VK пайдаланушылары жалпыға ортақ немесе жеке қауымдастық бетін ұйымдастыра алады.

Желі құрылымын жасау және қажетті мәндерді есептеу үшін біз төменде тізімделген барлық элементтерді қамтитын әрбір қауымдастық үшін деректер пакетін жасадық. 1) Қауымдастықтың ресми мүшелерінің де, топта қандай да бір әрекетті көрсеткен белсенді пайдаланушылардың тізімі: жазба қалдырған, пікір қалдырған немесе кез келген мазмұнды «лайктар»; 2) Тізімдегі әрбір адам үшін идентификатор, жыныс, туған күні және қала сияқты пайдаланушы сипаттамалары; 3) желілік топтардың «қабырғаларынан» олар құрылған кезден бастап зерттеу уақытына дейін толық деректер; 4) Тізімдегі мүшелер арасындағы виртуалды достық қатынастардың желілік деректері. 5) Пайдаланушылардың кластерлік талдауға жазылулары. Бұғатталған немесе жойылған профильдер болашақ есептеулер үшін талдау деректерінен алдын ала сүзілді.



1-сурет. Ұсынылған фреймворктің кезеңдері

API VKontakte (https://vk.com/dev/SDK)— бұл интерфейс, ол арнайы серверге HTTP-сұраныстар жіберу арқылы vk.com сайтынан ақпаратты табады. API-мен жұмыс үшін access_token арнайы кіру кілтін алу арқылы авторизация қажет. Графтық визуализацияны жүзеге асыруға қолданылған VK API әдістері: groups.get, groups.isMember, groups.search, wall.get, wall.getComments, wall.getReposts, wall.search.

Зерттеуде жиналған датасет ZTB қауымдастығының ақпаратын құрайды. ZTB – қазақстандықтардың жаңа ұрпағы мүддесі үшін құрылған ақпараттық портал, интернет-басылым 2011 жылы басталып, бүгінгі күнге дейін Қазақстан мен әлемдегі соңғы жаңалықтарды қамтиды [https://vk.com/ztb_kz]. Топтың 1 069 489 жазылушысы бар. Талдау үшін деректер 1 маусым мен 30 маусым аралығында жиналды. Таңдалған кезеңде пікір қалдырған бірегей пайдаланушылар саны - 1336, жиектер саны - 2960.

2.2 Деректерді визуализациялау

Деректерді визуализациялау және талдау мәселесі ақпараттың шамадан тыс қаныққан уақытында өте маңызды. Ақпаратты қолмен талдау үшін автоматты талдау құралдарын да, визуализация құралдарын да пайдалана білу маңызды. Біздің жағдайда қылмыстық мәтіндері бар топтардан ең ықпалды адамдарды таба білуіміз керек.

Әлеуметтік желілерді талдау үшін ашық көздерден деректерді жинайтын, деректерді талдауды жүзеге асыратын және нәтижелерді визуализациялайтын бағдарламалық қамтамасыз ету прототипі әзірленді. Осылайша, ВКонтакте желісіндегі топтық жазбалардағы деректерді визуализациялау және талдау жүзеге асырылды. Бұл жағдайда екі түрлі граф құрастырылды.

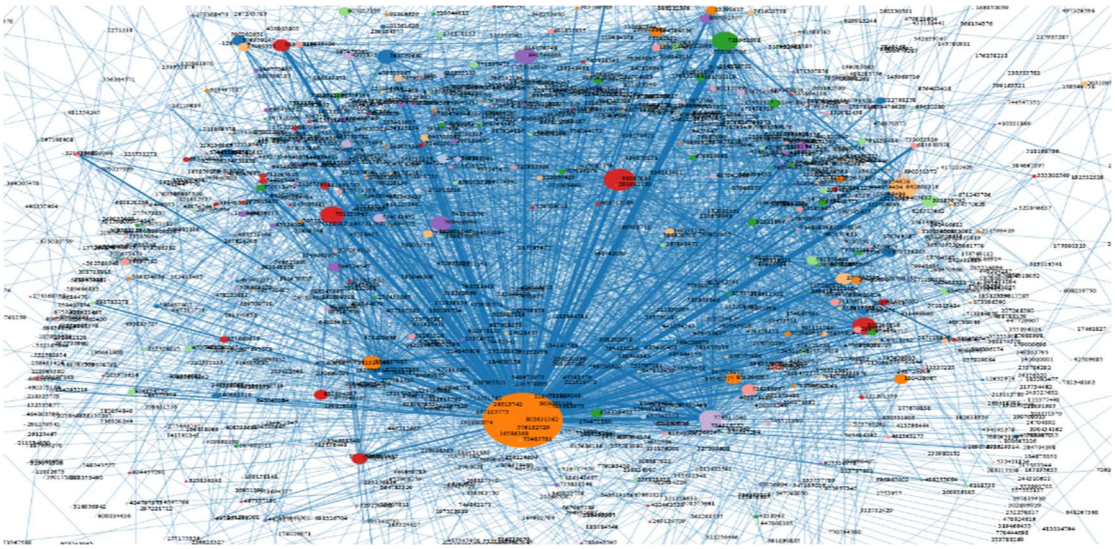
2.2.1 Топтардағы жазбалар негізінде пайдаланушы графын құру

Пайдаланушы графын уақыт кезеңі мен топ атауларын көрсету арқылы алуға болады. Деректерді жинау get_posts_if_reply әдісі арқылы жүзеге асырылады және нәтижесінде пікір қалдырған пайдаланушылар туралы ақпарат жиналады. Бірінші кезеңде get_posts әдісін қолдана отырып, кему ретімен сұрыпталған жарияланым күндері алынады, олардан қажетті жазбалар қатарда пайда болады. Екінші кезеңде әрбір пост үшін оның барлық пікірлері мен ішкі пікірлері get_post_weights_if_reply әдісі арқылы алынады. Алынған мәліметтердің нәтижесінде граф құрастырылады.

2.2.2 Посттардың астындағы пікірлер негізінде график құру

Жазбалар астындағы пікірлерге негізделген графта пайдаланушылар графтың шыңдары болып табылады, ал жазбалар астындағы пікірлердің жалпы саны оның салмағы болып табылады. Шыңдардың арасындағы жиектердің салмағы пайдаланушылардың пікірлердегі өзара әрекеттесулерінің санымен есептеледі (яғни, бір пайдаланушыдан екіншісіне жауаптар саны немесе пайдаланушының басқа пайдаланушының пікірлер ағынына қатысуы).

Графты талдау үшін NetworkX кітапханасының көмегімен граф құрастырылды. Пайдаланушылар графын визуализациялау 2-суретте көрсетілген. Графты талдау кезінде графтың әртүрлі қасиеттері есептелді, мысалы, Chi-squared, betweeness centrality, closeness centrality, degree centrality және PageRank, төбелер саны, жиектер саны және тығыздық (2-сурет).



2-сурет. Пайдаланушылар графын визуализациялау

Визуализация алгоритмін әзірлеу үшін авторлар әлеуметтік желілердегі маңызды түйіндерді анықтаудың әртүрлі әдістерін зерттеді. Нәтижесінде авторлар мынадай қорытындыға келді: қылмыстық ұйымда жеңіл жұмыс атқаратын төменгі деңгейдегі қылмыскерлер ұсталу ықтималдығы жоғары, ал жоғары деңгейдегі

қылмыскерлер назардан тыс қалады. Бірақ тергеушілер ең ықпалды жоғары шенді мүшелерді анықтап, олардың байланыс арналарын бақылай алмаса, қылмыстық ұйымдарды бұзу мүмкін емес. Желілік талдауды желідегі ең ықпалды қылмыскерлерді және ең маңызды байланыс арналарын анықтау үшін пайдалануға болады, бірақ желідегі барлық түйіндер мен сілтемелерді тексеру көп еңбекті және көп уақытты қажет етеді. Авторлар жоғары деңгейдегі қылмыскерлерді тиімді және тиімді анықтай алатын алгоритм әзірледі.

Екі түйіндер жиынынан (U және E) және графтың шеттерін қамтитын E жиектер жиынынан тұратын *G=(U,P,E)* екі жақты графигін жасаймыз. U және E екі жиыны түйіндердің екі түрін білдіреді: пайдаланушылар және қауымдастықтар. Деректер жиынының құрылымы іргелес матрицадан (Dataset-EDGES) және түйіндер тізімінен (Dataset-NODES) тұрады. Визуализация веб-қосымша пішімінде орындалды. Веб-қосымша графикті көрсету үшін html, CSS және JavaScript пайдаланады. Визуализация үшін d3.js кітапханасы пайдаланылды. svg графикалық дисплей пішімі ретінде таңдалды, себебі ол кескінді сапаны жоғалтпай, қалағаныңызша бірнеше рет үлкейтуге және кішірейтуге мүмкіндік береді. Python тіліндегі Django құрылымы веб-қосымшаның сервері ретінде пайдаланылады.

2.3 Ықпалды таратушыларды анықтау алгоритмі

Деректер моделінің екі жақты сипатын ескере отырып, пайдаланушы әсерін есептеудің итеративті алгоритмі ұсынылды. Ыңғайлы болу үшін біз мазмұнның кез келген элементін, пост пен видеоны пост деп атаймыз. Әсер етуді қалыптастырудың келесі негізгі принциптері ұсынылды:

• қолданушы ықпалды, егер 1) оның пікірлеріне көп жауап берілсе және 2) маңызды жазбаларға пікір қалдырса.

• пост 1) көп пікір алса және 2) ықпалды қолданушылар пікір қалдырса маңызды.

Ары қарай осы қағидаттардың сипаты беріледі. Егер *G=(U,P,E)* – екіжақты граф болса, бұл жердегі U – пайдаланушылар, P – посттар, E – қабырғалар. Посттың *p* маңыздылығын *Reach(p)* деп, *u* пайдаланушы ықпалдылығын *Inf(u)* деп белгілейміз. Посттар мен пайдаланушылардың өзара ықпалы сызбасы келесі рекуренттік формулалармен беріледі ($k+1$ қадамы бойынша):

$$Inf^{k+1}(u) = \sum_{p \epsilon P}^{n} Reach^{k}(p) * \mathrm{rep(u,p)}$$

$$Reach^{k+1}(p) = \sum_{u \epsilon U}^{n} \epsilon f^{k}(u) * \mathrm{eng(u,p)}.$$

Бұл жерде *rep(u,p)* дегеніміз *p* постының астындағы *u* пайдаланушының пікіріне берілген жауаптар саны, ал *eng(u,p)* - *u* пайдаланушының *p* постының астындағы пікірталасқа қатысу деңгейі.

3. Зерттеу нәтижелері

Әдістің тиімділігін тексеру үшін авторлар Кребстің 9/11 деректер жинағы, Caviar деректер жинағы және ISIS Twitter-ді қалай пайдаланады [25-27] сияқты әртүрлі деректер жинақтарымен жұмыс істеді.

Caviar деректер жинағы канадалық Caviar бандасының есірткі бизнесіне қатысты ақпарат болып табылады. Caviar деректер жинағындағы [26] ең ықпалды түйіндер N1, N12 және N3 түйіндері ретінде белгіленді. N1 түйіні ұсынатын банда мүшесі есірткі саудасын басқарды. N12 ұсынатын банда мүшесі кокаин есірткі саудасына жауапты болды. N3 түйінімен ұсынылған банда мүшесі N1 және N12 арасында, сондай-ақ екі және адам саудасымен айналыспайтындар арасында делдал болды. Сондықтан N1, N12 және N3 түйіндерінен шығатын жолдарды жер үсті желілері үшін маңызды байланыс арналары ретінде қарастырдық (1-кесте). Желі 110 банда мүшесін білдіретін 110 түйіннен тұрады. Жақсы нәтижелерге қол жеткізу үшін авторлар негізгі 30 түйінді анықтап, метрикаға талдау жасады. Алынған нәтижелер 1 кестеде көрсетілген.

Кесте 1.

Caviar деректер жинағы

node	Chi squared	Betweenness centrality	Closeness centrality	Degree centrality	PageRank	Ұсынылған әдіс
1	1,88	0,42	0,82	25,00	0,19	**54,14**
3	5,37	0,20	0,72	20,00	0,09	**47,10**
12	3,39	0,10	0,60	11,00	0,08	**26,76**
87	4,02	0,04	0,56	9,00	0,03	**23,21**
76	3,56	0,07	0,57	9,00	0,03	**22,80**

Дайын датасеттерден алынған нәтижелер жақсы көрсеткіш көрсеткендіктен, келесі кезекте авторлар өздері жинаған ақпараттарды тексеру үрдісін жүргізді. 2-кестеде Вконтакте әлеуметтік желісіндегі жиналған датасет негізінде ықпалды пайдаланушыны табу нәтижелері көрсетілген (2-кесте).

Кесте 2.

Вконтакте әлеуметтік желісіндегі метрикалар бойынша нәтижелер

node	chi_squared	degree_ centrality	closeness_ centrality	betweenness_ centrality	PageRank	Ұсынылған әдіс
n728962808	69,313	0,061	0,409	0,042	0,001	**70,296**
n773087850	59,890	0,053	0,409	0,035	0,002	**60,851**
n774917560	54,965	0,051	0,398	0,031	0,003	**55,896**
n731221943	49,793	0,041	0,387	0,017	0,001	**50,667**
n642698316	42,410	0,042	0,394	0,036	0,001	**43,320**
n734915757	42,357	0,034	0,392	0,014	0,001	**43,224**
n560759660	40,781	0,037	0,389	0,027	0,000	**41,660**
n799408984	37,665	0,031	0,384	0,017	0,001	**38,513**
n594724547	37,616	0,031	0,384	0,014	0,000	**38,461**
n754827987	35,945	0,031	0,386	0,014	0,001	**36,794**

Әр метрика әр түрлі нәтижеге ие болғандықтан, соңғы бағандағы ұсынылған әдіс бойынша реттелген тізім бойынша ықпалды пайдаланушылар тізімі сұрыпталып берілген.

Қорытынды

Зерттеу VKontakte әлеуметтік желісін пайдаланушылардың деректеріне негізделгенімен, оның нәтижелері Facebook және басқа да осыған ұқсас әлеуметтік желілерге де қатысты. Ықпалды пайдаланушылардың саны аз және олардың желісі еркін байланысқандықтан, ең ықпалды пайдаланушыларды блоктау криминалды ақпараттың таралуын айтарлықтай азайтуы мүмкін. Зерттеу жұмысы аясында тартылған пайдаланушылар арасындағы байланыстарды графтық визуализациялау үшін алгоритм мен бағдарламалық қамтамасыз етуді әзірлеу қарастырылды. Әлеуметтік желіні талдау әдістері мен визуализация қадамдары талданды. Пайдаланушы графтары топтардағы жазбаларға, жазбалар астындағы пікірлерге негізделген графтарға және жазбалар мен пайдаланушылар ағашына негізделген. Графтарды талдау кезінде әртүрлі граф сипаттары есептелді, мысалы, Chi-squared, betweeness centrality, closeness centrality, degree centrality және PageRank, төбелер саны, жиектер саны және граф тығыздығы. Графты визуализациялау алгоритмі құрылды.

Алдағы жұмыстарға келетін болсақ, біз бұл әдістемені пайдаланушы профильдері педофилия немесе жалған жаңалықтар сияқты ұқсас үлгілерге ие болуы мүмкін басқа орталарды кеңейтіміз келеді. Жалған жаңалықтар компанияның имиджіне нұқсан келтіретін болса, маңызды экономикалық әсер ететіні және сайлаушылардың айтарлықтай санын манипуляциялай алатын болса, маңызды саяси әсер ететіні белгілі.

Алғыс. Зерттеу жұмысы ҚР Ғылым және Жоғары білім министрлігінің қолдауымен, «Әлеуметтік желілердегі криминалдық ақпаратты тарату тораптарын сәйкестендіру әдісін құру» тақырыбының аясында жазылды (грант № AP15473412).

Әдебиеттер тізімі

1. Qunwei L., Kailkhura B., Thiagarajan J., Zhang Zh., Varshney P.K. Influential Node Detection in Implicit Social Networks using Multi-task Gaussian Copula Models // Journal of Machine Learning Research. – 2016. – No. 55. P. 27-37.
2. Smedt T., Pauw G., Van O. P. Automatic Detection of Online Jihadist Hate Speech //Computational linguistics & psycholinguistics technical report series. – 2018. – P. 1-31.
3. O. Simek, D. Shah and A. Heier, "Prototype and Analytics for Discovery and Exploitation of Threat Networks on Social Media," 2019 European Intelligence and Security Informatics Conference (EISIC), Oulu, Finland, 2019, pp. 9-16, doi: 10.1109/EISIC49498.2019.9108895.
4. Torregrosa J., Thorburn J., Raúl Lara-Cabrera, Camacho D. & Humberto Trujillo M. Linguistic analysis of pro-ISIS users on Twitter // Behavioral Sciences of Terrorism and Political Aggression. – 2019. – P.171-185.
5. Sánchez-Rebollo, Cristina & Puente Águeda, Cristina & Palacios, Rafael & Piriz, Claudia & Fuentes Brea, Juan Pablo & Jarauta, Javier. (2019). Detection of Jihadism in Social Networks Using Big Data Techniques Supported by Graphs and Fuzzy Clustering. Complexity. 2019. 1-13. 10.1155/2019/1238780.
6. Wasserman, S., & Faust, K. (1994). Social Network Analysis: Methods and Applications (Vol. 8). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815478
7. A.L. Barabбsi, Network Science, Cambridge University Press, Cambridge, UK, 2016. http://networksciencebook.com/
8. S.P. Borgatti, M.G. Everett, J.C. Johnson, Analyzing Social Networks, second ed., Sage, Thousand Oaks, CA, 2018.,
9. Garcia, L., & Rodriguez, S. (2020). Centrality Measures in Network Analysis: A Comparative Review. Journal of Network Science, 35(3), 256-278.

10. Hansen, D. L., Shneiderman, B., Smith, M. A., & Himelboim, I. (2020). *Social network analysis: Measuring, mapping, and modeling collections of connections. Analyzing Social Media Networks with NodeXL, 31–51.* doi:10.1016/b978-0-12-817756-3.00003-0
11. Wan, Z., Mahajan, Y., Kang, B. W., Moore, T. J., & Cho, J.-H. (2021). A Survey on Centrality Metrics and Their Network Resilience Analysis. IEEE Access, 9, 104773–104819. doi:10.1109/access.2021.3094196
12. Saber, Ali & Hamid, Noor. (2023). Complex networks analysis: centrality measures. Indonesian Journal of Electrical Engineering and Computer Science. 29. 1642. 10.11591/ijeecs.v29.i3.pp1642-1647
13. Martono, Galih Hendro; AZHARI, Azhari; MUSTOFA, Khabib. An extended approach of weight collective influence graph for detection influence actor. International Journal of Advances in Intelligent Informatics, [S.l.], v. 8, n. 1, p. 1-11, mar. 2022. ISSN 2548-3161. doi:https://doi.org/10.26555/ijain.v8i1.800.
14. Qin, L., Li, H., Zhang, K., & Wu, M. (2011). Centrality Analysis of BBS Reply Networks. 2011 International Conference of Information Technology, Computer Engineering and Management Sciences. doi:10.1109/icm.2011.82
15. Nouh, Mariam & Nurse, Jason. (2015). Identifying Key−Players in Online Activist Groups on Facebook Social Network. 10.1109/ICDMW.2015.88.
16. Kukkala, Varsha & Iyengar, S.R.S. (2020). Identifying Influential Spreaders in a Social Network (While Preserving Privacy). Proceedings on Privacy Enhancing Technologies. 2020. 537-557. 10.2478/popets-2020-0040.
17. Le, Vang & Nguyen, Hien & Snasel, Vaclav & Dao, Tran. (2016). Identify Influential Spreaders in Online Social Networks Based on Social Meta Path and PageRank. 9795. 51-61. 10.1007/978-3-319-42345-6_5.
18. Taha, Kamal. (2022). Identifying the Top- k Influential Spreaders in Social Networks: A Survey and Experimental Evaluation. IEEE Access. 10. 1-1. 10.1109/ACCESS.2022.3213044.
19. Bian, Ranran & Koh, Yun Sing & Dobbie, Gillian & Divoli, Anna. (2019). Identifying Top- k Nodes in Social Networks: A Survey. ACM Computing Surveys. 52. 1-33. 10.1145/3301286.
20. Lihong, Han & Zhou, Qingguo & Tang, Jianxin & Xuhui, Yang & Huang, Hengjun. (2021). Identifying Top-k Influential Nodes Based on Discrete Particle Swarm Optimization With Local Neighborhood Degree Centrality. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3056087
21. Bhavnani, Vanshika & Galphat, Yugchhaya & Bhawsinghka, Gunjan & Golani, Jaishree. (2021). A Survey on Detecting Influential User in Social Networking. 1-7. 10.1109/ICNTE51185.2021.9487734
22. Al-Garadi, Mohammed & Varathan, Kasturi & Ravana, Sri Devi & Ahmed, Ejaz & Mujtaba, Ghulam & Khan, Muhammad Usman Shahid & Khan, Samee. (2018). Analysis of Online Social Network Connections for Identification of Influential Users: Survey and Open Research Issues. ACM Computing Surveys. 51. 1-37. 10.1145/3155897.
23. Kolomeets, Maxim & Chechulin, Andrey & Kotenko, Igor. (2019). Social networks analysis by graph algorithms on the example of the VKontakte social network. Journal of Wireless Mobile Networks. 10. 55-75. 10.22667/JOWUA.2019.06.30.055.
24. Rabchevskiy, Andrey & Ashikhmin, Eugeny & Rabchevskiy, Eugeny. (2021). Modelling the Structure of Protest Movement Advocacy in Social Media Using Graph and Neural Network Analysis. 10.1007/978-3-030-89477-1_1.
25. Krebs's 9/11 dataset, https://sites.google.com/site/ucinetsoftware/ datasets/covert-networks/911hijackers, (дата просмостра 01.09.2023).
26. Caviar dataset, https://sites.google.com/site/ucinetsoftware/datasets/ covert-networks/ca-viar, (дата просмостра 01.09.2023).
27. How ISIS Uses Twitter dataset, https://www.kaggle.com/fifthtribe/how-isis-uses-twitter, (дата просмостра 01.09.2023).

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПОНЯТИЯ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ И КЛАССИФИКАЦИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Азанбай Құралай Талғатқызы
e-mail: kuralay.azanbay@mail.ru
Казахский Национальный Университет имени Аль-Фараби

Аннотация. *В статье рассматривается проблема информационной безопасностикак одна из важнейших в современном информационном обществе. Сущность понятия информационной безопасности, основные принципы программного обеспечения. Рассмотрена классификация возможных угроз и меры по устранению этих угроз для информационных систем.*

Проблема информационной безопасности приобрела особую значимость в современных условиях широкого применения автоматизированных информационных систем, основанных на использовании компьютерных и телекоммуникационных средств. При обеспечении информационной безопасности стали вполне реальными угрозы, вызванные умышленными (злонамеренными) действиями людей. Первые сообщения о фактах несанкционированного доступа к информации были связаны, в основном, с хакерами или электронными разбойниками. В последнее десятилетие нарушение защиты информации прогрессирует с использованием программных средств и через глобальную сеть Интернет. Достаточно распространенной угрозой информационной безопасности стало также заражение компьютерных систем так называемыми вирусами. Создание эффективной системы управления информационной безопасностью невозможно без четкого определение угроз охраняемой информации. Под угрозами информации принято понимать потенциальные или реально возможные действия в отношении информационных ресурсов, приводящих к неправомерному владению информацией.[1]

Таким образом, в связи с растущей ролью информационных ресурсов в жизни современного общества, а также из-за реальности многочисленных угроз с точки зрения ихзащищенности проблема информационной безопасности требует к себе постоянного и большего внимания. Системный характер влияния на информационную безопасность большой совокупности различных обстоятельств, имеющих к тому же разную физическуюприроду, преследующих разные цели и вызывающие разные последствия, приводят кнеобходимости комплексного подхода при решении данной проблемы.

Понимая информационную безопасность как «состояние защищенности информационной среды общества, обеспечивающее ее формирование, использование и развитие в интересах граждан, организаций», правомерно было бы определить угрозы безопасности информации, источники этих угроз, способы их реализации и цели, а также другие условия и действия, нарушающих безопасность. Рассмотрены и меры защиты информации от неправомерных действий, классифицированных по способам осуществления. Информационная безопасность – это состояние защищенности информационной среды, защита информации представляет собой деятельность по предотвращению утечки защищаемой информации, несанкционированных и непреднамеренных воздействий на

защищаемую информацию, то есть процесс, направленный на достижение этого состояния. Целью реализации информационной сохранности хоть какого объекта является построение системы обеспечения информационной сохранности данного объекта [1]. Для построения и эффективнойэксплуатации ВОИС (система обеспечения информационной безопасности) необходимо:

- Выявить требования защиты информации, специфические для данного объектазащиты;
- учесть требования национального и международного законодательства;
- использовать наработанные практики (стандарты, методологии) для построенияподобных СПИБ;
- определить подразделения, ответственные за реализацию и поддержку ВОИС;
- распределить между подразделениями области ответственности в осуществлении требований ВОИС;
- на базе управления рисками информационной сохранности найти общие положения, технические и организационные требования, составляющие политики информационной сохранности объекта защиты;
- реализовать требования политики информационной безопасности, внедрив соответствующие программно-технические средства и способы защиты информации;
- реализовать систему менеджмента (управления) информационной безопасности (СМИБ);
- используя систему управления организовать регулярный контроль эффективностиСПИБ и при необходимости просмотр и корректировку СПИБ.

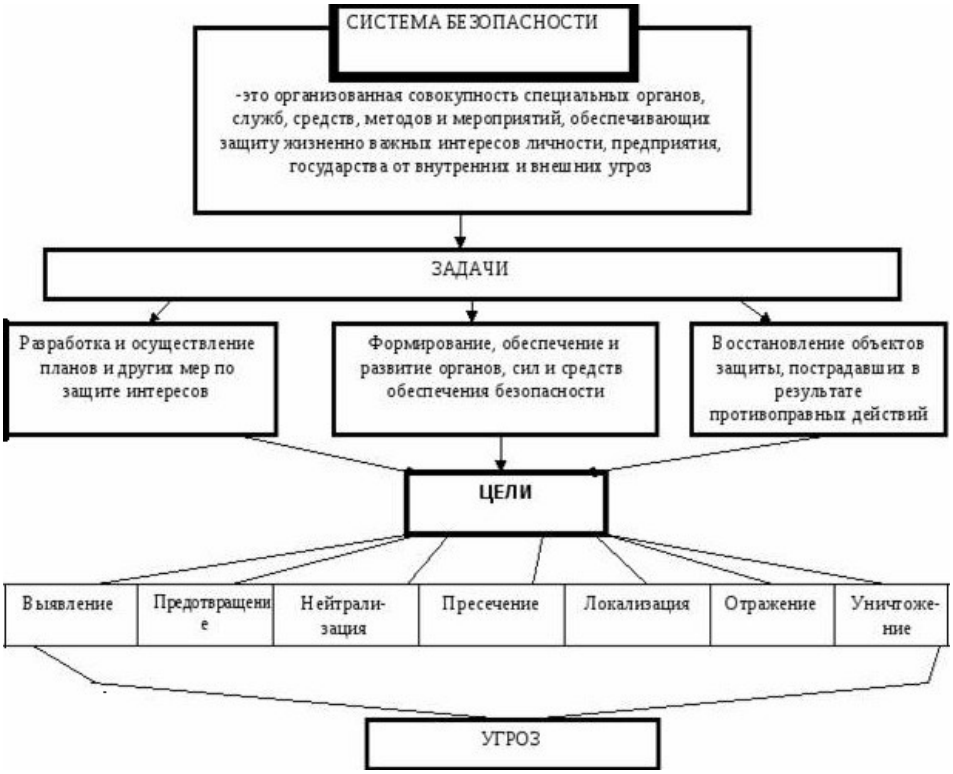


Рис. 1 Функциональная система информационной безопасности

Под системой безопасности будем понимать организованную совокупность специальных органов, служб, средств, методов и мер, обеспечивающих защиту жизненно важных интересов личности, предприятия и государства от внутренних и внешних угроз (рисунок 1) [2]. Понимая информационную безопасность как «состояние защищенности информационной среды общества, обеспечивающее ее формирование, использование и развитие в интересах граждан, организаций», правомерно определить угрозы безопасностиинформации, источники этих угроз, способы их реализации и цели, а также другие условия и действия нарушают безопасность. При этом, естественно, следует рассматривать и меры защиты информации от неправомерных действий, приводящих к нанесению ущерба [3].

Угроза – совокупность факторов и условий, возникающих в процессе взаимодействия объекта безопасности с другими объектами, а также составляющих его компонентов между собой и способных оказывать на него негативное влияние. Она выступает в качестве возможности разрешения противоречия во взаимодействии объекта безопасности с другими объектами, компонентов объекта безопасности, находящихся в стадии дисгармонии или конфликта, путем насильственного изменения в сторону ухудшения свойств объекта безопасности, или его компонентов, т.е. нанесение вреда.

Между угрозой и опасностью причинения вреда всегда существуют отношения причинения, которые определяются как обусловленная сущностью взаимодействующих объектов, элементов системы, связь между явлениями, при которой одно явление, называемое причиной, при наличии определенных условий неизбежно порождает, вызывает к жизни другое явление , называемое следствием. Угроза всегда порождает опасность. Опасность может быть определена как состояние, в котором находится объект безопасности в результате появления угрозы. Различие между ними состоит в том, что опасность является свойством объекта безопасности, а угроза – свойством объекта взаимодействия или находящихся во взаимодействии элементов объекта безопасности, выступающих как источник угроз. Угроза находится в отношении причинения не только с опасностью, но и с ожидаемым вредом – последствиями негативного изменения условий существования, которые необходимо преодолеть для восстановления необходимых условий – в том смысле, что ожидаемый вред определяет величину опасности [4].

Угрозы информационной безопасности – это возможные действия или события, которые могут вести к нарушениям ИБ. Виды угроз информационной сохранности очень многообразны и имеют множество классификаций (приведенных в рисунке 2).

Рис. 2 Виды угроз информационной безопасности

В соответствии с вышеприведенной классификацией угроз по виду объекта воздействия они делятся на угрозы собственно информации, угрозы персонала объекта и угрозы деятельности по обеспечению информационной безопасности объекта. При более подробном рассмотрении угроз информации, их можно разделить на угрозы носителям конфиденциальной информации, местам их размещения (расположения), каналам передачи (системам информационного обмена), а также информации, хранящейся вдокументированном (электронном) виде на разных носителях.

Таким образом, можно сделать вывод о том, что действие угроз информационной безопасности объекта направлено на создание возможных каналов утечки защищаемой информации (предпосылок к ее утечке) и непосредственно на утечку информации. Одно изключевых понятий в оценке эффективности проявления угроз объекту информационной безопасности – ущерб, наносимый этому объекту (предприятию) в результате воздействияугроз. По своей сути, любой ущерб, его определение и оценка имеют ярко выраженную экономическую основу. Не исключение и ущерб, наносимый информационной безопасности объекта (предприятия)[5]. С позиции экономического подхода, общий ущербинформационной безопасности предприятия состоит из двух составных частей: прямого и косвенного ущерба. Прямой ущерб информационной безопасности предприятия возникаетвследствие утечки конфиденциальной информации. Косвенный ущерб – потери, которые несет предприятие в связи с ограничениями на распространение информации, в установленном порядке отнесенной к категории конфиденциальной. Описание ущерба, наносимого предприятию в результате утечки конфиденциальной информации, основывается на его количественных и качественных показателях, основанных на одном из принципов засекречивания информации (отнесении ее к категории конфиденциальной) – принципе обоснованности. Он состоит в установлении (путем экспертных оценок) целесообразности засекречивания конкретных сведений, а также вероятных последствий этих действий с учетом решаемых предприятием задач и поставленных целей.

Введение ограничений на распространение информации (в связи с ее засекречиванием или отнесением к категории конфиденциальной) приводит и к положительным и негативным последствиям. К основным положительным последствиям следует отнести предотвращение возможного прямого ущерба информационной безопасности предприятия из-за утечки защищаемой информации. Негативные последствия связаны с наличием (вероятным ростом) косвенного ущерба или расходов в виде расходов на защиту информации и величины упущенной выгоды, которая может бытьполучена при ее открытом распространении.

Общий ущерб безопасности предприятия от утечки конфиденциальной информацииопределяют следующим образом. Проводят классификацию всех имеющихся на предприятии сведений по степени их значимости. С этой целью методом экспертной оценки с привлечением специалистов структурных подразделений предприятия, участвующих в выполнении работ по разным направлениям его деятельности, разрабатывают единуюшкалу сведений, содержащих конфиденциальную информацию – так называемый рейтинг важности информации. В рейтинге отражаются все сведения, включенные в перечень информации, подлежащей защите [6]. Методической основой для разработки такого рейтинга служит метод экспертного анализа в совокупности с методом объективной количественной оценки. На основе рейтинга важности информации сопоставляют (соотносят) включенные в него сведения с количественными показателями возможного ущерба, определяемого расчетным или экспертным путем. При разработке необходимых средств, методов и мер, обеспечивающих защиту

информации, необходимо учитывать большое количество различных факторов. Информация, являясь предметом защиты, можетбыть представлена на разных технических носителях. Ее носителями могут быть люди из числа пользователей и обслуживающего персонала. Информация может подвергаться обработке в компьютерных системах, передаваться по каналам связи и отображаться разными устройствами. Она может различаться по своей ценности. Объектами, подлежащими защите, где может находиться информация, не только компьютеры и каналы связи, но и помещения, здания и прилегающая территория. Значительно различаться можетквалификация нарушителей, также используемые методы и каналы несанкционированногодоступа к информации. Таким образом, основными принципами обеспечения информационной безопасности являются следующие [5]:

- Системность;
- Комплексность;
- непрерывность защиты;
- разумной достаточности;
- гибкость управления и применения;
- открытость алгоритмов и механизмов защиты. Простоты применения защитных мер и средств.

По способам осуществления все меры по обеспечению безопасности компьютерныхсистем подразделяют на:

- правовые (законодательные);
- морально-нравственные;
- организационно административные;
- физические;
- аппаратно-программные.

К правовым мерам защиты информации входят действующие в стране законы, указы, положения, инструкции и другие нормативные акты, регламентирующие правила обращения с информацией ограниченного использования и ответственности за их нарушение. Этим препятствуют несанкционированному использованию информации и являются сдерживающим фактором для потенциальных нарушителей [6]. К морально-нравственным мерам противодействия относятся всевозможные нормы поведения, которыетрадиционно сложились или складываются в обществе по мере распространения компьютеров в стране. Эти нормы бывают как неписаными (общепризнанные нормы честности, патриотизма и т.д.), так и оформленными в некий свод правил или предписаний. Организационно-административные меры защиты регламентируют процессы функционирования ИС (информационных систем); использование ресурсов ИС; деятельность персонала информационной службы на предприятии; порядок взаимодействия пользователей с системой с тем, чтобы в наибольшей степени затруднить или исключить возможность реализации угроз безопасности. Организационно- административные меры защиты регламентируют процессы функционирования ИС (информационных систем); использование ресурсов ИС; деятельность персонала информационной службы на предприятии; порядок взаимодействия пользователей с системой с тем, чтобы в наибольшей степени затруднить или исключить возможность реализации угроз безопасности.

Организационно-административные мероприятия включают в себя:

- разработку правил обработки информации в ИС;

- совокупность действий при проектировании и оборудовании вычислительных центров и других объектов ИС (учет влияния стихии, пожаров, охраны помещений и т.п.);
- совокупность действий при подборе и подготовке персонала (проверка новых сотрудников, ознакомление их с порядком работы с конфиденциальной информацией, с мерами ответственности за нарушение правил ее обработки; создание условий, при которых персоналу было бы невыгодно допускать злоупотребления и т.п.);
- организацию надежного пропускного режима;
- организацию учета, хранения, использования и уничтожения документов и носителей с конфиденциальной информацией;
- распределение реквизитов разграничения доступа (паролей, полномочий и т.п.);
- организацию скрытого контроля за работой пользователей и персонала ИС;
- совокупность действий при проектировании, разработке, ремонте и модификации оборудования и программного обеспечения (сертификация используемых технических и программных средств, строгое санкционирование, рассмотрение и утверждение всех изменений, проверка на удовлетворение требований защиты, документальная фиксация изменений и т.п.).

К физическим мерам защиты относятся различные механические, электро- и электромеханические устройства или сооружения, специально предназначенные для создания физических помех на возможных путях проникновения и доступа нарушителей (турникеты, колючая проволока, кодовые замки, системы охранно-пожарной сигнализациии т.п.).

К аппаратно-программным мерам защиты относятся разные электронные устройства и особые программы, которые реализуют без помощи других либо в комплексе с другими средствами следующие методы защиты[7]:

- идентификацию и аутентификацию субъектов ИС;
- разграничение доступа к ресурсам ИС;
- контроль целостности данных;
- обеспечение конфиденциальности данных;
- аудит событий, происходящих в ИС;
- резервирование ресурсов и компонентов ИС

Следовательно, в современных условиях сохранность информационных ресурсов может быть обеспечена только комплексной системой защиты информации. Комплексная система защиты информации должна быть: непрерывной, плановой, целеустремленной, конкретной, активной, надежной. Система защиты информации должна опираться на систему видов собственного обеспечения, способную реализовать ее функционирование нетолько в повседневных условиях, но и в критических ситуациях.

Список использованных источников:

1. Закон Республики Казахстан от 16 ноября 2015 года № 401-V «О доступе к информации» (с изменениями и дополнениями по состоянию на 18.11.2022 г.) // Интернет ресурс: https://online.zakon.kz/Document/?doc_id=39415981 (дата обращения 03.12.2022 г.)
2. Семененко В.А. Информационная безопасность: Учебное пособие. 2-е изд., стереот. - М.: МГИУ, 2005. - 215 с.

3. Корнюшин, П.Н. Информационная безопасность / П.Н. Корнюшин, С.С. Костерин. – Владивосток: ТИДОТ ДВГУ, 2003. – 154 с.
4. Конев И. Р. Информационная безопасность предприятия / И. Р. Конев, А. В. Беляев. – СПб. : БХВ-Петербург, 2003. – 747 с.
5. И.В. Аникин, В.И. Глова, Л.И. Нейман, А.Н. Нигматуллина Теорияинформационной безопасности и методология защиты информации // Учебное пособие. Казань: Изд-во Казан. гос. техн. ун-та, 2008 с. 358.
6. Курушин В. Д. Компьютерные преступления и информационная безопасность / В.Д. Курушин, В. А. Минаев. – М. : Новый юрист. – 2012. – 256 с.
7. Гатчин Ю. А. Теория информационной безопасности и методология защиты информации / Ю. А. Гатчин, В. В. Сухостат. – СПб. : СПбГУ ИТМО, 2010. – 98 с.

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ИДЕНТИФИКАЦИИ ЧЕЛОВЕКА ПО ГОЛОСУ

Шорманов Т.С., Мазакова А.Т., Джомартова Ш.А., Мазаков Т.Ж.
e-mail: shormanov@gmail.com, tmazakov@mail.ru
Казахский национальный университет имени аль-Фараби, Алматы, Казахстан

Аннотация. *Статья посвящена применению нейронных сетей для идентификации человека по голосу. В работе рассмотрены возможные подходы к биометрической идентификации: векторное квантование (VQ), модели гауссовой смеси (GMM), скрытые марковские модели (HMM), а также проведен анализ преимущества и недостатков по применению нейронных сетей. В процессе исследования выявлены основные проблемы, которые могут возникнуть в ходе ее решения. Дано общее описание проблемы и формулировка задач исследования.*

Введение

В последнее десятилетие развитие нейронных сетей привело к значительным прорывам в области идентификации и синтеза речи благодаря применению новых подходов и методов. Поскольку речь является основным способом общения между людьми, а также наиболее естественной и эффективной формой обмена информацией между людьми. Поэтому вполне логично желание использовать данный способ для идентификации личности. Поскольку голосовой отпечаток уникален для своего владельца, его можно использовать для уникальной идентификации. Образец голоса – это индивидуальная характеристика, и невозможно найти двух людей с абсолютно идентичной аудио спектрограммой голоса.

Идентификации по голосу, строятся по единой схеме:

• производится регистрация пользователя и вычисляется шаблон;
• выбираются участки речевого потока для дальнейшего анализа;
• осуществляется первичная обработка сигнала;
• вычисляются первичные параметры;
• строится «отпечаток» (шаблон) голоса;
• производится сравнение «отпечатков» голосов и формируется решение по идентичности голосов или «близости» голоса к группе голосов [1].

Распознавание речи можно определить как процесс поэтапного кодирования аналогового речевого сигнала в цифровой и состоит из следующих основных шагов: дискретизации, квантования и кодирования.

Дискретизация сигнала заключается в выборке отсчетов аналогового сигнала с заданной периодичностью и определяйся частотой дискретизации. Поскольку человек способен слышать звуки в частотном диапазоне от 20 Гц до 20 кГц, максимальная частота дискретизации исходного звукового сигнала должна составлять не менее 40 кГц, то есть отсчеты желательно проводить с частотой 40 000 раз в секунду. В связи с этим часто аудиозапись проводится с максимальной частота дискретизации звукового сигнала 44,1 или 48 кГц.

Квантование по амплитуде представляет собой измерение мгновенных значений амплитуды дискретного по времени сигнала и преобразование его в дискретный по времени и амплитуде. На Рисунке 1 показан процесс квантования по уровню аналогового сигнала.

Рис. 1. Пример дискретизации и квантования звукового сигнала

Кодирование заключается в преобразовании в цифровой код квантованного сигнала. При этом точность измерения при квантовании зависит от количества разрядов. Если значения амплитуды записать с помощью двоичного числа и задать длину N разрядов, число возможных значений кодовых слов будет равно 2^N, если значение амплитуды отсчета представляется 16-разрядным кодовым словом, максимальное число градаций амплитуды (уровней квантования) составит 2^{16}=65536. Для 8-разрядного представления соответственно получим 2^8= 256 градаций амплитуды.

Методы исследования. Биометрическая система идентификации личности по голосу должна извлекает персональные физиологических или поведенческих характеристик, которыми обладает целевой человек, и использует эти шаблоны для сопоставления с заданными данными.

К настоящему моменту для решения данной задачи используются различные подходы. Ранее использовались подходы назовем их «традиционными» к извлечению речевых характеристик включают: векторное квантование (VQ), GMM (модели гауссовой смеси), HMM (скрытые марковские модели) и так далее.

Векторное квантование (VQ): векторное квантование, алгоритм для сопоставления признаков и создания моделей с использованием векторов признаков. Тип VQ – это алгоритм обучения без учителя. Он создает кластеры, которые представляют модели зарегистрированных спикеров.

Скрытая марковская модель (HMM): HMM может моделировать статистические вариации характеристик, давать статистическое представление о том, как динамик воспроизводит звук.

Модель гауссовой смеси (GMM): GMM статистическая модель для представления нормально распределенных субпопуляций внутри общей популяции. Гауссова смесь распределений параметризируется двумя типами значений — смесь весов компонентов и средних компонентов или ковариаций (для многомерного случая).

При этом ранее считалось, что GMM и HMM подходы обеспечивают более эффективное распознание голоса, но с внедрением нейронных сетей для целей выделения речевых характеристик, дало значительный прирост эффективности распознавания голоса и речи.

Использование нейронных сетей

Поскольку звук обычно хранится в виде последовательности 16-битных целых чисел, требуется генеративная модель для вывода 2^{16}= 65 536 вероятностей на временной шаг для синтеза необработанного звука. Кроме того аудио частота дискретизации, превышающая десять тысяч, приводит к чрезвычайно большой длине последовательности, что делает ее более сложной для синтеза необработанного звука. Для решения данной задачи требуется квантование речи для сжатия целочисленных значений и длины последовательности. Преобразование по закону μ может квантовать каждый временной шаг до 256 значений и восстанавливать необработанный звук. Данной подход широко используется в моделях генерации речи, таких как широко известная модель WaveNet. [2], но скорость логического вывода данного подхода будет низкая, поскольку длина последовательности не уменьшается, также помимо низкой скорости обработки данных данный подход характеризуется необходимостью использования языковой модели для получения и проверки не только голосовой характеристики и текста которые произносится то есть данная модель идентифицирует только голос.

В последнее время векторное квантование широко применяется в моделях речи с самоконтролем для извлечения голосовых признаков, таких как vq-wav2vec данный подход показан в работах [3] и моделью HuBERT [4]. Также в работе [5] показан подход с использованием кодов из моделей с самоконтролем, данный подход позволяет реконструировать контент(речь) то есть могут предугадывать слова, а скорость вывода выше, чем у WaveNet. Однако существенный недостаток данного подхода, что личные голосовые особенности говорящего никак не учитываются, а качество реконструкции низкое, то есть данная модель идентифицирует голос и текст, но плохо учитывает персональные данные такие как манера говорить или акцент.

Модель AudioLM [6] обучает языковые модели преобразования речи как на токенах k-средних из модели с самоконтролем, так и на акустических токенах из модели нейронного кодека, что приводит к более высококачественному преобразованию речи. По сравнению с традиционными подходами к аудиокодекам кодек на основе нейронных сетей значительно лучше работает при низких битрейтах. В первые подход с адаптацией к говорящему спикеру был представлен в работе [7]

Преимущества. По сравнению с другими методами метод квантования аудиокодека имеет следующие преимущества:

- Метод содержит информацию как о самом говорящем так и дополнительную акустическую информацию, которая может помочь в более точном идентификации говорящего при реконструкции по сравнению с моделью HuBERT [4].
- Существует готовый декодер кодека для преобразования дискретных токенов в сигнал без дополнительных усилий по обучению, по сравнению с методом на основе VQ.
- Уменьшение длины временных шагов для эффективности решения проблемы преобразования μ-закона [2].

На основании выше указанных преимуществ модель нейронного речевого кодека позволяет более эффективно работать с голосовыми записями. Из-за остаточного квантования в модели нейронного кодека токены имеют иерархическую структуру: токены из предыдущих квантователей восстанавливают акустические свойства, такие как идентичность говорящего, в то время как последовательные квантователи изучают мелкие акустические детали. Каждый квантователь обучен моделировать остаток от предыдущих квантователей. На практике данный подход позволяет идентифицировать не только голос конкретного человека, но и позволяет также более точно предсказать манеру разговора такие как паузы между словами и ударения.

Недостатки. Надежность синтеза: Во всех вышеперечисленных подходах наблюдаются случае, когда некоторые слова могут быть неясными, пропущенными или дублироваться. В основном это связано с тем, что часть фонемы к акустическому языку представляет собой авторегрессивную модель, в которой существует беспорядочное выравнивание и нет ограничений. Общей проблемой является тот факт, что большинство моделей построены на ограниченном количество аудио данных, то есть недостаточный охвате говорящих с различными акцентами. Поскольку обычно данные представляют собой набор аудиокниг, в которых большинство высказываний относится к литературному стилю.

Заключение

Биометрические идентификаторы такие как отпечатки пальцев, распознавание лица, сетчатка или голоса более надежны по сравнению с методами, основанными на паролях и токенах, поскольку они обладают такими характеристиками, как универсальность, уникальность, постоянство и возможность сбора. Поскольку каждый из вышеперечисленных биометрических идентификаторов имеет свои плюсы и минусы по сравнению с другими, для более высокой степени надежности желательно применять одновременно несколько способов идентификации. Применение нейронных сетей для биометрической идентификации по голосу дает значительное преимущество, поскольку позволяет проводит идентификацию не только физиологической характеристики такой как голос человека, но и дополнительно проверять индивидуальные голосовые поведенческие характеристики.

Работа выполнена за счет средств НИИ математики и механики при КазНУ имени аль-Фараби и грантового финансирования научных исследований на 2023–2025 годы по проекту AP19678157 «Разработка программно-аппаратного комплекса мониторинга состояния уровня заполняемости водоема».

Литература

1. Singh N., Agrawal A., Khan R.A., Голосовая биометрия: технология аутентификации на основе голоса, передовая наука, инженерия и медицина / Июль 2018
2. Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew W. Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. In The 9th ISCA Speech Synthesis Workshop, page 125. ISCA, 2016.
3. Alexei Baevski, Steffen Schneider, and Michael Auli. vq-wav2vec: Self-supervised learning of discrete speech representations. In ICLM, 2020.
4. Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29:3451–3460, 2021.
5. Kushal Lakhotia, Evgeny Kharitonov, Wei-Ning Hsu, Yossi Adi, Adam Polyak, Benjamin Bolte, Tu Anh Nguyen, Jade Copet, Alexei Baevski, Adelrahman Mohamed, and Emmanuel Dupoux. Generative spoken language modeling from raw audio. CoRR, abs/2102.01192, 2021.
6. Zalán Borsos, Raphaël Marinier, Damien Vincent, Eugene Kharitonov, Olivier Pietquin, Matthew Sharifi, Olivier Teboul, David Grangier, Marco Tagliasacchi, and Neil Zeghidour. Audiolm: a language modeling approach to audio generation. CoRR, abs/2209.03143, 2022.
7. Sercan Ömer Arik, Jitong Chen, Kainan Peng, Wei Ping, and Yanqi Zhou. Neural voice cloning with a few samples. In NeurIPS, pages 10040–10050, 2018.

Application of Nir Soft Utilities in Computer Forensics for Analysis and Recovery of Digital Evidence

Tashatov N.N., Bakhytov B.A.
e-mail: tash.nur@mail.ru, bigasyr@gmail.com
L.N. Gumilyov Eurasian National University, Kazakhstan

Abstract. *The growing ubiquity of digital technology has led to an increase in cybercrimes and the need for effective digital forensic tools. This article explores the applications of NirSoft utilities in computer forensics, focusing on their role in the analysis and recovery of digital evidence. NirSoft utilities provide a valuable set of tools for investigators, offering insights into system activities, passwords, and network interactions. This article discusses key NirSoft utilities, their usage, and their significance in digital forensic investigations*.

KeywordsDigital forensics, Investigation, Analyze Digital Evidence, Password Recovery, Network Analysis, System Analysis

1. Introduction

In an era characterized by an unprecedented reliance on digital technology, the landscape of criminal investigations has undergone a profound transformation. The advent of digital devices and the ubiquity of the internet have presented both new opportunities and challenges for law enforcement agencies, necessitating the evolution of investigative techniques. At the heart of this transformation lies the critical discipline of digital forensics, which has emerged as an indispensable component of modern investigations.

1.1 Background on the Importance of Digital Forensics in Modern Investigations

The importance of digital forensics in contemporary investigative practices cannot be overstated. As our society becomes increasingly digital-centric, a vast treasure trove of information is generated, stored, and transmitted through electronic devices and networks. Criminal activities, ranging from cybercrimes and fraud to terrorism and intellectual property theft, frequently leave behind a digital footprint. This digital trail holds the key to unraveling complex criminal networks, establishing guilt or innocence, and ensuring the due process of law.

Moreover, the scope of digital evidence extends beyond cybercrimes. It encompasses a wide array of cases, including those related to financial fraud, corporate malfeasance, domestic disputes, and even violent crimes. In these diverse scenarios, digital forensics plays a pivotal role in uncovering critical evidence, reconstructing timelines, and providing the necessary insights to support or refute claims in a court of law.

1.2 The Need for Specialized Tools to Collect and Analyze Digital Evidence

As the significance of digital evidence continues to grow, so too does the demand for specialized tools and methodologies tailored to its collection, preservation, and analysis. Unlike

traditional forms of evidence, such as physical documents or tangible objects, digital evidence presents unique challenges due to its intangible, volatile, and easily alterable nature.

Specialized digital forensic tools are essential to address these challenges effectively. They provide investigators with the means to access, acquire, and examine digital artifacts, ranging from emails and text messages to log files and system data. These tools are designed not only to extract data but also to maintain the integrity and authenticity of evidence throughout the forensic process. Moreover, we examine the emergence of NirSoft utilities as valuable assets within the digital forensic toolkit, shedding light on their contributions to the analysis and recovery of digital evidence.

Digital forensics relies heavily on the availability of powerful and versatile tools capable of extracting, analyzing, and interpreting digital evidence. NirSoft, a reputable software developer, offers a comprehensive suite of utilities tailored to meet the diverse needs of computer forensics investigators.

2. NirSoft Utilities Overview.

NirSoft, a reputable software developer, offers a comprehensive suite of utilities tailored to meet the diverse needs of computer forensics investigators. In this section, we provide an introduction to NirSoft and offer a brief overview of key NirSoft utilities that have become indispensable in the realm of digital forensic investigations.

2.1 Introduction to NirSoft

NirSoft, founded by Nir Sofer, has established itself as a leading provider of small, yet incredibly effective, Windows-based utilities. These utilities cover a wide spectrum of functions, from system diagnostics and password recovery to network monitoring and data analysis. What sets NirSoft apart is its commitment to creating free, portable, and user-friendly software, making it accessible to a global community of users, including digital forensics practitioners.

2.2 Key NirSoft Utilities for Computer Forensics

2.2.1 NirSoft Password Recovery Utilities:

- **Mail PassView:** Designed for recovering email passwords from popular email clients.
- **WebBrowserPassView:** Allows the retrieval of passwords stored in web browsers.
- **WirelessKeyView:** Aids in recovering wireless network keys.

2.2.2 NirSoft Network Analysis Utilities:

- **WirelessNetView:** Monitors nearby wireless networks and provides essential information.
- **CurrPorts:** Displays a list of open network ports and associated processes.
- **SmartSniff:** Captures and analyzes TCP/IP packets in real-time.

2.2.3 NirSoft System Analysis Utilities:

- **USBDeview:** Provides information about USB devices connected to a computer.

- **USBLogView:** Logs USB device activity for later analysis.
- **RecentFilesView:** Lists recently opened files and their details.

2.2.4 NirSoftMiscellaneous Utilities:

- **AccessEnum:** Scans file system and registry permissions.
- **WinPrefetchView:** Examines Windows Prefetch files.
- **IECacheView:** Retrieves information about cached files in Internet Explorer.

These are just a few examples of the extensive range of NirSoft utilities that have found application in computer forensics. Each utility serves a specific purpose, and collectively they provide investigators with valuable insights into digital artifacts, system activities, and user interactions.

3. Examples of practical application of Nirsoft utilities for Password Recovery

3.1 NirSoft Utilities for Password Recovery.

Figure 1 shows the extraction of saved passwords from the Chrome browser using **ChromePass**utility.

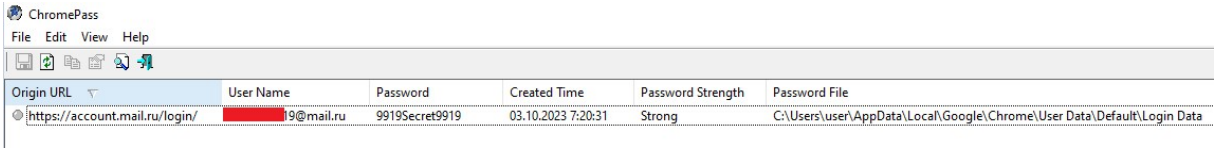


Fig. 1 ChromePass

Figure 2 shows the extraction of passwords for Wi-Fi networks to which the device was connected using **WirelessKeyView** utility.

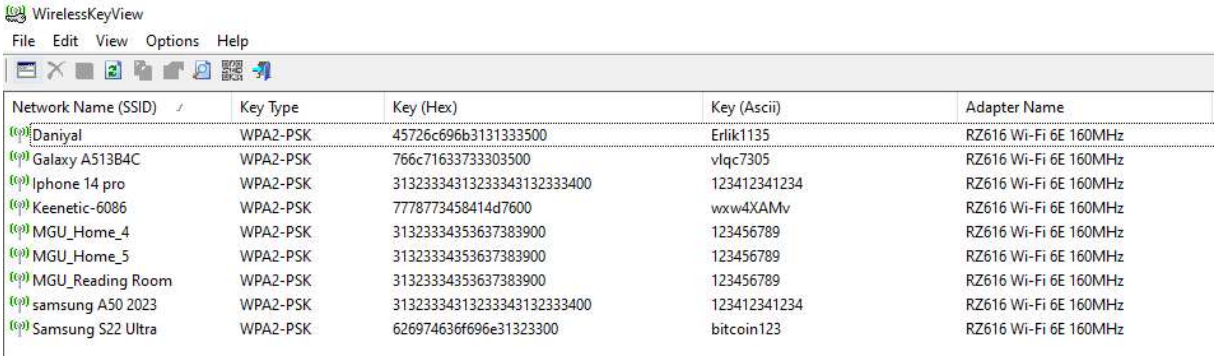


Fig. 2 WirelessKeyView

4.Ethical and Legal Considerations for Password Recovery

Password recovery, while indispensable, raises ethical and legal concerns:

- **Ethical Considerations:** Investigators must obtain appropriate authorization and informed consent before attempting password recovery. Unauthorizedaccesstoaccountsorsystemsviolatesethicalstandards.
- **Data Privacy:** Respecting data privacy rights is paramount. Recovered passwords may grant access to private information, necessitating stringent data protection measures.
- **Legal Compliance:** Password recovery must comply with local and international laws. In some jurisdictions, unauthorized access to accounts, even during an investigation, may have legal repercussions.

Statistics regarding password recovery success rates and legal challenges vary widely depending on the specific case, tools used, and jurisdiction. These factors underline the need for forensic experts to exercise caution, adherence to ethical guidelines, and strict legal compliance when employing NirSoft utilities for password recovery in digital forensic investigations.

5. Challenges using nirsoft utilities

The utilization of NirSoft utilities in computer forensics offers a plethora of advantages; however, it is essential to acknowledge and navigate the challenges that accompany their deployment in investigative practices.

5.1 Addressing Challenges and Limitations

While NirSoft utilities excel in many aspects of digital forensics, investigators must be cognizant of their inherent challenges and limitations:

- **Data Volatility:** NirSoft utilities often operate on live systems, which can introduce data volatility concerns. The act of extracting data or executing utilities may alter the state of the system or inadvertently delete crucial evidence.
- **Limited Mac and Linux Support:** NirSoft utilities are primarily designed for Windows environments, which may limit their application in cross-platform investigations involving Mac or Linux systems.
- **Anti-Forensic Measures:** Some malicious actors employ countermeasures to hinder forensic analysis. NirSoft utilities may encounter obstacles when facing encrypted or obfuscated data.
- **Legal Admissibility:** The admissibility of evidence obtained using NirSoft utilities may be subject to scrutiny in legal proceedings. Investigators must be prepared to demonstrate the reliability and integrity of the tools and techniques employed.

7. Best Practices and Recommendations

Effectively harnessing the power of NirSoft utilities in digital forensics requires a deep understanding of these tools and a commitment to best practices. In this section, we offer a set of recommendations and best practices for investigators and forensic analysts, ensuring the efficient and ethical utilization of NirSoft utilities in their investigative endeavors.

7.1 Best Practices for Using NirSoft Utilities Effectively in Digital Forensics

1. Familiarize Yourself with NirSoft Tools: Thoroughly understand the functionalities and applications of NirSoft utilities relevant to your investigation. Regularly update your knowledge as new tools and updates are released.

2. Use Portable Versions: Whenever possible, opt for portable versions of NirSoft utilities. These versions do not require installation, minimizing the risk of altering evidence on the target system.

3. Document Every Step: Maintain meticulous documentation of your investigative process. Record the tools used, commands executed, and their outcomes. A comprehensive trail of your actions enhances transparency and credibility.

4. Maintain Data Integrity: Ensure data integrity by utilizing read-only methods and verified copies of digital evidence. Avoid altering or tampering with original data during the investigation.

5. Verify Results: Cross-verify findings obtained through NirSoft utilities with other forensic methods or tools to enhance reliability and reduce the risk of false positives or negatives.

6. Stay Updated: NirSoft utilities are continually evolving. Stay updated with the latest versions and enhancements to benefit from improved features and bug fixes.

7. Adhere to Legal and Ethical Guidelines: Strictly adhere to legal and ethical guidelines, including obtaining proper authorization and consent, maintaining the chain of custody, and respecting data privacy regulations.

7.2 Recommendations for Investigators and Forensic Analysts

1. Training and Certification: Pursue training and certification in digital forensics to acquire the necessary skills and knowledge for effective tool usage.

2. Collaboration: Foster collaboration with other forensic experts and law enforcement agencies to share knowledge and experiences related to NirSoft utilities and other digital forensic tools.

3. Maintain a Toolkit: Build a comprehensive toolkit that includes a variety of NirSoft utilities tailored to different investigative scenarios. Customize your toolkit based on your specialization.

4. Practice Test Environments: Create controlled test environments to practice using NirSoft utilities without compromising real case data. This helps in refining your skills.

5. Continuous Learning: Engage in continuous learning and networking within the digital forensics community. Attend conferences, seminars, and webinars to stay updated on industry trends and best practices.

6. Legal Consultation: In cases involving complex legal issues, consult with legal experts to ensure that your investigative practices and the use of NirSoft utilities align with current legal standards.

By following these best practices and recommendations, investigators and forensic analysts can leverage NirSoft utilities as valuable assets in their digital forensic toolkits. These guidelines not only enhance the effectiveness of investigations but also ensure adherence to ethical and legal standards, ultimately contributing to the integrity of the digital forensic profession.

References

1. Smith, J. A Comprehensive Guide to Digital Forensics / J. Smith // International Journal of Cybersecurity. – London: Springer, 2021. – Vol. 5, N. 3. – P. 127-142.
2. Johnson, R. Digital Forensics and Cybersecurity: Investigating High-Tech Crimes / R. Johnson // Journal of Computer Security. – New York: McGraw-Hill, 2020. – Vol. 3, N. 1. – P. 55-68.
3. Brown, L. Cybercrime Investigations: Methods and Techniques / L. Brown // Digital Forensics Quarterly. – San Francisco: Pearson, 2019. – Vol. 8, N. 4. – P. 213-228.
4. Clark, S. Mobile Device Forensics: Unlocking the Evidence / S. Clark // Journal of Digital Investigation. – Boston: Jones & Bartlett, 2018. – Vol. 12, N. 2. – P. 79-94.
5. Davis, P. Forensic Analysis of Computer Data: Guidelines and Procedures / P. Davis // International Journal of Digital Evidence. – Chicago: Wiley, 2017. – Vol. 15, N. 3. – P. 101-116.
6. White, E. Network Forensics: Detecting and Investigating Cybercrime Incidents / E. White // Digital Evidence and Electronic Signature Law Review. – Toronto: CRC Press, 2016. – Vol. 9, N. 2. – P. 45-58.
7. Thomas, C. The Art of Memory Forensics: Detecting Malware and Threats in Windows, Linux, and Mac Memory / C. Thomas // Journal of Forensic Sciences. – Seattle: Syngress, 2015. – Vol. 22, N. 4. – P. 189-203.
8. Garcia, M. Investigating Digital Crime / M. Garcia // International Journal of Computer Science and Information Security. – Amsterdam: Springer, 2014. – Vol. 11, N. 1. – P. 33-48.
9. Anderson, D. File System Forensic Analysis / D. Anderson // Journal of Digital Forensics, Security, and Law. – Cambridge: Addison-Wesley, 2013. – Vol. 8, N. 3. – P. 147-162.
10. Patel, N. Cloud Forensics: Investigating Digital Evidence in Cloud Environments / N. Patel // Digital Investigation. – London: Academic Press, 2012. – Vol. 6, N. 4. – P. 221-236.
11. Rogers, A. Mobile Phone Forensics: Principles and Practice / A. Rogers // Journal of Digital Forensics, Security, and Law. – San Francisco: Wiley, 2011. – Vol. 6, N. 2. – P. 83-98.
12. Williams, S. Investigative Computer Forensics: The Practical Guide for Lawyers, Accountants, Investigators, and Business Executives / S. Williams // International Journal of Cybercrime and Digital Forensics. – Oxford: McGraw-Hill, 2010. – Vol. 2, N. 4. – P. 169-184.
13. Evans, B. Cyber Crime Investigations: Bridging the Gaps between Security Professionals, Law Enforcement, and Prosecutors / B. Evans // Journal of Digital Investigation. – Boston: Elsevier, 2009. – Vol. 5, N. 1. – P. 25-40.
14. Harris, D. Computer Forensics and Investigations / D. Harris // Digital Forensics Quarterly. – New York: Prentice Hall, 2008. – Vol. 7, N. 3. – P. 123-138.

15. Lee, T. Advanced Forensic Digital Imaging / T. Lee // International Journal of Cybersecurity Research. – London: CRC Press, 2007. – Vol. 4, N. 2. – P. 61-76.
16. Casey, E. Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet. – https://www.sciencedirect.com/book/9780121631048/digital-evidence-and-computer-crime (Accessed August 15, 2023).
17. National Institute of Standards and Technology (NIST). Computer Forensics. – https://www.nist.gov/topics/computer-forensics (Accessed August 20, 2023).
18. Digital Forensics Magazine. – https://www.digitalforensicsmagazine.com/ (Accessed August 25, 2023).
19. Cybercrime Magazine. – https://cybersecurityventures.com/cybercrime-magazine/ (Accessed August 30, 2023).
20. The Sleuth Kit and Autopsy. Open Source Digital Forensics Tools. – https://www.sleuthkit.org/ (Accessed September 5, 2023).

USER AUTHENTICATION AND DATA MANAGEMENT WITH USING BLOCKCHAIN SYSTEMS

A.Sh. Barakova, O.A. Ussatova
Institute of Information and Computing Technologies, Almaty, Kazakhstan,
Al-FarabiKazakh National University, Almaty, Kazakhstan
balia_79@mail.ru
uoa_olga@mail.ru
ORCID ID: *https://orcid.org/0000-0002-5276-6118*
ORCID ID: *https://orcid.org/0000-0002-0904-745X*

Annotation. *The development of technology has led to innovations in equipment, and the number of high-tech devices is growing every day. These devices are deployed via the Internet of Things. Internet of Things devices have security-related limitations at both the software and hardware levels. Secure user interfaces allow you to overcome limitations at the software level, for example, external user interfaces can be easily accessed through public and private networks. Interfaces are connected to a localized storage for*
storage of data generated by Internet of Things devices. Localized storage deployed in a closed environment, connected to Internet of Things devices, is more efficient from a security point of view than online servers.
Blockchain has emerged as a technology or method with the capabilities to provide secure administrative authentication and access to internet of things devices and their computing data decentralized in a way with high reliability and stability. This article offers methods of authentication of devices, end users and transactions using algorithms built into the blockchain.
Keywords: *blockchain; authentication; IoT; Internet of things*

Introduction

With the rapid growth in the number of smart gadgets and high-speed networks that are used to communicate with these smart devices, the Internet of Things (Internet of things, IoT) has attracted a lot of attention of researchers over the past few years [1]. These embedded devices or IoT
devices connect via public or private networks, gain remote access, and provide the desired functionality. Public and private networks use network protocols to exchange information and interact between IoT devices [2].

The Internet of Things is a resource–intensive device that is not able
to protect itself from malicious attacks, such as a man-in-the-middle attack, DOS attacks, etc., and can be easily hacked by intruders. To ensure It is important to implement proper methods for the user, as well as device authentication and computational transaction, to ensure that IoT devices are secure in all respects [3]. There is also a need for a system to ensure interaction between end
users and Internet of Things devices [4]. Numerous authentication methods [5] have been developed and deployed to ensure the security of Internet of Things devices, but they are all based on a centralized architecture and depend on a centralized management body, such as a database or system servers. The Control Center checks end users, IoT system devices, and

records communications between end users and IoT devices using various protocols. Mutual authentication,
certificate-based authentication, and token–based authentication are all these are centralized authentication methods. These methods have many disadvantages, such as high transactional computing costs, centralized trusted third parties, a single point of failure, lack of privacy, and a high probability of hacking. Since these methods rely on a trusted third party in this way
, double dependency problems arise [6]. Figure 1 shows the problem of double dependency.

The disadvantage of direct user authentication is that they only authenticate users with a username or password. Thus, the security of this authentication method is quite low System owners can send the same transaction to multiple
system participants.

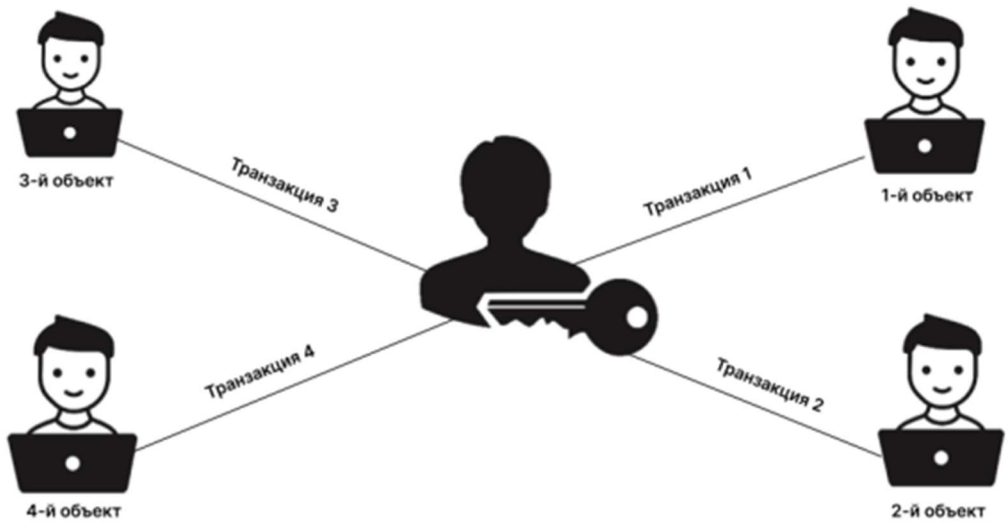


Figure 1. Authentication structure

The open authentication method for user authentication is a centralized method in which all calculations are performed using a central identifier.The group authentication method splits the authentication method on several nodes, but the group authentication method does not use hash key values, so each object in the system can perform some malicious actions. All the above methods have many disadvantages. They
only authenticate users and do not authenticate devices. internet of things. They also use centralized
authentication methods. With the centralized method, all computational data is stored on the server. Objects in the system rely too heavily on the server or central authority to perform the desired access transactions. If the centralized authority does not have sufficient reliability in terms of security, then it may violate the security of the entire system. To reduce the disadvantages of a centralized (trusted thirdparty a) authentication of Internet of Things devices, it is proposed a decentralized authentication scheme for the end user, Internet of Things devices and transactions using algorithms that provide blockchain technology. The proposed system provides the ability to authenticate the end user and Internet of Things devices. The proposed system also makes it easier for end users to match secure communications with Internet of Things devices, while ensuring

security without any requirements for centralized identification. The main purpose of the architecture of the system proposed in this article – ensure the security of Internet of Things devices at the hardware level. To
achieve this goal, it is necessary to use blockchain technology.

System architecture

The proposed system architecture has five main objects with access to web algorithms via the Internet: an administrator, end users who provide a direct connection to the system, a blockchain algorithm, a database (MySQL server) containing end user data and IoT devices. A more detailed description of the objects is given
below. During registration, unique hash addresses (public and private keys) are assigned to Internet of Things devices and end users. How Internet of Things devices and end users are registered on the
Internet. The administrator and databases are also part of the system. The detailed architecture of the system is shown in Fig. 2.

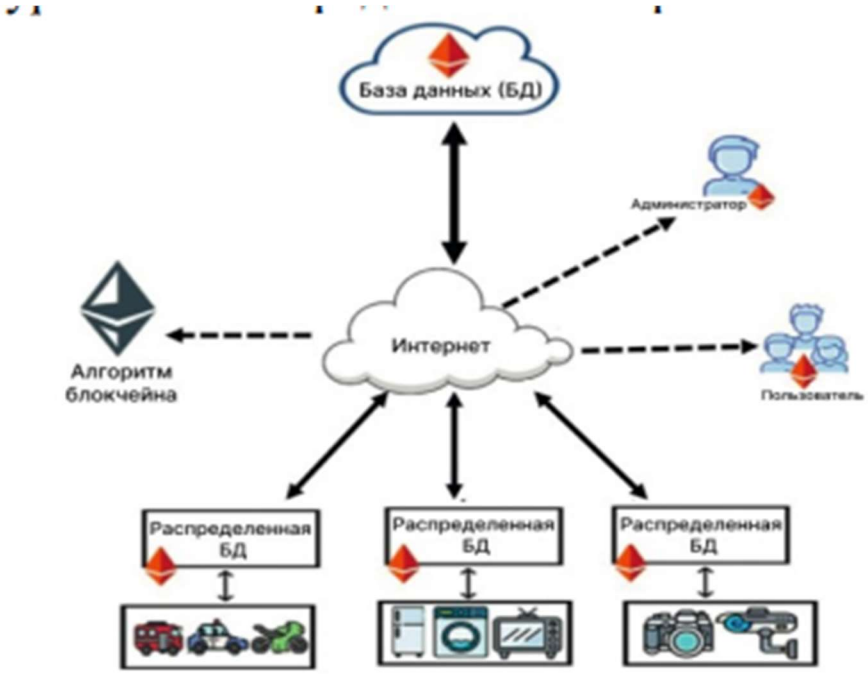


Figure 2. The detailed architecture of the system

Conclusion

This article proposes an architectural design of logical algorithms based on the blockchain using the hash key values algorithm for end users, Internet of Things devices and transaction authentication in a distributed manner without the intervention of third parties. Authentication of a large number of end users, IoT devices and transactions leads to data deployment using a MySQL server, which frees end users and IoT devices from the computational complexity of authentication in the blockchain network. If the distributed registry is not updated after each transaction, data loss occurs, which creates malware between the blocks of the distributed registry, since each block is linked to the previous block through the previous
value of the public hash key to create a chain. In the future, a decentralized architecture may be implemented to ensure the security of Internet of Things devices on the server side.

Literature

1. Almadhun R., Kadada M., Alhemeiri M., Alshehi M. and Salah K.
Authentication scheme for users of Internet of Things devices using fog nodes with blockchain support 2018, in
Proceedings of the 15th IEEE//ACS International Conference on Computer Technologies 2018 Systems and Applications (AICCSA), 2018. Aqaba, Jordan. – pp. 1-8.
2. Lally G. and Sgandurra D. On the way to creating a framework for consistent testing of the security of Internet
of Things devices, in the proceedings of the First International Seminar on ETAA, 2018, Barcelona, Spain, September.
3. Park N. Mutual authentication scheme in the technology of the secure Internet of Things for a comfortable lifestyle, Sensors (Switzerland), 2015. – vol. 16. – pp. 1-16.
4. Mishra S. and Paul A. Critical analysis of attack detection schemes in IoT and open calls, in Proceedings of the IEEE International Conference on Computing 2020,
Energy and Communication Technologies (GUCON). 2020. Noida, India, October. – p. 57-62.
5. Luu L., Chu D.H., Olikel H., Saxena P. and Hobor A. Making smart contracts smarter, in materials of ACM Proceedings // ACM Conference on Computer and
Security of communications, 2016. – pp. 254-69.
6. Zaramenskikh E.P. Internet of Things. Research and field of application. – M.:
IN-ta-M, 2016. – 188 p.

НЕЛИНЕЙНОЕ ПРЕОБРАЗОВАНИЕ ЛЕГКОВЕСНОГО АЛГОРИТМА ШИФРОВАНИЯ LBC

Капалова Н.А., Алгазы К.Т., Хаумен А.
e-mail: kapalova@ipic.kz, kunbolat@mail.ru, haumen.armanbek@gmail.com
Институт информационных и вычислительных технологий КН МНВО, Алматы, Казахстан

Аннотация. *В статье представлен новый легковесный симметричный алгоритм шифрования данных под названием LBC. Главной целью разработки легковесного алгоритма шифрования является необходимость создания отечественного алгоритма с ограниченными ресурсами. В Казахстане в сфере защиты информации применяются в основном зарубежные аппаратно-программные средства, которые являются прозрачными для их разработчиков. Поэтому разработка отечественных средств криптографической защиты информации, в том числе отечественных легковесных алгоритмов шифрования является актуальным вопросам. Разработанный легковесный алгоритм имеет простую схему реализации, и в тоже время все преобразования, используемые в данном алгоритме, показывает хорошие криптографические свойства. Одним из компонентов, который играет важную роль в криптостойкости, является S-блок, основные характеристики которого приведены в данной работе. Также, исследованы свойства нелинейности таблицы подстановок с вычислением соответствующих значений. На основании проведенных работ установлено, что предложенный алгоритм шифрования эффективен для обеспечения хорошего лавинного эффекта и бинарная последовательность, полученная после зашифрования, близка к случайной. Результаты исследований выявили хорошие криптографические свойства данного алгоритма.*

Ключевые слова: *алгоритм шифрования, легковесный алгоритм, криптографические преобразования, таблица подстановок, криптостойкость.*

Введение

В последнее годы наблюдается тенденция массового перехода от интернета персональных компьютеров к интернету вещей. В жизнь общества все больше входит электронных приспособлений, направленных на улучшение жизни и так или иначе взаимодействующих с интернетом. Мобильные телефоны, электронные ключи, бытовая техника - все они снабжены процессором и функционируют в составе большой сети Интернет. С растущим спросом легкие шифры привлекают все больше внимания. В 2018 году Национальный институт стандартов и технологий (NIST) опубликовал призыв к облегченным шифрам, принимая заявки на упрощенное аутентифицированное шифрование и хеширование для сред с программными или аппаратными ограничениями. В связи с этим разработка и анализ облегченных криптографических алгоритмов приобрели популярность среди исследователей. Эти исследования были сосредоточены на разработке легкого блочного шифра, анализе безопасности и оценке производительности и т.д. Такие легковесные алгоритмы предназначены для снижения потребления энергии и памяти для выполнения требований приложений с ограниченными ресурсами, таких как RFID и устройства Интернета вещей (IoT).

Существует достаточно много реализаций легковесной криптографии, программных и аппаратных. У них разные, а иногда и противоположные характеристики. Например, перестановочный бит в аппаратной реализации практически ресурсонезависим, тогда как в программной реализации он может значительно замедлить работу. Кроме того, большая таблица подстановок часто хорошо реализуется в программном обеспечении, в то время как аппаратные реализации могут быть относительно дорогими.

Блочный шифр считается достаточно безопасным для практического использования после того, как он подвергся обширному криптоанализу третьей стороной. Блочные шифры часто встречаются в качестве криптографических примитив, которые применяются в таких протоколах, как Open PGP [1], Transport Layer Security [2] и др. Эти алгоритмы шифрования с симметричным ключом, которые используют один секретный ключ, выполняют шифрование и расшифрование. Блочные шифры могут использоваться в качестве строительных блоков для создания генераторов случайных чисел, кодов аутентификации сообщений. Также ожидается, что блочные шифры будут безопасными в постквантовом мире, просто выбрав более длинные ключи и размеры рабочих блоков [3]. Блочные шифры быстро стандартизируется и упрощает как программную, так аппаратную реализацию. Следовательно, в информационной безопасности блочный шифр имеет основные значения и является основным шифром с ограниченными ресурсами. В последнее время, с развитием встроенных систем с ограниченными ресурсами, встроенные блочные шифры стали важными примитивами для криптографических приложений, таких как беспроводные сенсорные сети (WSN), метки RFID и сенсорные узлы. Облегченный блочный шифр предназначен для рассмотрения его применения на целевой платформе. Среди этих облегченных блочных шифров некоторые предназначены для аппаратной реализации, некоторые - для программного обеспечения, а некоторые полностью разработаны как для аппаратного, так и для программного обеспечения. При разработке аппаратного обеспечения его цель состоит в том, чтобы уменьшить необходимое количество логических элементов. Для разработки программного обеспечения облегченный блочный шифр рассматривает типичную платформу реализации как микропроцессор и его цель состоит в том, чтобы сэкономить больше места для хранения, исключить вычисления и максимизировать пропускную способность. Программная реализация позволяет нам приблизительно оценить потребляемую мощность, умножив время обработки на среднюю потребляемую мощность устройства [4].

Легковесные блочные шифры обычно имеют размер блока от 32 до 64 бит с размером ключа 80 или 128 бит. К настоящему времени было предложено большое количество алгоритмов шифрования. В случае ограниченных ресурсов каждый из них имеет различные преимущества в эффективности программного и аппаратного обеспечения. Эти алгоритмы включают PRESENT [5], Piccolo [6], LED [7] и другие. Компромисс между безопасностью, стоимостью и эффективностью является ключевым вопросом, который необходимо решить при разработке легковесного блочного шифра. В частности, длина ключа облегченного блочного шифра, общее количество раундов и алгоритм структуры влияют на безопасность, стоимость и эффективность, но наиболее существенную роль играет безопасность самого криптографического алгоритма. Однако в процессе проектирования между различными алгоритмами можно выделить некоторые с одинаковой или похожей операцией, но алгоритмы различны, что занимает значительные ресурсы области служебных данных, что приводит к высокой стоимости аппаратной реализации и значительному снижению эффективности

производительности. Алгоритм блочного шифрования по структуре представляет две группы: SP сеть и сеть Фейстеля. Преимущество применения сети SP заключается в уровне диффузии. Его циклическая функция может изменять все блочные сообщения в итеративном цикле; следовательно, его безопасность относительно высока. Для решения этих проблем шифры в традиционных структурах Фейстеля обычно требуют много раундов, таким образом, это увеличивает потребление энергии.

Целью данной работы является разработка отечественного легковесного алгоритма шифрования данных, а также изучение и оценка его надежности с использованием методов криптоанализа.

Для достижения этой цели были поставлены следующие задачи:

– разработать легковесный алгоритм симметричного блочного шифрования, который отвечает основным требованиям к криптографическому преобразованию и обеспечивает высокую производительность и гибкость в программно-аппаратной реализации;

– провести исследование и оценить надежность разработанного алгоритма шифрования методами статистического и криптографического анализа.

Нелинейное преобразование в алгоритмах шифрования

В каждом современном алгоритме шифрования используется нелинейное преобразование в виде таблицы подстановок. Доказано, что оно является сильным криптографическим примитивом против линейного и дифференциального криптоанализа [8]. В легковесных блочных алгоритмах для нелинейных преобразований также используются табличные подстановки – *S*-блоки. Но по требованию к легковесным алгоритмам S-блоки не должны занимать большой объем памяти (ROM и RAM). Чтобы хранить таблицу, заменяющую 8-битовые фрагменты данных, необходимо 256 байт, а для хранения 4-битного S-блока требуются всего 16 байтов [9]. Поэтому в легковесной криптографии в основном используются 4-битные S-блоки. К ним относятся такие алгоритмы, как PRESENT и SERPENT.

Большинство современных шифров содержат один или несколько нелинейных компонентов, что имеет важное значение для стойкости шифра. Среди всего множества S-блоков особенно интересны биективные. Для повышения стойкости блочного шифра к линейному и дифференциальному криптоанализу применяют два основных подхода:

1. увеличение числа активных S-блоков;
2. использование S-блоков с сильными криптографическими свойствами [10].

Процедура S-блочного преобразования в шифре обеспечивает рассеивание и перемешивание байтов открытого текста, что в свою очередь значительно повышает криптостойкость шифртекста к реализации различных криптоаналитических атак [11-13]. Важными криптографическими свойствами, которыми должен обладать S-блок, являются сбалансированность, высокая нелинейность, высокая алгебраическая степень, низкая дифференциальная однородность и низкая автокорреляция.

Во многих известных работах по анализу S-блоков используются криптографические булевые функции. При этом каждый S-блок представлен компонентами булевых функций, характеризующих эффективность нелинейной замены.

Описание алгоритма

Алгоритм LBC-3 разработан для шифрования данных блочного типа длиной 64 бит с ключом 80 бит. LBC-3 при шифровании выполняет 20 раундов. Каждый раунд включает в себя 4 вида преобразования:

- преобразование S,
- преобразование RL,
- преобразование L,
- преобразование K.

Процесс шифрования

При шифровании исходный блок открытого текста делится на 4 подблока по 16 бит. Шифрование начинается с добавления первых 64 битов мастер ключа на исходный текст (рисунок 1). Далее выполняются раундовые преобразования.

Если обозначить раундовый процесс шифрования как E, то получим:

$$E(A) = [K(L(S(A_0)||RL(S(A_0)) \oplus S(A_1)||S(A_2)||S(A_3)))]_{rc} = B, \qquad (1)$$

где $A = \{A_0, A_1, A_2, A_3\}$ – 64 битный исходный текст, $\{A_i\}$– 16 битные подблоки, B – 64 битный шифртекст, rc – количество раундов, $\oplus$ - сложение по модулю 2.
Функция RL выполняется только для подблока $\{A_0\}$. То есть $\{A_1\} = RL(A_0) \otimes A_1$.

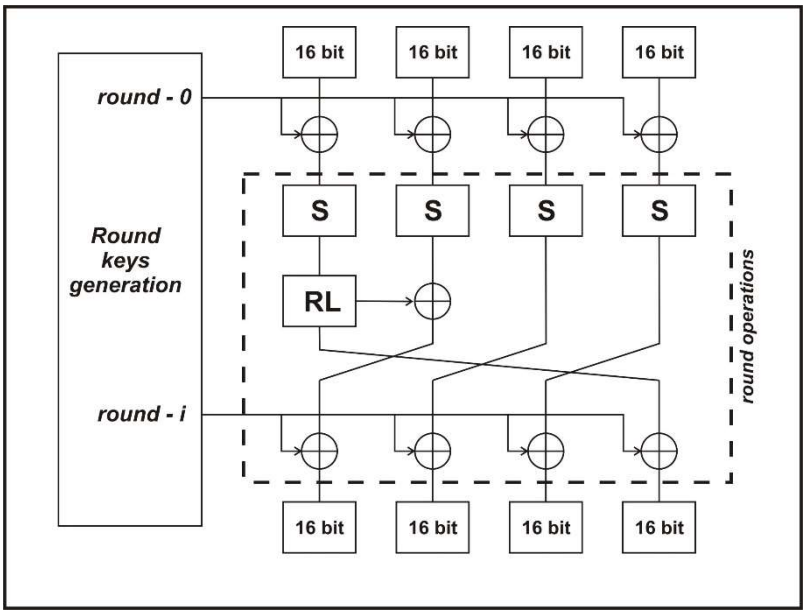


Рис. 1 Общая схема алгоритма LBC

Преобразование S

Преобразование S было сконструирован для замены битов блока на другие биты с помощью таблиц (S-блок). Это нелинейное преобразование алгоритма использует операции с полубайтами (4 бита). К каждому полубайту применяется нелинейная биективная подстановка, задаваемая одномерным массивом, состоящим из 16 элементов (Таблица 1).

Таблица 1. Таблица замены

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
S(X)	9	2	A	4	0	6	7	D	5	1	8	3	E	F	B	C

16-битный подблок поступает на вход преобразования S и разбивается на 4 группы по 4 бита. Каждый полубайт входного блока – это индекс значения, находящегося в таблице подстановок (Таблица 1). Таким образом, на выходе преобразования S будет получен

набор полубайтов, находящихся по соответствующим индексам в заданной таблице подстановок:

$$S(A) = S(a_0)||S(a_1)||S(a_2)||S(a_3), \quad S: V_4 \rightarrow V_4, \qquad (2)$$

где A – входные 16 бит, a_i – полубайты подблока.

S-блок генерирован путем получения обратного элемента по модулю $x^4 + x^3 + 1$. После этого к каждому элементу таблицы было применено следующее аффинное преобразование:

$$\left\|\begin{matrix} b'_3 \\ b'_2 \\ b'_1 \\ b'_0 \end{matrix}\right\| = \left\|\begin{matrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{matrix}\right\| * \left\|\begin{matrix} b_3 \\ b_2 \\ b_1 \\ b_0 \end{matrix}\right\| + \left\|\begin{matrix} 1 \\ 0 \\ 0 \\ 1 \end{matrix}\right\| \qquad (3)$$

где b_i – биты полубайта, b'_i - новые биты полубайта.

Преобразование RL

Линейным преобразованием алгоритма LBC-3 является циклический сдвиг битов и подблоков на определенную позицию. 64 бита входного блока делятся на 4 подблока по 16 битов. Линейное преобразование на уровне подблока осуществляется с помощью функции *RL*, которая применяется только к первому подблоку (Рисунок 1). Результат функции *RL* суммируется со вторым подблоком по модулю 2 (операция XOR). Функция *RL* имеет следующий вид:

$$RL(a) = a \oplus (a \lll 7) \oplus (a \lll 10), \quad a \in F_2^{16}, \qquad (4)$$

где «⋘» - оператор циклического сдвига битов подблока влево.

Преобразование L

Это линейное преобразование осуществляется над целым блоком. При этом подблоки циклически сдвигаются влево на одну позицию, т.е. на первой позиции окажется второй подблок, на месте второго – третий подблок, на месте третьего – четвертый подблок, а первый подблок переместится на четвертую позицию (Рисунок 1).

Преобразование K

После циклических сдвигов подблоков к блоку данных добавляются раундовые ключи по модулю 2. Раундовый ключ, состоящий из 64-х битов, делится на 4 подключа по 16 бит и суммируется с соответствующими подблоками данных.

Процесс расшифрования

Процесс расшифрования выполняется в обратном порядке. При этом вместо преобразований *S* и *L* используются обратные преобразования S^{-1} и L^{-1}, а преобразования *RL* и *K* остаются такими же, как и в процессе шифрования:

$$D(B) = [S^{-1}(L^{-1}(K(B_0)||RL(K(B_0))\oplus K(B_1)||K(B_2)||K(B_3)))]_{rc} = A \qquad (5)$$

где $\{B_i\}$ – подблоки шифртекста. A – расшифрованный исходный текст. Функция *RL* выполняется только для подблока $\{B_0\}$. То есть, $\{B_1\} = RL(B_0) \oplus B_1$.

Преобразование S^{-1}

Это обратное преобразования к табличной замене полубайтов. Оно осуществляется с помощью таблицы 2.

Таблица 2. Обратная таблица замены

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**

S(X)	4	9	1	B	3	8	5	6	A	0	2	E	F	7	C	D

Алгоритм LBC имеет секретный ключ длиной 80 бит. Из этой последовательности битов генерируются раундовые ключи алгоритма. Процесс генерации ключей состоит из нескольких преобразований. Алгоритм генерации раундовых ключей описаны в работе [14].

Исследование свойств S-блока

В алгоритме LBC используется 4-битный S-блок, который содержит 16 уникальных и различных элементов, которые варьируются от 0 до F в шестнадцатеричном формате (Таблица 1). Рассмотрим показатели эффективности S-блока, влияющие на уровень стойкости разработанного шифра к различным криптоатакам. В таблице 3 приведены свойства используемого S-блока.

Таблица 3. Характеристика используемого S-блока

Свойства		Значение
Сбалансированность		да
Вес Хемминга		8
Расстояние Хемминга		8
Нелинейность	максимум	12
	минимум	4
Корреляционный значение	максимум	8
	минимум	-8
$\lvert AC \rvert_{max}$		8
SSI		640
SAC		нет
CI		нет
t-устойчивость		нет
Циклическая структура		1: - (0; 4; 16)

При разработке стойких систем шифрования сложно достичь оптимальных значений всех исследуемых характеристик, поэтому разработчикам приходится решать сложную многокритериальную оптимизационную задачу выбора булевой функции f. В связи тем, что актуальность данных задач высока, имеется множество работ, устанавливающих или уточняющих соотношения между разными криптографическими параметрами.

Дифференциальный криптоанализ алгоритма LBC

Для проведения дифференциального криптоанализа берутся две пары заранее подобранных шифртекстов A_1 и A_2, где атакующий вычисляет дифференциал: $\Delta A = A_1 \oplus A_2$, и с помощью вычисленного дифференциала пытается определить, каким должен быть дифференциал шифртекстов $\Delta B = B_1 \oplus B_2$.

В большинстве случаев вероятность подбора злоумышленником точного значения ΔB крайне низкая. Атакующий способен определить частоту возвращения ΔB для заданного ΔA, что в свою очередь дает ему возможность получить часть ключа или целый ключ.

В работе [15] представлены результаты проведенного анализа алгоритма шифрования на лавинный эффект и показано, что после седьмого раунда все биты выходного блока зависят от всех входных битов. Поэтому в первую очередь целесообразно провести исследование алгоритма шифрования на семи раундах. Перед проведением дифференциального криптоанализа алгоритма шифрования необходимо рассмотреть дифференциальные свойства используемых преобразований по отдельности.

Операция побитового циклического сдвига и xor не оказывают никакого влияния на разность, это легко проверит. Однако для построения многораундовых характеристик важно определить, как именно будут преобразованы значения нужных байтов, которые будут получены после других преобразований. При сложении раундового ключа по модулю 2 его биты будут взаимно уничтожены. Поэтому значение ключа также не влияет на дифференциалы. В связи с этим, изменение разности зависит в основном от дифференциальных свойств S-блоков.

В алгоритме LBC используется 4-х битный S-блок. Для S-блока необходимо построить таблицу распределения разности. Методология построения такой таблицы приведена в работах [16, 17]. При проведении дифференциального криптоанализа, необходимо отслеживать все комбинации двоичных векторов, складывать все возможные входные и выходные элементы S-блока, считать его нули и полученные результаты заносить в таблицу распределения разности (таблица 4).

Таблица 4. Таблица распределение разности для S- блока

ΔX/ΔY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	2	0	0	2	0	2	2	0	0	2	4	0	0	2	0
2	0	2	4	0	2	2	2	0	0	0	2	0	2	0	0
3	2	2	0	2	0	2	0	2	2	0	0	0	4	0	0
4	0	0	2	2	0	0	0	0	4	0	2	0	2	2	2
5	0	2	2	2	0	0	2	2	0	2	0	0	0	0	4
6	0	2	0	0	0	2	0	0	0	2	0	2	2	4	2
7	0	0	0	4	2	0	2	0	2	2	0	2	2	0	0
8	2	2	2	0	0	0	2	0	2	0	0	4	0	2	0
9	0	0	0	0	0	2	2	4	2	0	2	2	0	0	2
10	4	2	0	0	2	0	0	0	2	2	2	0	0	0	2
11	2	0	2	0	0	0	0	2	0	4	2	2	2	0	0
12	2	0	0	0	2	0	4	2	0	0	0	0	2	2	2
13	2	0	2	2	2	4	0	0	0	0	0	2	0	0	2
14	0	4	0	2	2	0	0	2	0	0	2	2	0	2	0
15	0	0	2	0	4	2	0	2	2	2	0	0	0	2	0

В построенной таблице распределения разностей, максимальным значением вероятности является значение $4/16 = 1/4$. В таблице 5 приведены входные и соответстующие им выходные разности с максимальными значениями.

Таблица 5. Результат преобразование разности для S-блока

ΔX	ΔY	ΔX	ΔY	ΔX	ΔY	ΔX	ΔY	ΔX	ΔY
0x01	0x0b	0x02	0x03	0x03	0x0d	0x04	0x09	0x05	0x0f
0x07	0x04	0x08	0x0c	0x09	0x08	0x0a	0x01	0x0b	0x0a
0x0d	0x06	0x0e	0x02	0x0f	0x05	0x06	0x0a	0x0c	0x07

В таблице 5 ΔX – является входом в блок подстановок, а ΔY – является значением, получаемым на выходе из блока подстановок, соответствующим ΔX.

Обсуждение разработанного легковесного алгоритма шифрования

При разработке и исследований криптографических свойств алгоритма LBC главным эталоном был взять легковесный алгоритм PRESENT. Алгоритм PRESENT в 2012 году был включен в международный стандарт облегченного шифрования ISO/IEC 29192-2:2012.

Разработанный легковесный алгоритм LBC структурно отличается от существующих на сегоднейшей день легковесных алгоритмов.

У алгоритма LBC-3 своя 4-х битная таблица замены (Таблица 1). Для линейного преобразования была использована функция RL. Эта функция добавлена для достижения хорошего качества «лавинного эффекта». А также циклический свдиг подблоков тоже обеспечивает хорошие перемещивания битов вместе с функцией RL.

Количество раундов алгоритма LBC-3 – 20. По результатам исследования хороший лавинный эффект достигается, начиная с 7-го раунда [15]. Поэтому 20-ти раундовые преобразования обеспечивает достаточную и хорошую криптостойкость.

Также самой отличительной особенности от Present у алгоритма LBC – это генерация раундовых ключей. Здест используются функций сдвига Ri похожая на RL, которая испольуется в преобразованиях основного алгоритма шифрования. Если в алгоритме Present только крайние левые 4 бита мастер ключа заменяются через S-блок, то в алгоритме LBC все биты ключа заменяются с помощью S-блока, используемого в основном процессе шифрования. Здесь функция Ri и циклический сдвиг подблоков мастер ключей обеспечивают хороший лавинный эффект, тем самым улучая свойства раундовых ключей.

Теоретические и экспериментальные испытания показали, что алгоритм полностью соответствует основным криптографическим требованиям.

Первым исследованием была проверка показателей лавинного эффекта. Для этого была создана специальная программа. Программа считала лавинный критерий и значения лавинного параметра, при изменении каждого бита входного блока данных. В работе [15] приведены результаты исследования лавинного эффекта на первом раунде при изменениях первого бита входного блока в каждом раунде. Среднее значения лавинного эффекта – 0.8766, максимальное значения – 0.969. После седьмого раунда среднее значения составляет – 0.0992. Здесь можно заметить, что, начиная с 7-го раунда среднее значения приближается к 0.09, свою очередь это свойство показывает хорошее качество лавинного эффекта. Чем меньше значение лавинного параметра, тем сильнее лавинный эффект в преобразовании. Из результата проведенного эксперимента следует, что алгоритм шифрования LBC удовлетворяет требованиям лавинного критерия после седьмого раунда. Поэтому мы считаем, что 20-и раундовые преобразования достаточным образом обеспечивает хороший криптостойкость.

Вторым этапом исследования свойства алгоритма LBC был дифференциальный криптоанализ. При проведении дифференциального криптоанализа установлено, что для входной разности $0x1111$ наиболее вероятной выходной разностью будет $0xA745$. Нетрудно подсчитать, что S-блок встречается в выходном выражении для второго подблока 13 раз, т.е. вероятность искомой разности равна $\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{16}=\left(\frac{1}{2}\right)^4$ и как следствие, достаточный уровень безопасности не обеспечивается. При проведении

анализа по перемещению каждого элемента и отслеживая на каждом шаге выполнение подстановки, было обнаружено, что после двенадцатого раунда на каждом подблоке S-блок будет задействован 127, 49, 67 и 92 раз. Повторно выбирая дифференциалы с максимальными значениями вероятностей получается, что вероятность нахождения ключа для двенадцати раундов шифра будет равна $\frac{1}{2^{98}}$.

Заключение

На основании проведенных тестов и исследований установлено, что нелинейное преобразование алгоритма шифрования LBC имеет надежную оценку стойкости. Характеристика используемого S-блока показывает наилучшее значения, не хуже других известных легковесных алгоритмов шифрования. Полученное значение вероятности по дифференциальному криптоанализу превосходит сложность полного перебора, т.е. после одиннадцатого раунда алгоритм шифрования является стойким к дифференциальному криптоанализу. В дальнейшем как продолжение данной работы, будут исследованы другие криптографические свойства линейных и нелинейных преобразовании используемых в алгоритме LBC.

Благодарность

Научно-исследовательская работа выполнена в рамках проекта AP09259570 «Разработка и исследование отечественного легковесного алгоритма шифрования при ограниченности ресурсов» в Институте информационных и вычислительных технологий КН МНВО РК.

Литература

1. Shaw D. The camellia cipher in OpenPGP, Tech. Rep. RFC5581 RFC Editor (Jun. 2009), https://doi.org/10.17487/rfc5581

2. Rescorla E. The transport layer security (TLS) protocol version 1.3: Tech. Rep. RFC8446 RFC Editor (2018), https://doi.org/10.17487/RFC8446

3. Daniel J. Bernstein, JohannesBuchmann, ErikDahmen Post-quantum cryptography, Springer, Berlin Heidelberg (2009)

4. Lang Li, Botao Liu, Hui Wang, QTL: A new ultra-lightweight block cipher, Microprocessors and Microsystems, Volume 45, Part A, 2016, pp. 45-55.

5. Bogdanov, A. *et al.* (2007). PRESENT: An Ultra-Lightweight Block Cipher. In: Paillier, P., Verbauwhede, I. (eds) Cryptographic Hardware and Embedded Systems - CHES 2007. CHES 2007. Lecture Notesin Computer Science, vol 4727, pp. 450-466. Springer, Berlin, Heidelber. https://doi.org/10.1007/978-3-540-74735-2_31

6. Shibutani, K., Isobe, T., Hiwatari, H., Mitsuda, A., Akishita, T., Shirai, T. (2011). Piccolo: An Ultra-Lightweight Block cipher. In: Preneel, B., Takagi, T. (eds) Cryptographic Hardware and Embedded Systems – CHES 2011. CHES 2011. Lecture Notesin Computer Science, vol 6917, pp 342-357. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23951-9_23

7. Guo, J., Peyrin, T., Poschmann, A., Robshaw, M. (2011). The LED Block Cipher. In: Preneel, B., Takagi, T. (eds) Cryptographic Hardware and Embedded Systems – CHES 2011. CHES 2011. Lecture Notesin Computer Science, vol 6917, p_p 326-341. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23951-9_22

8. K. Kazlauskas and J. Kazlauskas, “Key-dependent S-box generation in AES block cipher system,” Informatica, vol. 20, pp. 23-34, 2009.

9. Preneel B. Perspectives on Lightweight Cryptography," Inscript 2010, Shanghai, China, 20-24 October 2010.

10. Ivanov, G., Nikolov, N., Nikova, S. (2016). Cryptographically Strong S-Boxes Generated by Modified Immune Algorithm. In: Pasalic, E., Knudsen, L. (eds) Cryptography and Information Security in the Balkans. BalkanCryptSec 2015. Lecture Notes in Computer Science(), vol 9540. Springer, Cham. https://doi.org/10.1007/978-3-319-29172-7_3

11. Горбенко І.Д., Горбенко Ю.І. Прикладна криптологія. Теорія. Практика. Застосування: монографія.– Харків: Видавництво «Форт», 2012. – 870 с.

12. Sankhanil Dey, Ranjan Ghosh A Review of Existing 4-Bit Crypto S-Box Cryptanalysis Techniques and Two New Techniques with 4-Bit Boolean Functions for Cryptanalysis of 4-Bit Crypto S-Boxes*, Advances in Pure Mathematics, Vol.8 No.3, 2018, PP. 272-306. https://doi.org/10.4236/apm.2018.83015

13. Ardabek Khompysh, Nursulu Kapalova, Kunbolat Algazy, Dilmukhanbet Dyusenbayev & Kairat Sakan. Design of substitution nodes (S-Boxes) of a block cipher intended for preliminary encryption of confidential information, Cogent Engineering, No 9:1, (2022) pp. 1-14, https://doi.org/10.1080/23311916.2022.2080623

14. Kapalova, N., Algazy, K., & Haumen, A. Development of a new lightweight encryption algorithm. *Eastern-European Journal of Enterprise Technologies*, vol. *3, №*9(123), pp.6–19. https://doi.org/10.15587/1729-4061.2023.280055 /

15. Капалова Н.А., Хаумен А., Алгазы К.Т., Оценка алгоритма генерации раундовых ключей легковесного шифра LBC-3 // Вестник АУЭС №2, 2023г., - стр. 66-81. https://doi.org/10.51775/2790-0886_2023_61_2_20

16. Бабенко Л.К. Ищукова Е.А. Современные алгоритмы блочного шифрования и методы их анализа // Москва, «Гелиос АРВ», 2006 г. – 376 с.

17. K.T. Algazy, L.K. Babenko, R.G. Biyashev, E.A. Ishchukova, N.A. Kapalova, S.E. Nysynbaeva, Andrzej Smolarz Differential Cryptanalysis of New Qamal Encryption Algorithm // International journal of electronics and telecommunications, № 4, 2020, pp. 647-653.

СОДЕРЖАНИЕ

ПРОГРАММНЫЙ КОМИТЕТ 3
Приветственное слово участникам конференции 7

СЕКЦИЯ 3

Training a machine learning model for food product image recognition and model performance evaluation 10
Чат-боты, голосовые помощники и умные колонки: обзор и перспективы 19
Application of statistical models for forecasting assets of oil and gas companies.... 28
Automated hate speech classification through "Emotional" comments of network users 35
Разработка системы автоматического слежения за положением солнца 44
Разработка когнитивного человеко-машинного интерфейса для интеллектуальной системы управления сложными объектами и диагностики состояния оборудования 52
Пластмасса өнімдерін қайта өңдеуді басқарудың автоматтандырылған жүйесін жасау 59
Ғимараттың микроклиматын басқарудың автоматтандырылған жүйесін әзірлеу 67
Обзор архитектуры поисковых систем реального времени 74
Идентификация именованных сущностей на примере узбекского текста: методы и применение..... 80
Реляционные модели обработки тюркских языков..... 85
Automatic speech recognition for the kazakh language..... 90
Еда - подготовка данных для машинного обучения и прогнозирования динамики вредителей пшеницы..... 95
Развитие технологий искусственного интеллекта в современных информационно-телекоммуникационных системах 103
«Құқыққа қайшы интернет-контент» көптілді онтология қосымшасын құру .. 107
A new automatically controlled wind power station of the parallel manipulator 115
Morphological feature extraction for language identification: a comprehensive study 124
Оценка изменения пространственно-временного состояния окружающей среды региона 133
Методы машинного обучения для исследования зон пластового окисления на урановых месторождениях 137
Исследование использования нейронных сетей и приложений с ИИ для автоматизации процессов написания текстового описания продуктов 144
Онтологиялық инженеринг негізінде әскері-ғарыш қызметі саласындағы

стандарттарды әзірлеу және үйлестіру үшін зиялды басқару жүйесі 153

СЕКЦИЯ 4

Ақпараттық қауіпсіздік үшін жасанды интеллект әдісін қолдану 164
Управление рисками информационной безопасности мультисерверных web-систем 171
Безопасность и прозрачность данных в технологиях блокчейн 177
Protecting confidential data in the information business: strategies for the digital age 182
Қазақ тіліндегі мәтінге психо-эмоционалдық талдау жүргізуге арналған әдісті құру және зерттеу 188
Innovative approaches to protecting distributed computer systems and internet servers: using active methods 195
О стандартизации криптографических показателей безопасности СКЗИ 200
О подходах к установлению нормативных требований к процессам жизненного цикла СКЗИ 208
Current areas of security ensuring in wireless sensor networks.....217
Ашық кілт инфрақұрылымындағы биометрияға негізделген аутентификация әдісі 225
Структура алгоритма хеширования HAS03.....233
Интеграция кибербезопасности в BIM: современные тренды и перспективы ..240
Современные методы обеспечения информационной безопасности в BIM: анализ и сравнение 245
Исследование алгоритмов консенсуса для обеспечения стабильности и надежности блокчейн-системы.....251
Реализация квантово-устойчивого криптографического алгоритма HORS с хеш-функцией HAS01-256.....259
Әлеуметтік медиадағы экстремистік мазмұнды анықтау және зерттеу 265
LWC шифрлау алгоритміне дифференциалдық криптоталдау 273
Применение sat-решателей для оценки стойкости криптографических алгоритмов 282
Development of a user authentication user authentication method for web resource287
Жаңа хештеу алгоритмінің статитикалық қауіпсіздік қасиеттерін зерттеу 293
Исследование и реализация протокола квантового распределения ключей E91 299
Әлеуметтік желілердегі суреттер мен бейнелерді өңдеудің ерекшеліктері 305
Using multi-factor authentication in ensuring the security of medical data 311
Алгоритм электронного голосования с применением технологии блокчейн 319
AL04 шифрлау алгоритмінің статистикалық қауіпсіздігі бағалау 327
Теңгерімсіз деректермен желіге рұқсатсыз енуді анықтауға арналған терең оқыту модельдеріне талдау 334
Применение методов ролевой модели авторизации в здравоохранении через безопасные облачные вычисления 341

Позициялық емес полиномды санау жүйесіне негізделген аутентификация жүйесі үшін саусақ ізін тану алгоритмін таңдау 350
Зиянды бағдарламаларды анықтау үшін қолданылатын әдістерін талдау 355
Semantic analysis and classification of violent extremism texts: challenges and opportunities for society and security 360
Synthesis of a binary adder with a conditional sum 364
Development of a homomorphic encryption library for IoT protection based on the Atmelavr Kernel 369
Интернеттегі әлеуметтік желілердегі қылмыстық ақпараттың ықпалды таратушыларын анықтау..... 377
Теоретические основы понятия угроз информационной безопасности и классификация информационной безопасности 387
Применение нейронных сетей для идентификации человека по голосу 394
Application of Nir Soft Utilities in Computer Forensics for Analysis and Recovery of Digital Evidenc ……..……..398
User authentication and data management with using blockchain systems 405
Нелинейное преобразование легковесного алгоритма шифрования lbc..... 409

МАТЕРИАЛЫ

VIII Международной научно – практической конференции
«Информатика и прикладная математика»,
посвященной памяти 85 д.т.н, профессора Бияшева Р.Г.

26-27 октября 2023 года, Алматы, Казахстан

(ЧАСТЬ II)
(Сборник статей)

Под редакцией ***М.Н. Калимолдаева***

Компьютерная верстка ***Ж.А.Байсаринова***

Подписано в печать 18.10.2023 г. Формат А4
Печать цифровая. Бумага офсетная. Усл. печ. л. 48,94.
Тираж 300 экз. Заказ № 006604.
Отпечатано в ИИВТ МНВО РК.
Алматы, ул. Пушкина, 125