ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ

әл-Фараби атындағы Қазақ ұлттық университетінің

ХАБАРЛАРЫ

ИЗВЕСТИЯ

НАЦИОНАЛЬНОЙ АКАДЕМИИ
НАУК РЕСПУБЛИКИ КАЗАХСТАН
Казахский национальный
университет имени аль-Фараби

N E W S

OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF
KAZAKHSTAN
al-Farabi Kazakh National University

SERIES
PHYSICO-MATHEMATICAL

3 (343)

JULY – SEPTEMBER 2022

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

«ҚР ҰҒА Хабарлары. Физика-математикалық сериясы».
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **No 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.
Тақырыптық бағыты: *ақпараттық коммуникациялық технологиялар сериясы.*
Қазіргі уақытта: *«ақпараттық технологиялар»бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*
Мерзімділігі: *жылына 4 рет.*
Тиражы: *300 дана.*
Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

«Известия НАН РК. Серия физика-математическая».
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*
Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **No 16906-Ж** выданное 14.02.2018 г.
Тематическая направленность: *серия информационные коммуникационные технологии.*
В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*
Периодичность: *4 раз в год.*
Тираж: *300 экземпляров.*
Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

Адрес типографии: ИП «Аруна», г. Алматы, ул. Муратбаева, 75.

News of the National Academy of Sciences of the Republic of Kazakhstan.
Physical-mathematical series.
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan **No. 16906-Ж,** issued 14.02.2018
Thematic scope: *series information technology.*
Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*
Periodicity: *4 times a year.*
Circulation: *300 copies.*
Editorial address: *28, Shevchenko str., of. 219, Almaty, 050010, tel. 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

Address of printing house: ST «Aruna», 75, Muratbayev str, Almaty.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 5-18
https://doi.org/10.32014/2022.2518-1726.137
УДК 28.23.29
МРНТИ 28.17.27
А.С. Аканова[1*], А.А. Макашев[1], С.А. Наурызбаева[1], Н.Н.Оспанова[2]

[1]Казахский агротехнический университет им. С. Сейфуллина,
Казахстан, Астана;

[2]Павлодарский университет имени С. Торайгырова,
Казахстан, Павлодар.

E-mail: akerkegansaj@mail.ru

МОДЕЛИРОВАНИЕ ТЕМАТИЧЕСКОГО ИЗВЛЕЧЕНИЯ ДАННЫХ ИЗ ИНТЕРНЕТА

Аннотация. Процесс подготовки для результативного извлечения данных из Интернета по различным тематикам сталкивается с проб-лемой структурирования и организации процесса поиска данных и и их извлечения. Для решения данной проблемы можно успешно при-менить моделирование действий, производимых во время поиска и извлечения информации из Интертнета. Были исследованы веб-пар-синги с разных предметных областей, таких как финансовые данные, пси хологические исследования и другие. Описаны особенности работы веб-парсеров, способы хранения собранных данных. Иссле-до ваны понятия применяемые в области извлечения данных с Интер-нета. Также, в статье говорится о смежных темах, таких как NLP, глубокое и машинное обучение, и как они непосредственно связаны с процессом парсинга. В статье приведена модель поиска и извле че ния текста из Интернета, работа программы описывается в виде диаграм-мы прецедентов и диаграммы активности. Данные диаграммы исполь-зуются в первоначальном выполнении проекта и описании требований заказчика аналитиком. Для упрощения работы разработчика приме-няются различные виды диаграмм, но в большинстве случаев удобно использовать выше названные диаграммы для моделирования прог-рам много продукта. Также приведена схемаработы метода doc2bow,

который используется в машинном обучении при извлечения текста по темам. Также проведен обзор на современные инструменты парсинга, работающие с языком программирования Python. А именно библиотека BeautifulSoup, фреймворк Scrapy и набор инструментов для автоматизации тестирования Selenium. В конечном результате, были построены UML-диаграммы модели, которые подробно показывают процесс веб-парсинга. Представленная модель извлечения данных из Интернета является визуализацией действий производимых приложением. Предлагаемая диаграмма может использоваться при разработке приложений по извлечению данных из Интернет ресурса.

Ключевые слова: веб-парсинг, веб-парсер, моделирование, извлечение данных, диаграмм активности.

А.С. Ақанова[1*], А.А. Макашев[1], С.А. Наурызбаева[1], Н.Н. Оспанова[2]

[1]С. Сейфуллин атындағы Қазақстан агротехникалық университеті, Қазақстан, Астана;
[2]С. Торайғыров атындағы университеті, Қазақстан, Павлодар.
E-mail: akerkegansaj@mail.ru

ИНТЕРНЕТТЕН ТАҚЫРЫП БОЙЫНША ДЕРЕКТЕРДІ АЛУДЫ МОДЕЛДЕУ

Аннотация. Интернеттен әртүрлі тақырыптар бойынша деректерді тиімді алуға дайындық процесі деректерді іздеу және оларды алу процесін құрылымдау және ұйымдастыру проблемасына тап болады. Бұл мәселені шешу үшін Интертнеттен ақпаратты іздеу және алу кезінде жасалған әрекеттерді модельдеу арқылы әр-түрлі жолдарды тиімді қолдануға болады. Қаржылық деректер, психологиялық зерттеулер және басқалар сияқты әртүрлі тақырыптардағы веб-парсингтер зерттелді. Веб-парсерлердің ерекшеліктері ол жиналған деректерді сақтау әдістері арқылы сипатталды. Интернеттен деректерді алу саласында қолданылатын ұғымдар зерттелді. Сондай-ақ, мақалада NLP, терең және машиналық оқыту сияқты байланысты тақырыптар және олардың талдау процесіне тікелей қатысы туралы айтылады. Мақалада интернеттен мәтінді іздеу және шығару моделі берілген, бағдарлама прецеденттер диаграммасы және белсенділік диаграммасы түрінде сипатталған. Бұл диаграммалар жобаның бастапқы орындалуында және тапсырыс берушінің талаптарын сипаттап әзірлеушіге дайындау

үшін аналитик (талдаушы) жұмыс жасайды. Әзірлеушінің жұмысын жеңілдету үшін әртүрлі диаграммалар қолданылады, бірақ көп жағдайда бағдарламалық өнімді модельдеу үшін жоғарыда аталған диаграммаларды қолдану ыңғайлы. Сондай-ақ, тақырып бойынша мәтін шығару кезінде машиналық оқытуда қолданылатын doc2bow әдісінің сызбалары келтірілген. Python бағдарламалау тілімен жұмыс істейтін заманауи талдау құралдарына шолу жасалды. Атап айтқанда, BeautifulSoup кітапханасы, Scrapy шеңбері және Selenium тестін автоматтандыруға арналған құралдар жиынтығы туралы ақпарат берілген. Нәтижесінде веб-талдау процесін егжей-тегжейлі көрсететін UML диаграммалары арқылы модель жасалды. Интернеттен деректерді шығарудың ұсынылған моделі қолданба жасаған әрекеттерді визуализациялау болып табылады. Ұсынылып отырған диаграмма Интернет ресурстан деректерді шығару бойынша қосымшаларды әзірлеу кезінде пайдаланылуы мүмкін.

Түйін сөздер: веб-парсинг, веб-парсер, модельдеу.

A.S. Akanova[1*], A.A. Makashev[1] , C.A. Наурызбаева[1], N.N. Ospanova[2]

[1]Kazakh agrotechnical university, Kazakhstan, Astana;
[2]S.Toraigyrov University, Kazakhstan, Pavlodar.
E-mail: akerkegansaj@mail.ru

MODELING OF THEMATIC DATA EXTRACTION FROM THE INTERNET

Abstract. The process of preparing to extract data from the Internet on various topics faces the problem of structuring and organizing the data retrieval process. To solve the problem, modeling of actions performed when searching and extracting information from the Internet is used. Web parsings from such subject areas as financial data, psychological research and others were investigated. The features of the work of web parsers, methods of storing the collected data are described. The concepts used in the field of data extraction are investigated. The article describes NLP, deep and machine learning and presents a model for searching and extracting text from the Internet. The work of the program is described in the form of a precedent diagram and an action diagram. Diagrams are used by the analyst at the beginning of the project to outline the customer's requirements. The model of the doc2bow method, which is used in machine learning when

extracting text by topic, is shown. There is also an overview of parsing tools, namely, the BeautifulSoup library, the Scrapy framework and a set of tools for automating Selenium testing. As a result, UML diagrams of the model were built, which show in detail the process of web parsing. A model of data extraction actions from the Internet is presented. The proposed model can be used in the development of applications for data extraction.

Key words: web-parsing, web-parser, parser, parsing, modeling.

Введение. В эпоху накопления и обработки больших данных большое место в науке имеют исследования процесса извлечения данных из открытых Интернет источников. Такие технологии как NLP (Natural Language Processing) – обработка естественного языка требуют огромного количества данных, для того чтобы эффективно применить алгоритмы машинного обучения. В связи с требованием большого количества данных, не только для NLP, но и других сфер, появились инструменты для извлечения данных, которые называются «веб-парсеры», «парсеры», либо «веб-скрейперы». Инструменты извлечения данных, занимают особое место и в сфере психологических исследований (Speckmann, 2021). Каждое действие человека в Интернете оставляет «следы»: посты, комментарии, понравившиеся статьи и т.д. Speckman F. отметил, что такие «следы» непременно помогут в психологических исследованиях, поскольку каждый «след» описывает поведение человека. Инструментов извлечения текстов кроме применения в психологических исследованиях, использовались и в сфере финансов (Krotov et al., 2018). В статье показан сбор данных с помощью написания на языке программирования R веб-парсера. Термин «парсинг» обозначает синтаксический анализ. Веб-парсер можно запустить с помощью написания специального скрипта, например, на языке Python. При написании скрипта, задаются необходимые условия, для извлечения определенных блоков текста. Тем самым, посредством некоторых предустановленных правил производится анализ текста, в случае с веб-сайтами это весь HTML-документ. Далее, текст, который попал под условие скрипта необходимо сохранить. Для хранения можно использовать как обычные текстовые файлы, так и базы данных (Mahmood et al., 2018), о том, как делать сбор данных в какое-либо хранилище данных. Таким образом, с помощью веб-парсера можно автоматизировать сборку огромного количества данных, экономя на этом ресурсы и время. На сегодняшний день актуальность и достоверность информации являются приоритетом, что своей работой и

выполняет парсер. Путем запуска скрипта по расписанию, можно всегда получать свежую информацию.

Одной из особенностей при написании парсеров является то, что имеется возможность собирать данные не с одного веб-сайта, а с нескольких. Гибкость работы парсера напрямую зависит от разработчика, который пишет скрипт.

Извлечение данных, это систематизированный процесс извлечения и комбинирование необходимой информации из веб-ресурсов (Glez-Peña et al, 2014). Насколько нам известно веб-парсер имитирует действия человека при извлечении данных с веб-сайтов. А также информацию можно извлекать только с открытых веб-ресурсов.

В последнее время становится актуальным применение технологий глубокого обучения при анализе текста (Rekha et al., 2022). В статье сравниваются традиционные методы извлечения текста от сложных процессов. При извлечении текста из Интернета важно применение синтаксического анализа, в том числе универсальной зависимости, который рассматривается с помощью метода извлечения текста с межязыковыми отношениями (Taghizadeh et al., 2022). Синтез и извлечение структурированных данных с помощью неглубокого синтаксического анализа и сегментации предложений были предложены учеными и имел успех при работе с рускоязычным патентом (Korobkin et al., 2019). Веб-парсинг страниц можно отнести к более традиционным методам, поскольку процесс парсинга обходится без использования сложных технологий таких как нейронные сети и глубокое обучение. Но эффективное использование технологий веб-парсинга также не является простым. Для того, чтобы добиться нужного результата необходимо разработать корректную модель, отстроить архитектуру приложения, что в итоге будет вести себя согласно ожидаемым результатам.

Одним из примеров извлечения текста с источников является научная работа (Frisoni et al., 2021). В работе описывается то, как извлечение текста из публикаций, поможет справиться с их постоянным ростом. Необходимость извлечения полезной, структурированной информации является одной из основ веб-парсинга. Для этого необходимо определить четкие правила в алгоритме работы скрипта, применять регулярные выражения и корректные условия. Анализ HTML-разметки, а также использование XPath выражений являются актуальными и надежными при использования регулярных выражений (Antonov et al., 2020). Регулярные выражения являются неким шаблоном с определенными заданными внутри условиями. С помощью этих выражений можно найти нужные строки или подстроки. XPath запросы часто используются при

веб-парсинге. С помощью них можно находить нужные в DOM-структуре элементы, с которых далее будет происходить вычитка данных.

На данный момент не существует прямого законодательного органа, который бы занимался мониторингом веб-парсеров (Silva et al., 2018), но существуют нормативные документы в котором защищены авторские права. Существует еще одно правило «этики» при запуске веб-парсера на веб-сайт. Веб-парсер не должен нагружать сервер, на котором находится веб-сайт, то есть необходимо соблюдать интервал, при котором происходит парсинг, чтобы не навредить серверу веб-сайта.

Самой популярной библиотекой на языке Python является BeautifulSoup (Vargiu et al., 2013). С помощью данной библиотеки, можно написать парсер статичных страниц. Библиотека не имеет возможности парсить данные с динамически подгружающихся страниц, что является главным недостатком данной библиотеки.

В статье (Sirisuriya et al., 2015) говорится о том, что большинство веб-сайтов не дают доступа сохранить копию данных. Такая проблема действительно существует, поскольку многие веб-сайты используют для рендера страниц JavaScript или AJAX, который в свою очередь подгружает весь контент страницы динамический с помощью асинхронных запросов к серверу. Существуют библиотеки, которые решают данную проблему. Одна из них написана на языке Python: scrapy-splash. Это фреймворк, которые дает возможность создавать парсеры способные собирать данные, даже если веб-сайт используют загрузку контента через JavaScript или AJAX.

Многие крупные веб-сайты, такие как Twitter, iTunes, TikTok и др. предоставляют специальные API интерфейсы для того, чтобы получать определенные данные с их веб-сайта. В Twitter это посты, в iTunes информацию о треке, его продолжительность, наименование, альбом и т.д., в TikTok получение информации о профиле пользователя, искать пользователей и посты. Но у этих API интерфейсов существуют различные ограничения: требуется авторизация, ограничение на кол-во запросов в определенный интервал времени. О том как парсить данные веб-сайты в обход этим ограничениям описано в статье (Hernandez-Suarez., 2018). Он предложил использовать вышеописанную библиотеку написанную для Python – Scrapy. Движок scrapy отправляет запрос по ссылке поиска постов, также передаются необходимые параметры. Движок получает ответ от сервера и отправляет полученную страницу HTML элементами загрузчику. Далее, паук (в фреймворке Scrapy их принято так называть) начинает анализировать и парсить необходимые данные, которые в конце попадают в определенное хранилище данных.

Веб-парсинг используется для собра текстовой информации, которая в свою очередь требует обработки. Одной из проблем в компьютерной лингвистике является извлечение из текста информации по каким-либо темам.

Следовательно, исследование в области веб-парсинга требуют разработки модели извлечения информации, для добычи нужных данных, то есть избежать как можно меньше «мусора» при извлечении данных.

Для разработки модели извлечения информации необходимо обозначить следующие задачи:

- выполнить обзор и анализ научных статей
- разработать модель сбора целевых данных при веб-парсере на основе языка UML

Материалы и методы исследования. Моделирование извлечения данных по тематике. В данном исследование были применены такие методы исследования как анализ, синтез, сравнение.

Для того, чтобы понять, как работает парсер, и откуда собираются данные необходимо ознакомиться с понятием DOM-дерево (Document Object Model). DOM-дерево полностью показывает структуру веб-документа и иерархию всех HTML-элементов. Каждый элемент DOM-дерева – это то, что видит пользователь, находясь на веб-сайте: кнопка, меню, баннеры, картинки, видео (Uzun, 2020). Все эти элементы на странице считаются объектами. Эти элементы называют «тегами». По иерархии теги делятся на два вида – родительские и дочерние (Донова и др., 2019).

Основными действиями при извлечении данных по тематике является моделирование приложения. Для моделирования процесса веб-парсинга необходимо использовать UML (Unified Modeling Language) – унифицированный язык моделирования.

Чтобы определить какие роли непосредственно участвуют в процессе веб-парсинга и каким поведением они обладают, нужно смоделировать Use-case диаграмму (диаграмма прецедентов). Модель, которой показана на рисунке 1.

Рисунок 1. Use-case диаграмма (диаграмма прецедентов)

В диаграмме прицедентов (рисунок 1) показано распределение основных ролей объектов при извлечении данных. Основными исполнителями являются веб-сайт, сервер на котором работает веб-сайт, и, собственно, сам веб-парсер (инструмент, с помощью которого извлекают данные). Веб-парсер получает нужную страницу и начинает извлекать из нее данные которые требуются. Сервер отвечает за хранение БД. Веб-сайт отображает данные, приходящие с сервера. Между ними есть действие, которое связывает их – получение и отправка запросов. С помощью запросов веб-сайт и веб-парсер получают данные с сервера, в котором все данные и хранятся. После успешного выполнения запроса на получение содержимого HTML-страницы, происходит парсинг страницы нужных блоков данных.

Для того, чтобы подробно понять логику поведения системы необходимо построить диаграмму активностей (Activity diagram), которая показана на рисунке 2.

В диаграмме активностей можно подробно рассмотреть алгоритм действий веб-парсинга. Здесь можно наблюдать каждый шаг, перед тем как придет ожидаемый результат в виде нужного текста. Данная диаграмма показывает какая роль у объектов (веб-парсер, веб-сайт и сервер.) и за что они ответственны в системе.

Рассмотрев, вышеописанные UML-модели, разработчику не сложно определиться этапами разработки программного продукта предназначенный для тематического извлечения данных.

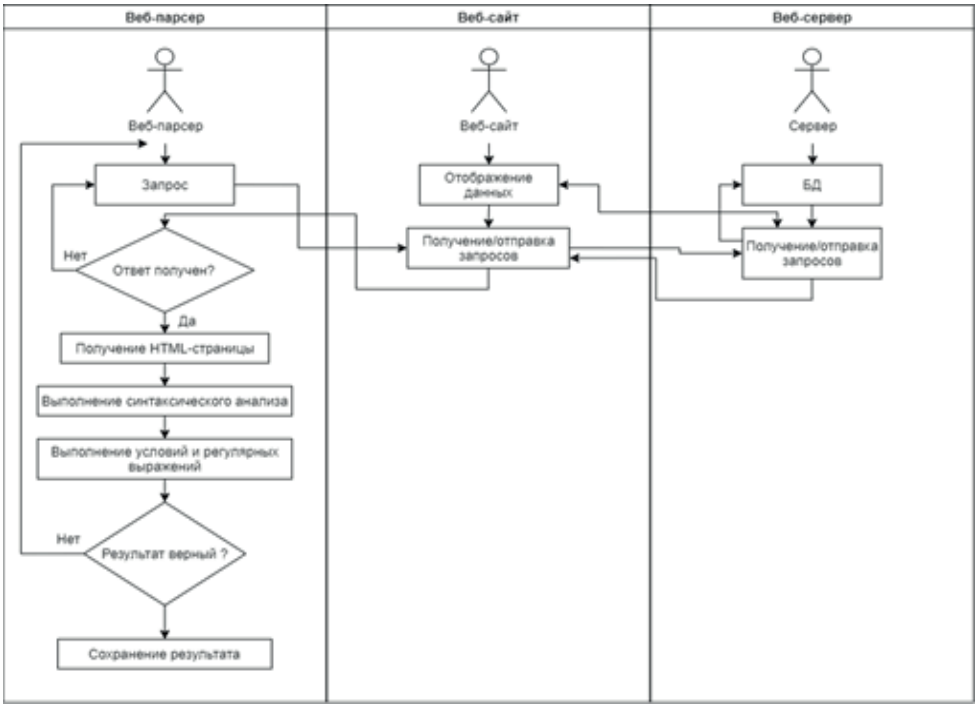


Рисунок 2. Activity диаграмма (диаграмма активностей)

Данная модель помогает решить некоторые проблемы разработчика:

1) Визуально увидеть полноценную картину программного продукта для аврсинга

2) Определиться с классами, которые ему потребуются в работе

3) Определить связи между объектами и их действия.

4) Легкая модернизация программного продукта

5) Выявление ошибко при тестирование

6) Выделение основных задач перед разработкой программного продукта

Архитектура программного продукта состит из веб-парсера (это может быть инструмент, либо приложение) и хранилища данных. Веб-парсер делает запрос к веб-сайту, получает ответ в виде HTML-страницы, далее производится синтаксический анализ по условиям, которые указал пользователь, в поиске нужных элементов помогают BeautifulSoup, Selenium или Scrapy, с помощью регулярных выражений можно повысить точность поиска определенных слов, фраз в тексте. Далее полученные результаты сохранять в БД, таблицы или текстовые файлы, для дальнейшей работы с ними.

При работе с веб-парсингом нужно помнить, что парсинг страниц на веб-сайтах, где имеется конфиденциальная информация, обход запретов на копирование контента с веб-сайтов будет являться нарушением закона. Необходимо соблюдать «этику» веб-парсинга, не нагружать сервера множеством параллельных запросов.

Элементы можно искать с помощью тега, класса, либо id элемента, также по CSS-селектору. Класс тега – это идентификатор, который используется при задании каких-либо стилей с помощью CSS. С помощью поиска через класс тега, можно сузить область поиска необходимого нам элемента, тем самым ускоряя веб-парсинг. Id – это уникальный идентификатор, который используется для наложения стилей на тег с помощью CSS, либо при наложении JavaScript скрипта на тег. CSS-селектор – специальный синтаксис, который используется в css файлах, при указывании тегов. Например: body p – означает, что внутри тега «body», необходимо найти тег «p»; p.paragraph-1 – означает, что необходимо найти тег p с классом «paragraph-1». С помощью определения данных условий парсинг находит нужные для сбора данных элементы.

Результаты. Из исследованных статьей следует отметить, что авторы использовали две библиотеки веб-парсинга – BeautifulSoup и Scrapy. Представляем описание некоторых характеристик библиотек веб-парсинга (таблица 1) и приемлемую библиотеку для применения его в программном продукте.

Таблица 1 - Характеристика BeautifulSoup, Scrapy и Selenium

Критерий\ Наименование парсера	BeautifulSoup	Scrapy	Selenium
Размер библиотеки	Легковесная библиотека, которая не требует много свободного места	Поскольку Scrapy является фреймворком для веб-парсинга, то занимает больше места	Selenium изначально был разработан для автоматизации тестирования веб-приложений, поэтому требует много места
Производительность	Данная библиотека работает медленно, но с применением технологии многопоточности производительность повышается	Scrapy является очень быстрым парсером, поскольку оптимизирован для извлечения больших объемов данных	Selenium является быстрым инструментом для парсинга, но уступает Scrapy

Парсинг JavaScript и AJAX страниц	BeautifulSoup не имеет возможности парсить страницы подгруженные динамически с помощью JavaScript или AJAX	У Scrapy есть возможность парсить динамически подгруженные страницы путем добавления библиотеки scrapy-splash	Selenium умеет парсить страницы подгружаемые JavaScript-ом, поскольку это приложение для автоматизации тестирования веб-приложений такая функция в нем имеется
Гибкость	BeautifulSoup имеет множество зависимостей, из-за чего его сложнее масштабировать	Scrapy легко масштабируется, имеет множество функционала для того чтобы увеличивать поток приходящих данных	Selenium может работать на больших проектах, но требует тщательной настройки ограничения лимита данных
Кроссплатформенность	Имеется кроссплатформенность поскольку это библиотека Python	Имеется кроссплатформенность поскольку это библиотека Python	Имеется кроссплатформенность
Документация	BeautifulSoup является популярным среди сообщества и у него хорошо расписанная, доступная для новичков документация	Scrapy имеет более сложную и менее детальную документацию, что является сложностью для новичков	Selenium имеет хорошо расписанную документацию, которая будет понятна как новичкам, так и опытным разработчикам
Доступность	Данная библиотека является рекомендацией для начинающих разработчиков веб-парсеров	Данная библиотека требует опыта от разработчика, используется в больших и сложных проектах	Данный инструмент требует опыта от разработчика

Из вышеописанной таблицы 1, можно сделать вывод, что BeautifulSoup и Scrapy написаны на языке Python, а значит он свободно распространяется. Selenium использует несколько инструментов. BeautifulSoup является библиотекой основанной на множестве зависимостей, что делает ее менее гибкой, не позволяет легко масштабироваться, но имеет упрощенный порог вхождения для разработчиков. Scrapy – фреймворк, разработанный для полноценного выполнения крупных процессов веб-парсинга и используется в больших проектах. Порог вхождения высокий. Selenium является «средним» между BeautifulSoup и Scrapy. С помощью него можно быстро парсить веб-страницы основанные на JavaScript.

Моделирование, применяя абстракцию при представлении данных показывает информацию в доступ визуальном формате. Тематическое извлечении текста на языке моделирования может выглядеть в виде сигналов, знаков, картинок и других объектов визуализации. Извлекая текст по ключевым словам из Интернет, можно определить их тематику и распределить их по «мешкам» или иначе применить метод Doc2bow из библиотеки Gensim. Doc2bow обычно используется в обучении данных. Слова объединяются в одну тематику (или как говорят оказываются в одном мешке) с помощью выбора слов с частотой образования друг с другом биграмм или триграмм.

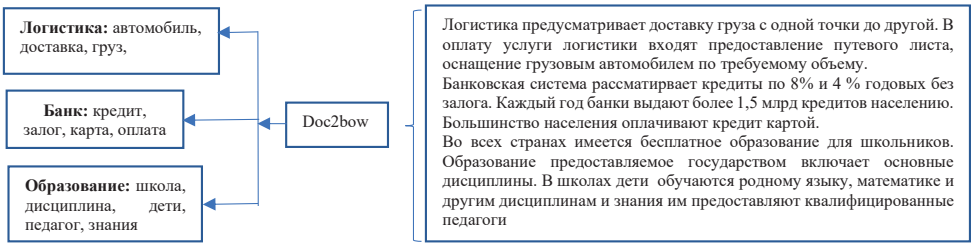


Рисунок 3. Модель схема работы Doc2bow

Отсюда, для тематического извлечения данных с Интернет ресурса предлагаем UML-диаграммы модели (рисунок 1, рисунок 2), которые подробно показывают процесс веб-парсинга.

Обсуждение. Представленные в статье диаграмма прецедентов и диаграмма активити являются одними из самых популярных видов диаграм, используемых в моделирование бизнес процессов программных продуктов. В данном случае данная гепотеза коррелирует с результатами исследований Донова М.М., Лозина Е.Н., Веретенникова Е.Г. и Барклаевской Н.В. (Донова и др., 2019, Барклаевская, 2015). Учитывая интенсивное развитие IT в Республике Казахстан была предложена одна из перспективных направлений проектной работы – моделирование, в котором решаются проблемы и совершенствования взаимодействия объектов программного продукта. Чем, обоснована решающая роль моделирования в разработке програмных продуктов. Перспективами дальнейших исследований являются применение концептуальной модели проекта. Предложен оптимальный инструмент для извлечения данных из Интернет ресурсов.

Итого, в результате были выполнены обзор и анализ научных статей, была разработана модель извлечения данных при веб-парсере на основе языка UML.

Заключение. Основными проблемами моделирования в разработке

программного продукта является отсутсвие аналитика либо аналитического мышления и не знание диаграмм языка UML. В результате данной работы была построена диаграмма прецедентов, в котором выделены актеры и их действие, что является ключевым моментом для разработчика. Разработчик в свою очередь может выделить себе без затруднений классы и методы в них. Диаграмме активности показывает последовательность работы выполнения действий актерами (классами).

Таки образом, моделирование извлечение текста из Интернета представлена в виде двух диаграмм на языке UML и схемы-модели работы doc2bow, что позволяет не только разработчику наглядно видеть процесс, но и заказчику.

Information about authors:

Akanova Akerke Saparovna – S. Seifullin Kazakh agrotechnical university, PhD, Senior lecturer, +77054480680, akerkegansaj@mail.ru; ORCID ID: http://orcid.org/0000-0003-2783-186X;

Makashev Ahmadi – S. Seifullin Kazakh agrotechnical university, Master student, +77086749153, *ahmadi98ahmadi@gmail.com*;

Nauryzbaeva Saya Amanzholovna – S. Seifullin Kazakh agrotechnical university, Master of Informatic, assistant lecturer, + 77016831139, ORCID ID: https://orcid.org/my-orcid?orcid=0000-0002-8544-0528;

Ospanova Nazira Nurgazyevna – Toraighyrov University, Associate Professor, +77011664573, nazira_n@mail.ru; ORCID ID: http://orcid.org/0000-0001-9416-0993.

REFERENCES:

Biegert J., Hofer H., Schultz A.H. (2015) A comparative study on web scraping. 86 DOI:10.1159/000433586 (in Eng).

Antonov E., Lopatina E., Ionkina K., Tretyakov E. Agent data merging Procedia Computer Science Том 169, Страницы 473 - 4782020 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 Seattle15 August 2019 DOI 10.1016/j. (procs.2020.02.).

Frisoni G., Moro G., Carbonaro A. A Survey on Event Extraction for Natural Language Understanding: Riding the Biomedical Literature Wave IEEE Access Открытый доступ Том 9, Страницы 160721 – 1607572021 DOI 10.1109/ACCESS.2021.3130956 (in Eng).

Glez-Peña D., Lourenço A., López-Fernández H., Reboiro-Jato M., Riverola F. F. (2014) Web scraping technologies in an API world (2014) // Briefings in bioinformatics, 15(5) 788-797. https://www.semanticscholar.org/paper/Web-scraping-technologies-in-an-API-world-Glez-Peña-Lourenço/b4951eb36bb0a408b02fadc12c0a1d8e680b589f (in Eng).

Hernandez-Suarez A. et al. Hernandez-Suarez Aldo, Sánchez-Pérez G., Toscano-

Medina K., Martínez-Hernández V., Sanchez Victor, Meana H. (2018) A web scraping methodology for bypassing twitter API restrictions. *https://arxiv.org/pdf/1803.09875.pd* (in Eng).

Korobkin D.M., Vasiliev S.S., Fomenkov S.A., Lobeyko V.I. (2019) Extraction of structural elements of inventions from Russian-language patents. 4th International Conference on Big Data Analytics, Data Mining and Computational Intelligence and the 8th International Conference on Theory and Practice in Modern Computing, Porto, DOI: 10.33965/tpmc2019_201907l020 (in Eng). Krotov V., Tennyson M. (2018) Research note: scraping financial data from the web using the R language. Journal of Emerging Technologies in Accounting, 15 (1), 169-181. DOI: https://doi.org/10.2308/1558-7940-15.1.i (Published: 01 July 2018) (in Eng).

Krotov V., Silva L. (2018) Legality and ethics of web scraping. Americas Conference on Information Systems 2018: Digital Disruption. https://www.researchgate.net/publication/324907302 Legality_and_Ethics_of_Web_Scraping (in Eng).

Mahmood A., Khan H.U., Alarfaj F.K., Ramzan M., Ilyas M. (2018) A multilingual datasets repository of the Hadith content. International Journal of Advanced Computer Science and Applications, 9 (2), 165 – 172. DOI 10.14569/IJACSA.2018.090224 (in Eng).

Rekha D., Sangeetha J., Ramaswamy V. (2022) Digital document analytics using logistic regressive and deep transition-based dependency parsing. Journal of Supercomputing, 78 (2), 2580 – 2596, DOI 10.1007/s11227-021-03973-4 (in Eng).

Speckmann F. (2021). Web scraping: A useful tool to broaden and extend psychological research // Zeitschrift für Psychologie. 229(4), 241-247, DOI: 10.1027/2151-2604/a000470 (in Eng).

Taghizadeh N., Faili H. (2022) Cross-lingual transfer learning for relation extraction using Universal Dependencies Computer Speech and Language, 71, 101265. DOI 10.1016/j.csl.2021.101265 (in Eng).

Uzun E. (2020) A novel web scraping approach using the additional information obtained from web pages //IEEE Access. 8. 61726-61740. (in Eng).

Vargiu E., Urru M. (2013) Exploiting web scraping in a collaborative filtering-based approach to web advertising. Artifical Intellegent Resours, 2 (1), 44-54. (in Eng).

Barclayevskaya N.V. The use of the unified modeling language UML in the project approach when teaching students in the specialty “business informatics. Materials of the scientific and methodological conference of the SZIU RANEPA, RANEPA under the President of the Russian Federation, Moscow. *https://elibrary.ru/download/elibrary_25659228_48815743.pdf.*

Dontsova M.M., Lozina E.N., Veretennikova E.G. Analysis and modeling of tutors’ work processes. Problems of design, application and security of information systems in the digital economy Materials of the XIX International Scientific and Practical Conference. Rostov-on-Don. https://elibrary.ru/download/elibrary_43785016_98237740.pdf.

DOM tree. [electronic resource]. URL: https://learn.javascript.ru/dom-nodes.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022),18-51
https://doi.org/10.32014/2022.2518-1726.138
UDC 621.39:004.05
IRSTI 81.93.29

Zh. Avkurova[1*], S. Gnatyuk[2], B. Abduraimova[3], L. Kydyralina[4]

[1]Karaganda Industrial University NJSC, Kazakhstan, Temirtau;
[2]National Aviation University, Ukraine, Kyiv;
[3]L.N.Gumilyov ENU, Kazakhstan, Astana;
[4]"ShakarimUniversity in Semey" NJSC, Kazakhstan, Semey.
E-mail: *zhadyra.avkurova.83@mail.ru*

MODELS OF STANDARDS AND GOVERNING RULES FOR THE SYSTEMS OF EARLY DETECTION OF APT-ATTACKS AND IDENTIFICATION OF VIOLATORS IN CYBERSPACE

Abstract. The violator of cybersecurity acting on the system changes some of its parameters, initializes or blocks the processes inherent in it. By evaluating these parameters, it is possible to detect the fact of intrusion into the system. On this principle the work of modern systems for the early detection of attacks and the identification of violators is based. In previous works, the authors described the parameters by which the intruder is identified - these are host and network parameters. Since the process of detecting and identifying an intruder takes place under conditions of uncertainty, and a number of certain parameters of systems for early detection of attacks are fuzzy, the functioning of such a system should be based on fuzzy logic.

Thus, in this work, on the basis of the proposed parameters, the linguistic variables were introduced and models of parameter standards were created. Membership functions were calculated for each variable and graphs of their terms were plotted. Also, standards have been formed that are necessary for the development of a system of logical rules to ensure the functioning of the system for early detection of attacks.

The results obtained will be further used to create an IDS / IPS system based on honeypot technology. In addition, examples of rules were developed

to identify the activities of different categories of cybersecurity intruders: disinformer, spammer, cracker, hacker, spam bot and hacker bot. These results can be used to improve existing IDS / IPS system or develop a new security system for early detection of APT-attacks directed on the critical information infrastructure of the state or other important objects.

Key words: intruder identification, cybersecurity, attack detection, linguistic variables, security policy.

Ж.С. Авкурова[1*], С.А. Гнатюк[2], Б.К. Абдураимова[3], Л.М. Кыдыралина[4]

[1]Қарағанды индустриялық университеті, КеАҚ, Қазақстан, Теміртау;
[2]Ұлттық авиациялық университеті, Украина, Киев;
[3]Л.Н.Гумилев атындағы Еуразия ұлттық университеті,
Қазақстан, Астана;
[4]Семей қаласының Шәкәрім атындағы университеті КеАҚ,
Қазақстан, Семей.
E-mail: *zhadyra.avkurova.83@mail.ru*

КИБЕРКЕҢІСТІКТЕГІ АРТ-ШАБУЫЛДАРДЫ ЕРТЕ АНЫҚТАУ ЖӘНЕ БҰЗУШЫЛАРДЫ СӘЙКЕСТЕНДІРУ ҮШІН ЭТАЛОН МОДЕЛЬДЕРІ АНЫҚТАУШЫ ЕРЕЖЕЛЕР

Аннотация. Жүйеде әрекет ететін киберқауіпсіздікті бұзушы оның кейбір параметрлерін өзгертеді, өзіне тән процестерді бастайды немесе блокқа қояды. Осы параметрлерді бағалау арқылы жүйеге ену фактісін анықтауға болады. Дәл осы принцип шабуылдарды ерте анықтаудың және құқық бұзушыларды сәйкестендірудің заманауи жүйелерінің жұмысына негізделген. Алдыңғы жұмыстарда авторлар бұзушыны анықтайтын параметрлерді сипаттады - бұл хост және желі параметрлері. Құқық бұзушыны анықтау және нақтылау процесі белгісіздік жағдайында жүретіндіктен және шабуылдарды ерте анықтау жүйелерінің бірқатар параметрлері анық емес болғандықтан, мұндай жүйенің жұмыс істеуі анық емес логикаға негізделуі керек. Осылайша, бұл жұмыста ұсынылған параметрлер негізінде лингвистикалық айнымалылар енгізіліп, параметрлер стандарттарының модельдері жасалды. Әрбір айнымалы үшін тиістілік функциялары есептелді және олардың графиктері салынды. Сондай-ақ, шабуылдарды ерте анықтау жүйесінің жұмыс істеуін қамтамасыз етуге мүмкіндік беретін логикалық

ережелер жүйесін әзірлеуге қажетті стандарттар қалыптастырылды. Алынған нәтижелер одан әрі honeypot технологиясы негізінде IDS / IPS жүйесін құру үшін пайдаланылатын болады. Сонымен қатар, киберқауіпсіздікті бұзушылардың әртүрлі санаттарының қызметін сәйкестендіру және анықтау үшін ережелер мысалдары әзірленді: теріс ақпарат беруші (дезинформатор), спаммер, крекер, хакер, спам-бот және крекер бот. Бұл нәтижелерді қолданыстағы IDS/IPS жүйелерін жақсарту немесе мемлекеттің және басқа маңызды объектілердің маңызды ақпараттық инфрақұрылымына жасалған APT шабуылдарын ерте анықтауға бағытталған жаңа қауіпсіздік жүйелерін әзірлеу үшін пайдаланылуы мүмкін.

Түйін сөздер: шабуылдаушыны сәйкестендіру, киберқауіпсіздік, шабуылды анықтау, лингвистикалық айнымалылар, қауіпсіздік саясаты.

Ж.С. Авкурова[1*], С.А. Гнатюк[2], Б.К. Абдураимова[3], Л.М. Кыдыралина[4]

[1]НАО Карагандинский индустриальный университет,
Казахстан, Темиртау;
[2]Национальный авиационный университет, Украина, Киев;
[3]ЕНУ им.Л.Н.Гумилева, Казахстан, Астана;
[4]НАО «Университет имени Шакарима города Семей»,
Казахстан, Семей.
E-mail:*zhadyra.avkurova.83@mail.ru*

МОДЕЛИ ЭТАЛОНОВ И ОПРЕДЕЛЯЮЩИЕ ПРАВИЛА ДЛЯ СИСТЕМ РАННЕГО ВЫЯВЛЕНИЯ APT-АТАК И ИДЕНТИФИКАЦИИ НАРУШИТЕЛЕЙ В КИБЕРПРОСТРАНСТВЕ

Аннотация. Нарушитель кибербезопасности, действуя на систему, меняет некоторые ее параметры, инициализирует или блокирует присущие ей процессы. Оценивая эти параметры, можно провести выявления факта вторжения нарушителя в систему. Именно на таком принципе основывается работа большинства существующих современных систем раннего выявления атак и идентификации нарушителей. В предыдущих работах авторами были описаны параметры, по которым осуществляется идентификация нарушителя – это хостовые и сетевые параметры. Поскольку процесс выявления и

идентификации нарушителя происходит в условиях неопределенности, а ряд определенных параметров систем раннего выявления атак носят нечеткий характер, то функционирование такой системы должно основываться на нечеткой логике.

Таким образом, в этой работе на основе предложенных параметров, были введены лингвистические переменные и построены модели эталонов параметров. Для каждой переменной были рассчитаны функции принадлежности и построены графики их термов. Также сформированы стандарты, необходимые для разработки системы логических правил, позволяющих обеспечить функционирование системы раннего выявления атак.

Полученные результаты в дальнейшем могут быть использованы для построения IDS / IPS систем на базе технологии honeypot. Кроме того, авторами были разработаны примеры правил для выявления и идентификации деятельности разных категорий нарушителей кибербезопасности: дезинформатора, спаммера, крекера, хакера, спам-бота и бота взломщика. Эти результаты могут использоваться для усовершенствования существующих IDS / IPS систем или для разработки новых систем безопасности, которые направлены на ранее выявление APT-атак на объекты критической информационной инфраструктуры государства и других важных объектов.

Ключевые слова: идентификации нарушителей, кибербезопасность, выявления атак, лингвистические переменные, политика безопасности.

Introduction. During the attack, the intruder, acting on the system, changes some of its parameters, creates or stops its inherent processes, and the like. All these actions affect the state of the system. By evaluating these parameters, it is possible to detect the fact of intrusion into the system. On this principle the work of modern systems for early detection of attacks (SEDA) and identification of violators is based (Avkurova, et all, 2020). This is how the NIDES system audits such processes as logging in, working with files and processes, administration and fixing errors and failures. The work (Iashvili, et all, 2021) describes the parameters by which the intruder is identified by the developed system. These include (host parameters):

- Username when logging in, UID;
- Login time, Tlog;
- Frequency of login requests, Nlog;
- Time taken to login, TSlog;
- Intensity of actions, I;
- CPU time/CPU load, CPU;

- The amount of loaded RAM, Muse;
- Number of executable files, NEF;
- The type of files used in the attack, AtEF;
- Number of failures and errors, NEr;
- Process/file execution time, RTPr/F;
- Unusual processes, UPr;
- File transfer to the system, TrFin;
- Modifying files, ModF,
- Copy/transfer files from the system, TrFout;
- Keyboard keystrokes, KS;
- Characteristics of ARP-, IP-, ICMP- and TCP- packets (network parameters).

Since the process of detecting and identifying an intruder takes place under conditions of uncertainty, and a number of the given parameters of the SEDA are fuzzy, the operation of such a system should be based on fuzzy logic.

Analysis of modern research and problem statement. In order to identify the intruder, we can use the logical-linguistic approach and the basic model of parameters, partially described in (Iashvili, et all, 2021), which will be the basis for creating of the developed SEDA. For example, to detect port scans, papers (Zhang, et all, 2020) use linguistic variables (LV) «Number of virtual channels» and «Age of virtual channels», and in paper (Zuzčák, et all, 2019) LV «Number of simultaneous connections», «Request processing speed», «Delay between requests» and «Number of packets with the same source and destination address» - to detect DDOS attacks and spoofing.

The process of detecting and identifying an intruder requires determining the necessary parameters and their properties. In this regard, the main goal of this work is to create models of parameter standards and a system of defining rules (DR) necessary for the effective functioning of SEDA and the identification of violators in a poorly formalized environment.

Materials and methods: Rationale for the approach. Let us consider the method of linguistic terms using statistical data (MLTS), where as a measure of membership of an element in a set, an estimate of the frequency of using a concept, which is given by a fuzzy set to characterize the element, is taken. To do this, the value of LV X={x1, x2, . . ., xn} is placed on the universal scale [0,1]. The method is based on the condition that the same number of experiments fall into each interval of the scale, but in practice this is usually not observed. In real conditions, there is an empirical table in which experiments can be unevenly distributed over intervals. Some of them

may not be involved at all, then the data is processed using a hint matrix. Let it be necessary to estimate in LV values the deviations of the parameter Δ BO [0, B] (B - the maximum possible deviation), which characterizes the current measurements. Next, for n = 5, we determine the value of the LV {x1, x2, x3, x4, x5}. The interval [0, B] and B/B (estimated ratio) are divided into k segments (for example, 5), according to which statistics are collected that characterize the frequency of using the LV value by an expert to display his conclusions. Further, the data is entered into the table and processed in such a way as to reduce the errors introduced during the experiment: individual elements are removed from the table, on the left side and on the right side of which there are zeros in the line. The hint matrix is a line whose elements are calculated by the formula:

$$k_j = \sum_{i=1}^{n} b_{ij} = \sum_{i=1}^{5} b_{ij},\ j = \overline{1,5}. \qquad (5)$$

Next, in the resulting row of the matrix, the maximum element $k_{\max} = \max k_j$ is selected, and then all elements of the table are converted by the expression

$$c_{ij} = b_{ij} k_{max} / k_j,\ i = \overline{1,5};\ j = \overline{1,5}, \qquad (6)$$

and for columns where $k_j = 0$ linear approximation is applied $c_{ij} = (c_{ij-1} + c_{ij+1})/2,\ i = \overline{1,5};\ j = \overline{1,5}$. Next, the value of the LV is calculated by the formula

$$\mu_{ij} = c_{ij} / c_{imax},$$

where $c_{imax} = \max\limits_{j} c_{ij},\ i = \overline{1,5};\ j = \overline{1,5}$. (7)

The described method uses data from statistical studies. Their processing is quite laborious, since to construct the FV of one term, it is necessary to carry out statistical studies of all terms of the LP (Gizun et all, 2013).

Let us create a model of standards of linguistic variables for fuzzy parameters of intruder identification from the set of parameters defined in (Iashvili, et all, 2021).

Login time, Tlog. This parameter is based on the fact that the activity of IS and users of these systems depends on the time of day. Usually, a lot of user activity when logging into the system is manifested in the daytime, less - at night, but other statistics are possible, which is determined by the mode of operation of the organization, which includes IS. The nature of this parameter is unclear, because it is impossible to unequivocally draw a conclusion about the illegal activity of the violator. So in organizations with

working hours from 08.00 to 16.00, the probability that the user who logged in is the intruder is the lowest at 08.00 and increases over time, reaching a maximum in the hours after 16.00. However, it should be noted that in the concept of honeypot technologies, this parameter loses its weight somewhat, since any activity on them is considered malicious.

Let's evaluate the LV «The level of legitimacy over time». Let's define the value of the linguistic variable $\{x_1, x_2, x_3\}$ corresponding to {legitimate, suspicious, illegitimate}. That is $T_{T\log} = \bigcup_{i=1} T^i_{T\log}$ = {legitimate, suspicious, illegitimate}

We use statistics on B = 24 hours. It is advisable to divide the total interval into 4 intervals [00:00;06:00], [06:00;12:00], [12:00;18:00], [18:00;24:00].

Data for LV Tlog Table 1

LV value	Interval			
	№1	№2	№3	№4
High	0	8	6	1
Medium	2	1	2	3
Low	6	1	1	4

Using expression (5), we determine $k_j = \|8\ 10\ 9\ 8\|$, where $k_{max} = 10$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 0 & 8 & 6{,}66 & 1{,}25 \\ 2{,}5 & 1 & 2{,}22 & 3{,}75 \\ 7{,}5 & 1 & 1{,}11 & 5 \end{Vmatrix}.$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 0 & 1 & 0{,}83 & 0{,}16 \\ 0{,}66 & 0{,}26 & 0{,}59 & 1 \\ 1 & 0{,}13 & 0{,}15 & 0{,}66 \end{Vmatrix}.$$

For $\bigcup_{i=1}^{3} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1} \Delta B_i / B =$ $= \{0{,}25;\ 0{,}5;\ 0{,}75;$ and get the following fuzzy numbers:

$$Л = \{0/0{,}25;\ 1/0{,}5;\ 0{,}83/0{,}75;\ 0{,}16/1\},$$
$$П = \{0{,}66/0{,}25;\ 0{,}26/0{,}5;\ 0{,}59/0{,}75;\ 1/1\},$$
$$Н = \{1/0{,}25;\ 0{,}13/0{,}5;\ 0{,}15/0{,}75;\ 0{,}66/1\}.$$

Graph of FP for the LP terms. The login time is shown on Fig. 1.

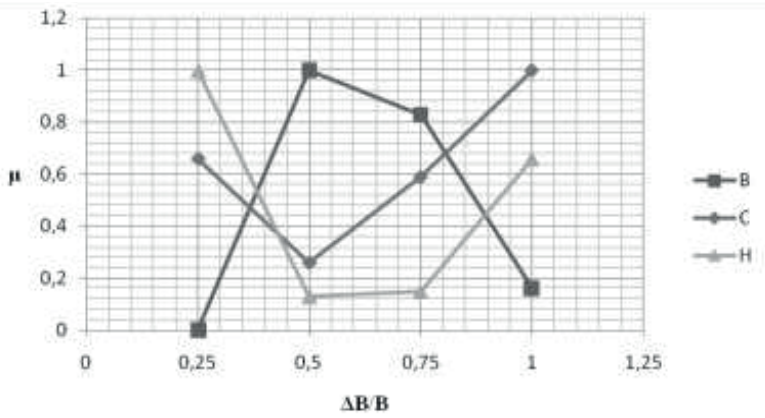


Fig. 1. Linguistic patterns of fuzzy numbers for Tlog

The frequency of login requests, Nlog. It is clear that a high frequency of login requests will be noted when the system is attacked by bots (in particular by hacking bots, since spammers do not require a login). A human attacker is also characterized by an increased frequency of requests as a result of an attempt to bypass the protection and the theoretical assumption that he does not have a legitimate login and password, so he will be forced to make at least several attempts. Moreover, the greater the number of attempts, the more likely it is that the intruder is really trying to enter the IS. It is clear that this parameter is also fuzzy.

Let's estimate the LV «Frequency of requests to enter the system.» Let's define the value of the linguistic variable {x1, x2, x3, x4, x5} corresponding to {low, below average, average, above average, high}. That is

$$T_{N\log} = \bigcup_{i=1}^{5} T_{N\log}^{i} = \{low, belowaverage, average, aboveaverage, high\}$$

The frequency of login requests for an ordinary user is usually minimal (more often, a legitimate user enters a login and password once), and modern password guessing programs are able to take 5310986 passwords /s (Golub et all, 2009). However, to determine the terms of this LV, it will be sufficient to limit ourselves to the value B = 100 requests/s, because a person is not able to go through the authentication procedure more than 10-15 times per minute. It is advisable to divide the general interval into 5 intervals [0, 1], [1, 10], [10, 40], [40, 80], [80, 100]

Data for LV Nlog Table 2

LV value	Interval				
	№1	№2	№3	№4	№5
Low	8	0	0	0	0
Below average	5	2	0	0	0
Average	1	6	4	0	0
Above average	0	2	8	1	0
High	0	0	1	6	6

Using expression (5), we determine $k_j = \|14 \quad 10 \quad 13 \quad 7 \quad 6\|$, where $k_{max} = 14$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 8 & 0 & 0 & 0 & 0 \\ 5 & 2{,}8 & 0 & 0 & 0 \\ 1 & 8{,}4 & 4{,}31 & 0 & 0 \\ 0 & 2{,}8 & 8{,}62 & 2 & 0 \\ 0 & 0 & 1{,}08 & 12 & 16 \end{Vmatrix}.$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0{,}56 & 0 & 0 & 0 \\ 0{,}12 & 1 & 0{,}51 & 0 & 0 \\ 0 & 0{,}32 & 1 & 0{,}23 & 0 \\ 0 & 0 & 0{,}07 & 0{,}75 & 1 \end{Vmatrix}.$$

For $\bigcup_{i=1}^{5} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{5} \Delta B_i / B = \{0{,}01;\ 0{,}1;\ 0{,}4;\ 0{,}8;\ 1\}$ and get the following fuzzy numbers:

$$H = \{1/0{,}01;\ 0/0{,}1;\ 0/0{,}4;\ 0/0{,}8;\ 0/1\},$$
$$HC = \{1/0{,}01;\ 0{,}56/0{,}1;\ 0/0{,}4;\ 0/0{,}8;\ 0/1\},$$
$$C = \{0{,}12/0{,}01;\ 1/0{,}1;\ 0{,}51/0{,}4;\ 0/0{,}8;\ 0/1\},$$
$$BC = \{0/0{,}01;\ 0{,}32/0{,}1;\ 1/0{,}4;\ 0{,}23/0{,}8;\ 0/1\},$$
$$B = \{0/0{,}01;\ 0/0{,}1;\ 0{,}07/0{,}4;\ 0{,}75/0{,}8;\ 1/1\}.$$

Graph of FP for the LP terms. The frequency of login requests is shown on Fig. 2

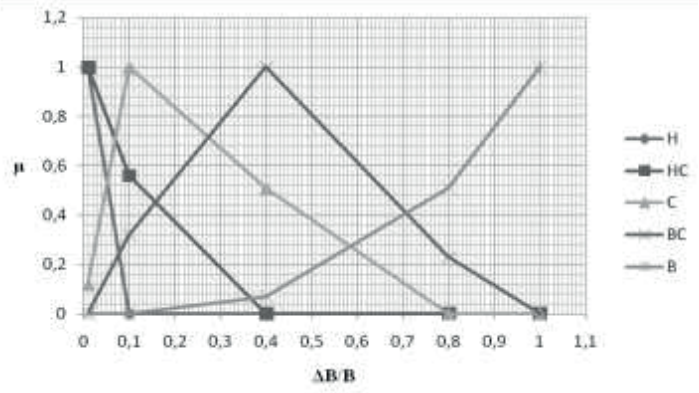


Fig. 2. Linguistic patterns of fuzzy numbers for Nlog

Time taken to login, TSlog. A parameter that is closely related to the previous one. The time spent by the intruder is in most cases greater than the time spent by the legitimate user. But it is fuzzy, because it does not allow an unambiguous identification.

Let's estimate the LV «Time spent on logging in». Let's define the value of the linguistic variable {x1, x2, x3, x4, x5} corresponding to {very small, small, medium, large, very large}. That is

$$T_{S\log} = \bigcup_{i=1}^{5} T_{S\log}^{i} = \{verysmall, small, medium, l\arg e, veryl\arg e\}$$

A legitimate user in a password-protected IS spends from several seconds to several minutes on identification. But the time spent by illegitimate users to break the password system is relatively large. So modern password guessing systems for cracking an 8-character password, consisting of a combination of letters, numbers and special characters, spend up to 61 days (Niu, et all, 2017). Therefore, we take the value of B = 60 days = 5184000 s. It is advisable to divide the total interval into 5 intervals [0 s; 30 s], [30 s; 5 min], [5 min; 1 h], [1 h; 1 day], [1 day; 60 days].

Data for LV TSlog Table 3

LV value	Interval				
	№1	№2	№3	№4	№5
Very small	9	3	0	0	0
Small	5	10	1	0	0
Medium	1	7	5	0	0
Large	0	1	2	9	2
Very large	0	0	1	6	9

Using expression (5), we determine $k_j = \|14 \quad 21 \quad 9 \quad 15 \quad 11\|$, where $k_{max} = 21$ and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 13,5 & 3 & 0 & 0 & 0 \\ 7,5 & 10 & 2,33 & 0 & 0 \\ 1,5 & 7 & 11,67 & 0 & 0 \\ 0 & 1 & 4,67 & 12,6 & 3,82 \\ 0 & 0 & 2,33 & 8,4 & 17,18 \end{Vmatrix}$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,22 & 0 & 0 & 0 \\ 0,75 & 1 & 0,23 & 0 & 0 \\ 0,13 & 0,6 & 1 & 0 & 0 \\ 0 & 0,08 & 0,37 & 1 & 0,3 \\ 0 & 0 & 0,14 & 0,49 & 1 \end{Vmatrix}$$

For $\bigcup_{i=1}^{5} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{5} \Delta B_i / B = \{5,79 \cdot 10^{-6};\ 5,79 \cdot 10^{-5};\ 6,94 \cdot 10^{-4};\ 0,02;\ 1\}$ and get the following fuzzy numbers:

$$ОМ = \{1/5,79 \cdot 10^{-6};\ 0,22/5,79 \cdot 10^{-5};\ 0/6,94 \cdot 10^{-4};\ 0/0,02;\ 0/1\},$$
$$М = \{0,75/5,79 \cdot 10^{-6};\ 1/5,79 \cdot 10^{-5};\ 0,23/6,94 \cdot 10^{-4};\ 0/0,02;\ 0/1\},$$
$$С = \{0,13/5,79 \cdot 10^{-6};\ 0,6/5,79 \cdot 10^{-5};\ 1/6,94 \cdot 10^{-4};\ 0/0,02;\ 0/1\},$$
$$Б = \{0/5,79 \cdot 10^{-6};\ 0,08/5,79 \cdot 10^{-5};\ 0,37/6,94 \cdot 10^{-4};\ 1/0,02;\ 0,3/1\},$$
$$ОБ = \{0/5,79 \cdot 10^{-6};\ 0/5,79 \cdot 10^{-5};\ 0,14/6,94 \cdot 10^{-4};\ 0,49/0,02;\ 1/1\}.$$

The graph of the FP for the LP terms is shown on Fig.3.

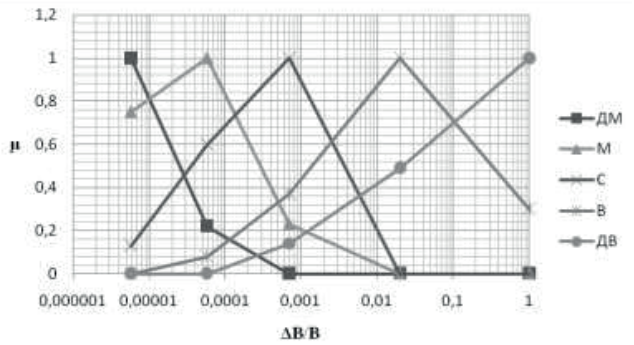


Fig. 3. Linguistic patterns of fuzzy numbers for TSlog

Intensity of actions, I. This refers to the number of any user actions, including logging in/out of the system, transferring, changing, copying files, starting/stopping processes, etc., per unit of time. The intensity may not differ for a human intruder and for a legitimate user, however, for bots it is much higher, therefore, it is the most significant for identifying and distinguishing between human-robot categories. Although a significant excess of the norm indicates the activity of unauthorized automatic intruder systems (bots), however, I is a fuzzy parameter, since it is very difficult to determine the normal value of the intensity indicator.

Let's evaluate the LV «Intensity of actions». Let's determine the value of the linguistic variable {x1, x2, x3} corresponding to {low, medium, high}.

That is $T_I = \bigcup_{i=1}^{i} T_I^i = \{low, medium, high\}$

The intensity of the actions of an ordinary person is very low, usually from 3 to 10 actions per minute. The intensity of the bots is ten times greater. For research, we will take the upper limit of 100 actions/min, although this figure may be higher for robots. It is advisable to divide the total interval into 4 intervals [0; 5], [5, 10], [10, 50], [50, 100].

Data for LV I Table 4

LV value	Interval			
	№1	№2	№3	№4
Low	7	5	1	0
Medium	0	7	4	0
High	0	1	5	7

Using expression (5), we determine $k_j = \|7\ 13\ 10\ 7\|$, where $k_{max} = 13$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 13 & 5 & 1,3 & 0 \\ 0 & 7 & 5,2 & 0 \\ 0 & 1 & 6,5 & 13 \end{Vmatrix}.$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,38 & 0,1 & 0 \\ 0 & 1 & 0,74 & 0 \\ 0 & 0,08 & 0,5 & 1 \end{Vmatrix}.$$

For $\bigcup_{i=1}^{3} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{3} \Delta B_i / B = \{0,05;\ 0,1;\ 0,5;\ 1\}$ and get the following fuzzy numbers:

$$H = \{1/0,05;\ 0,38/0,1;\ 0,1/0,5;\ 0/1\},$$
$$C = \{0/0,05;\ 1/0,1;\ 0,74/0,5;\ 0/1\},$$
$$B = \{0/0,05;\ 0,08/0,1;\ 0,5/0,5;\ 1/1\}.$$

The graph of the FP for the LP terms. The intensity of actions is shown on Fig.4.

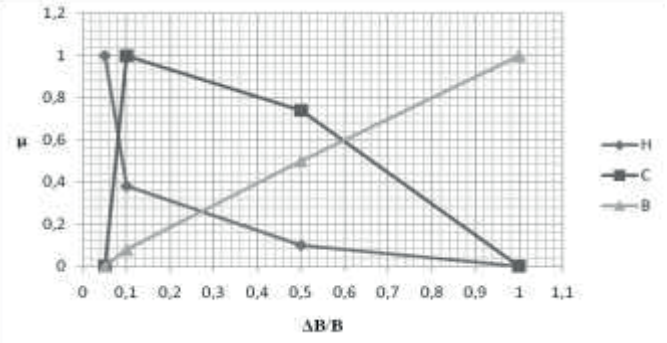


Fig. 4. Linguistic patterns of fuzzy numbers for I

Processor time / processor load, CPU. Since the number of active processes on honeypot systems should be kept to a minimum, any increase in load is a sign of intruder activity on the system. In real ISs, the probability

that the activity is caused by the intruder is somewhat lower, and, of course, the normal amount of processor time is higher. However, this parameter can still be effectively used to identify the fact of a violation in intrusion detection systems and SEDA. Since it is impossible to give an unambiguous answer about the intruder for this parameter, primarily due to the possible activity of viruses, the processor CPU is a fuzzy parameter.

Let's estimate the LV «Processor time/processor load». Let's determine the value of the linguistic variable {x1, x2, x3} corresponding to {low, medium, high}.

That is $T_{CPU} = \bigcup_{i=1}^{3} T^{i}_{CPU} = \{low, medium, high\}$

Under normal conditions of average use of PC capacities and in the absence of intruders or malware, the average processor load is 20-30%. Of course, the norm may vary somewhat depending on the OS, installed software and the production tasks of the organization. The maximum possible percentage of CPU load is B = 100%. It is advisable to divide the total interval into 4 intervals [0; 20], [20, 50], [50; 75], [50, 100].

Data for LV CPU Table 5

LV value	Interval			
	№1	№2	№3	№4
Low	9	6	0	0
Medium	3	8	1	0
High	0	1	5	8

Using expression (5), we determine $k_j = \|12\ 15\ 6\ 8\|$, where $k_{max} = 15$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 11,25 & 6 & 0 & 0 \\ 3,75 & 8 & 2,5 & 0 \\ 0 & 1 & 12,5 & 15 \end{Vmatrix}$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,53 & 0 & 0 \\ 0,47 & 1 & 0,31 & 0 \\ 0 & 0,07 & 0,83 & 1 \end{Vmatrix}$$

For $\bigcup_{i=1}^{3} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{3} \Delta B_i / B = \{0,2;\ 0,5;\ 0,75;\ 1\}$, and get the following fuzzy numbers:

$$H = \{1/0,2;\ 0,53/0,5;\ 0/0,75;\ 0/1\},$$
$$C = \{0,47/0,2;\ 1/0,5;\ 0,31/0,75;\ 0/1\},$$
$$B = \{0/0,2;\ 0,07/0,5;\ 0,83/0,75;\ 1/1\}.$$

The graph of the FP for the LP terms. Processor time / processor load is shown on Fig.5.

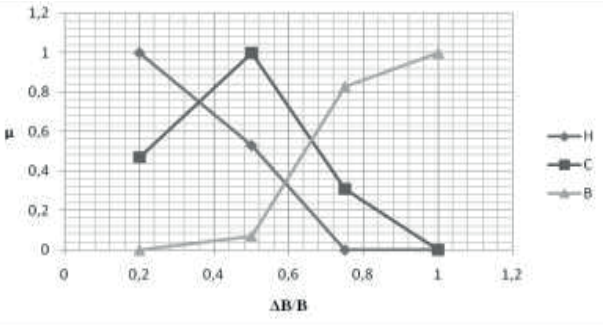


Fig. 5. Linguistic patterns of fuzzy numbers for CPU

The amount of loaded RAM, Muse. Similar in essence to the previous one and is also fuzzy. The FV for this parameter is almost identical to the LV «Processor time/processor load».

Number of executable files/processes, NEF. Also included in the group of fuzzy parameters. The fact of an attacker's actions on this parameter is determined by a deviation from the norm. This parameter includes only user processes and files, and system files are not taken into account. So in each organization, in accordance with the security policy and job responsibilities, each legitimate user can use certain files at a certain moment, and the simultaneous use of several files or processes at once is practically excluded. This allows to identify both external and internal Intruders, but with a certain probability.

Let's estimate the LV «Number of executable files». Let's define the value of the linguistic variable {x1, x2, x3} corresponding to {very small, small, normal, large, very large}.

That is $T_{NEF} = \bigcup_{i=1}^{5} T^{i}_{NEF} = \{\text{very small, small, normal, large, very large}\}$

The normal number of executable files, as already noted, is very dependent on the industry in which a particular IS operates, and on the set of security policy rules and job descriptions of the organization. Basically, this value ranges from 7 to 10 processes. Considering also the technological limitations of IS and typical security policies, we will take the maximum value of 20 user processes for research. It is advisable to divide the total interval into 4 intervals [0; 4], [4, 8], [8, 12], [12, 16], [16; 20].

Data for LV NEF Table 6

LV value	Interval				
	№1	№2	№3	№4	№5
Very small	8	5	1	0	0
Small	4	7	2	0	0
Normal	1	4	9	3	1
Large	0	0	4	8	2
Very large	0	0	1	5	8

Using expression (5), we determine $k_j = \|13 \quad 16 \quad 17 \quad 16 \quad 11\|$, where $k_{max} = 17$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 10,46 & 5,31 & 1 & 0 & 0 \\ 5,23 & 7,44 & 2 & 0 & 0 \\ 1,31 & 4,25 & 9 & 3,19 & 1,55 \\ 0 & 0 & 4 & 8,5 & 3,09 \\ 0 & 0 & 1 & 5,31 & 12,36 \end{Vmatrix}$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,51 & 0,1 & 0 & 0 \\ 0,7 & 1 & 0,27 & 0 & 0 \\ 0,15 & 0,47 & 1 & 0,35 & 0,17 \\ 0 & 0 & 0,47 & 1 & 0,36 \\ 0 & 0 & 0,08 & 0,43 & 1 \end{Vmatrix}.$$

For $\bigcup_{i=1}^{5} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{5} \Delta B_i / B = \{0,2;\ 0,4;\ 0,6;\ 0,8;\ 1\}$ and get the following fuzzy numbers:

$$ОМ = \{1/0,2;\ 0,51/0,4;\ 0,1/0,6;\ 0/0,8;\ 0/1\},$$
$$М = \{0,7/0,2;\ 1/0,4;\ 0,27/0,6;\ 0/0,8;\ 0/1\},$$
$$Н = \{0,15/0,2;\ 0,47/0,4;\ 1/0,6;\ 0,35/0,8;\ 0,17/1\},$$
$$Б = \{0/0,2;\ 0/0,4;\ 0,47/0,6;\ 1/0,8;\ 0,36/1\},$$
$$ОБ = \{0/0,2;\ 0,0/0,4;\ 0,08/0,6;\ 0,43/0,8;\ 1/1\}.$$

The graph of the FP for the LP terms. The number of executable files/ processes is shown on Fig.6.

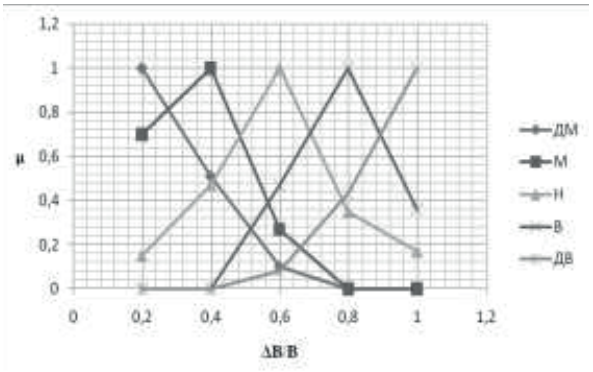


Fig. 6. Linguistic patterns of fuzzy numbers for NEF

Number of failures and errors, NEr. This parameter is fuzzy, since failures and errors can occur when working in the authorized user mode and intruder. However, with frequent repeated failures or errors, it can be concluded with a certain degree of probability that the system has been attacked. This group includes a wide range of events from authorization errors to failures when executing certain processes or files. With the active work of the intruder, regardless of his class and category, the frequency of occurrence of malfunctions will be somewhat higher. It should also be noted that it is quite possible when identifying an intruder-work, this frequency will be even higher.

Let's estimate the LV «Number of failures and errors». Let's determine the value of the linguistic variable {x1, x2, x3} corresponding to {low, medium, high}.

That is $T_{NEr} = \bigcup_{i=1}^{3} T^{i}_{NEr} = \{\text{low, medium, high}\}$

The functionality of an IS is considered normal if there are no errors or failures in its operation at all. However, the occurrence of a small number

of failures is still possible, mainly due to insufficient user qualifications, his negligence or the use of low-quality hardware / software. There are no official statistics on this issue, so it is difficult to determine the normal value of this parameter. As part of the study, we set the maximum number of errors and failures per day B = 10. It is advisable to divide the total interval into 4 intervals [0, 1], [1, 4], [4, 8], [8, 10].

Data for LV NEr Table 7

LV value	Интервал			
	№1	№2	№3	№4
Low	5	1	0	0
Medium	0	4	3	0
High	0	0	1	6

Using expression (5), we determine $k_j = \|5\ 5\ 4\ 6\|$, where $k_{max} = 6$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 6 & 1,2 & 0 & 0 \\ 0 & 4,8 & 4,5 & 0 \\ 0 & 0 & 1,5 & 6 \end{Vmatrix}$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,2 & 0 & 0 \\ 0 & 1 & 0,94 & 0 \\ 0 & 0 & 0,25 & 1 \end{Vmatrix}$$

For $\bigcup_{i=1}^{3} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{3} \Delta B_i / B = \{0,1;\ 0,4;\ 0,8;\ 1\}$ and get the following fuzzy numbers:

$$H = \{1/0,1;\ 0,2/0,4;\ 0/0,8;\ 0/1\},$$
$$C = \{0/0,1;\ 1/0,4;\ 0,94/0,8;\ 0/1\},$$
$$B = \{0/0,1;\ 0/0,4;\ 0,25/0,8;\ 1/1\}.$$

The graph of the FP for the LP terms. The number of failures and errors is shown on Fig.7.

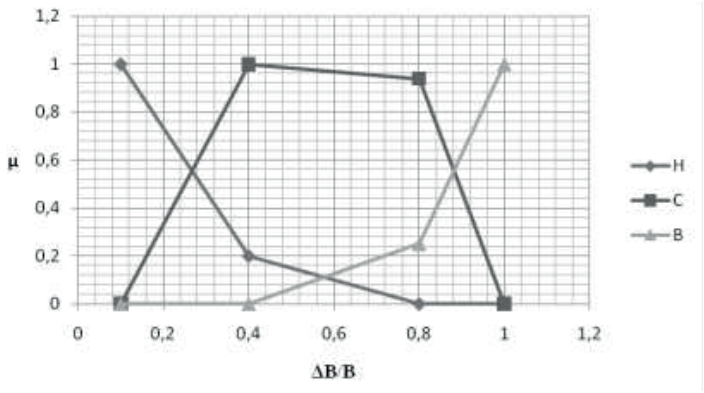


Fig. 7. Linguistic patterns of fuzzy numbers for NEr

Process/file execution time, RTPr/F. Examining the statistics of the work of IS of various enterprises and organizations, it is easy to notice that, depending on the specifics of the work, the time spent on performing a certain operation is approximately the same for the same type of IS and their tasks. Honeypot systems mainly run system processes, that support the operation of the honeypot itself, or administrator processes that run at a certain time for a certain period. Thus, when identifying such processes, it can be concluded that the system is attacked by an intruder. Since this can be caused by the negligence of the employee, the conclusion is ambiguous and, accordingly, the parameter is fuzzy.

Let's estimate the LV «Process/file execution time». Let's define the value of the linguistic variable {x1, x2, x3, x4, x5} corresponding to {very small, small, medium, large, very large}.

That $T_{RT\,\mathrm{Pr}/F} = \bigcup_{i=1}^{5} T^{i}_{RT\,\mathrm{Pr}/F} == \{\text{very small, small, medium, large, very large}\}$

A legitimate user in IS, in the course of performing his job duties, works with a certain file or process in the course of a certain time. So the average worker works with one file or process for a period of time from 30 minutes to 3 hours. If this indicator is significantly less or more, then this may indicate suspicious activity. It is advisable to set the maximum value of this variable B = 24 hours, the total interval is divided into 5 intervals [0 s; 1 min], [1 min; 30 min], [30 min 3 h], [3 h; 6 h], [6 h; 24h].

Data for LV RTPr/F Table 8

LV value	Interval				
	№1	№2	№3	№4	№5
Very small	9	4	1	0	0

Small	5	7	2	0	0
Medium	0	3	8	3	0
Large	0	0	3	9	6
Very large	0	0	1	4	9

Using expression (5), we determine $k_j = \|14 \quad 14 \quad 15 \quad 16 \quad 15\|$, where $k_{max} = 16$, and, in accordance with (6), calculate:

$$\|c_{ij}\| = \begin{Vmatrix} 10,29 & 4,57 & 1,07 & 0 & 0 \\ 5,71 & 8 & 2,13 & 0 & 0 \\ 0 & 3,43 & 8,53 & 3 & 0 \\ 0 & 0 & 3,2 & 9 & 6,4 \\ 0 & 0 & 1,07 & 4 & 9,6 \end{Vmatrix}$$

Let us calculate the FV by the formula (7):

$$\|\mu_{ij}\| = \begin{Vmatrix} 1 & 0,44 & 0,1 & 0 & 0 \\ 0,71 & 1 & 0,27 & 0 & 0 \\ 0 & 0,4 & 1 & 0,35 & 0 \\ 0 & 0 & 0,36 & 1 & 0,71 \\ 0 & 0 & 0,11 & 0,42 & 1 \end{Vmatrix}$$

For $\bigcup_{i=1}^{5} \mu_{ij}$ respectively, we find the estimated relations $\bigcup_{i=1}^{5} \Delta B_i / B =$ $= \{6,94 \cdot 10^{-4};\ 0,02;\ 0,125;\ 0,25;\ 1\}$ and get the following fuzzy numbers:

$$ОМ = \{\ 1/6,94 \cdot 10^{-4};\ 0,44/0,02;\ 0,1/0,125;\ 0/0,25;\ 0/1\ \},$$
$$М = \{\ 0,71/6,94 \cdot 10^{-4};\ 1/0,02;\ 0,27/0,125;\ 0/0,25;\ 0/1\ \},$$
$$С = \{\ 0/6,94 \cdot 10^{-4};\ 0,4/0,02;\ 1/0,125;\ 0,35/0,25;\ 0/1\ \},$$
$$Б = \{\ 0/6,94 \cdot 10^{-4};\ 0/0,02;\ 0,36/0,125;\ 1/0,25;\ 0,71/1\ \},$$
$$ОБ = \{\ 0/6,94 \cdot 10^{-4};\ 0/0,02;\ 0,11/0,125;\ 0,42/0,25;\ 1/1\ \}.$$

The graph of the FP for the LP terms. The execution time of the process/ file is shown on Fig.8.

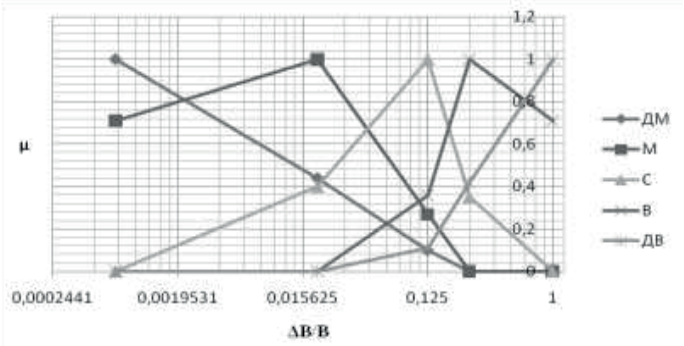


Fig. 8. Linguistic patterns of fuzzy numbers for RTPr/F

Thus, in the work with the use of SEDA, LV was introduced and models of the standards of parameters Tlog, Nlog, TSlog, I, CPU, Muse, NEF, NEr, RTPr/F were created. Also, for each LV, the FV were calculated and the graphs of their terms were plotted. Formed standards are necessary for the formation of logical rules to ensure the functioning of the SEDA. The results obtained will be further used to create an IDS system based on honeypot technology. Firstly, we will form in the next subsection of the DR to identify the fact of an IS violation and identify the identity of the attacker

DR system for efficient implementation of virtual honeypots.

To date, the following classes of SEDA are being developed:

 systems that detect the fact of an intrusion based on a comparison of the functional state of the IS with a set of specific signatures (templates) and

 systems based on used so-called DR (Gizun et all, 2013- Siddiqui, et all, 2016).

The main disadvantage of the first class of SEDA, which DR-based systems are deprived of, is the impossibility of their use in unknown attacks and, as a result, the impossibility of practical use in conditions of uncertainty and a fuzzy formalized environment. Therefore, despite the extremely low percentage of false positives of signature systems, the further development of second-class SEDA, in our opinion, is much more promising. The existing decision-principle SEDA are mainly focused on the use of complex mathematical models and require a lot of time for the formation of statistical data. However, expert approaches do not have such a requirement, which greatly simplifies the use of this method in the field of constructing SEDA. In this regard, an urgent task in the development of SOPs is to create models for detecting an anomalous state of IS caused by the activity of an intruder, based on the use of fuzzy logic methods, expert assessments and models of parameter standards necessary to identify an intruder. The use of these

models in the construction of the decisive type of SEDA is associated with the need to form rules aimed at identifying the intruder and his identification. That is why the purpose of this work is to develop a mathematical model that is used in the formation of the corresponding DR to identify the intruder.

In previous works (Avkurova, et all, 2020- Iashvili, et all, 2021), two groups of parameters for detecting and identifying an intruder are defined: with fuzzy and clear nature, respectively. So fuzzy parameters (Login time, Tlog; Frequency of login requests, Nlog; Time spent on login, TSlog; Intensity of action, I; Processor time / processor load, CPU; Amount of loaded RAM, Muse; Number of executable files, NEF; Number of failures and errors, NEr; Process/file execution time, RTPr/F) at the first stage of the SEDA operation make it possible to identify the presence of an intruder in the IS and to carry out their preliminary identification in a poorly formalized environment. Clear parameters should be used to confirm the fact of activity of the violation and the final assignment of the intruder to a certain category (Login user name, UID; Type of files used in the attack, AtEF; Unusual processes, UPr; Transfer file to the system, TrFin; modifying files, ModF; copy / transfer files from the system, TrFout; Keyboard keystrokes, KS) at the second stage of the system operation.

To solve the problem, it is necessary to create sets of DR, which are some statements that are based on the result of generalization of certain theoretical and experimental knowledge (data) and reflect the intuitive judgments of experts to ensure the search for a rational semantic solution to weakly formalized problems.

The construction of the DR can be carried out using the corresponding model, for the creation of which we introduce a set of linguistic identifiers:

$$LI = \bigcup_{i=1}^{d} LI = \{LI_1, LI_2, ..., LI_d\}, \qquad (8)$$

where d - the number of elements of the set required to detect the anomalous state, a LI_i (i = 1,d) – elements LI, each of which takes one of the text values that characterize in linguistic form the level of the anomalous state of the system, which can be generated by attacking actions. For example, for d = 5, expression (1) can be defined as follows:

$$LI = \bigcup_{i=1}^{5} LI = \{LI_1, LI_2, LI_3, LI_4, LI_5\} = \{Н, БНВ, БВН, В, К\}, \qquad (9)$$

where LI_1=Н, LI_2=БНВ, LI_3=ББН, LI_4=В and LI_5=К respectively reflected by the text values «Low», «More low than high», «More high than low», «High» and «Critical».

Further, based on the sets of identifiers LI and the set of linguistic or extended by the name - logical-linguistic connection LC, we will construct a set of PV:

$$ER = \left\{ \bigcup_{i=1}^{n} ER_i \right\} = \{ER_1, ER_2, ..., ER_n\}, \tag{10}$$

where ER_i (i=1,n) - a subset of possible rules for detecting the n-th anomalous state generated by the n-th attack, while

$$\begin{aligned} &\bigcup_{i=1}^{n} ER_i = \bigcup_{i=1}^{n} \left\{ \bigcup_{j=1}^{r_n} ER_{ij} \right\} = \{ER_{11}, ER_{12}, ..., ER_{1r_1}\}, \\ &\{ER_{21}, ER_{22}, ..., ER_{2r_2}\}, ..., \{ER_{n1}, ER_{n2}, ..., ER_{nr_n}\}, \end{aligned} \tag{11}$$

where ER_{ij} (i=1,n, j=1,r_n) - j-th rule of the n-th subset of possible rules, r_i (i=1,n) - the total number of possible rules aimed at detecting the n-th anomaly.

Note that for each ER_{ij} there is a corresponding DR:

$$\begin{aligned} &\{ER_{11} = (LC_{11} \to LI_{11}), ER_{12} = (LC_{12} \to LI_{12}), ..., \\ &ER_{1r_1} = (LC_{1r_1} \to LI_{1r_1})\}, \{ER_{21} = (LC_{21} \to LI_{21}), \\ &ER_{22} = (LC_{22} \to LI_{22}), ..., ER_{2r_2} = (LC_{2r_2} \to LI_{2r_2})\}, \\ &... \\ &\{ER_{n1} = (LC_{n1} \to LI_{n1}), ER_{n2} = (LC_{n2} \to LI_{n2}), ..., \\ &ER_{nr_n} = (LC_{nr_n} \to LI_{nr_n})\}. \end{aligned} \tag{12}$$

Generalizing expression (5), taking into account (10) and (11), we obtain

$$\begin{aligned} &ER = \bigcup_{i=1}^{n} \left\{ \bigcup_{j=1}^{r_i} ER_{ir_j} \right\} = \bigcup_{i=1}^{n} \left\{ \bigcup_{j=1}^{r_i} (LC_{ir_j} \to LI_{ir_j}) \right\} = \\ &\left\{ \bigcup_{i=1}^{n} \left\{ \bigcup_{j=1}^{r_i} ER_{ir_j} = (LC_{ir_j} \to LI_{ir_j}) \right\} \right\}, \end{aligned} \tag{13}$$

where ER_{ir_j} is r_j-th anomaly detection rule generated by the n-th attack, which is literally interpreted as: «If LC_{ir_j} is true, then the level of the abnormal state that can be generated by the n-th attack will be LI_{ir_j}».

The creation of rules is usually carried out on the basis of an expert approach, this is especially important in cases where it is necessary to give preference to one of the alternatives, for example, in which LC_{ir_j} (13) is the result associated with LI_{ir_j} and most objectively reflect the state of the system. Let consider the process of forming a choice for a set of alternatives using a specific example.

Let r_i logical-linguistic links and d (1) linguistic identifiers be used to create a set of rules, one of which can most objectively reflect the state of the environment about the presence of an anomaly. Therefore, the total number of possible alternative solutions is dxr_1, that is, for the assembly of each rule ER_{1j} ($j = 1,r_1$), it is necessary to consider d alternative variants of the rules, to select one of which we will use the methods of determining the coefficients of importance (CI). We will use the rank transformation method (RT), since it allows using the services of several experts, tabular forms are used as input, the initial function is linear, and the complexity is low.

Next, as an example, we define d=5, r_1=3, then

$$LC_I = \left\{ \bigcup_{j=1}^{r_1} LC_{1j} \right\} = \{LC_1, LC_2, LC_3\} = \{(t_{Tlog} \cong П,$$
$$t_{Nlog} \cong BC, t_{TSlog} \cong B, t_{CPU} \cong B, t_{NEF} \cong B, t_{NEr} \cong B,$$
$$t_{RTPr/F} \cong ДМ), (t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong B,$$
$$t_{CPU} \cong B, t_{NEF} \cong ДВ, t_{NEr} \cong B, t_{RTPr/F} \cong ДВ),$$
$$(t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong ДВ, t_{CPU} \cong B, t_{NEF} \cong ДВ,$$
$$t_{NEr} \cong B, t_{RTPr/F} \cong ДВ)\},$$

and as values LI_{1k}(k = 1,..,5) use the data from the formula (9). Thus, for each LC_{1j}(j = 1,..,3) there are possible d=5 finals of detection of anomalies associated with specific values of linguistic identifiers in (9). The most objective result will be determined using the method of average ranks (AR) (Gornitska et all, 2012).

According to this method, as an example, we will use the judgments of 4 experts on d=5 of possible results ER^k_{1j}(k = 1,d, j = 1,r_1) for each j-th rule.

For example, for the first rule, the set of alternative solutions will be

$$\bigcup_{k=1}^{d} ER_{11}^{k} = \left\{ ER_{11}^{1}, ER_{11}^{2}, ER_{11}^{3}, ER_{11}^{4}, ER_{11}^{5} \right\} =$$

$$\{(t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong B, t_{CPU} \cong B, t_{NEF} \cong B,$$

$$t_{NEr} \cong B, t_{RTPr/F} \cong ДМ) \rightarrow H, (t_{Tlog} \cong П, t_{Nlog} \cong BC,$$

$$t_{TSlog} \cong B, t_{CPU} \cong B, t_{NEF} \cong B, t_{NEr} \cong B, t_{RTPr/F} \cong ДМ)$$

$$\rightarrow БНВ, (t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong B, t_{CPU} \cong B,$$

$$t_{NEF} \cong B, t_{NEr} \cong B, t_{RTPr/F} \cong ДМ) \rightarrow БВН, (t_{Tlog} \cong П,$$

$$t_{Nlog} \cong BC, t_{TSlog} \cong B, t_{CPU} \cong B, t_{NEF} \cong B, t_{NEr} \cong B,$$

$$t_{RTPr/F} \cong ДМ) \rightarrow B, (t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong B,$$

$$t_{CPU} \cong B, t_{NEF} \cong B, t_{NEr} \cong B, t_{RTPr/F} \cong ДМ) \rightarrow K\}.$$

Further, on the basis of RT, we determine the CI, which are reflected by the parameter λ. Its minimum value indicates a greater advantage of the alternative, that is, its CI is higher. Fo the rule ER_{11} let's calculate the values x_{1j}^{k} i λ_{1j}^{k} for each of the possible results $ER^{k}{}_{11}$ (k = 1,5): x_{11}^{1}=(1+3+1+2)/4=1,75; x_{11}^{2}=(2+1+3+2)/4=2; x_{11}^{3}=(3+2+2+2)/4=2,25; x_{11}^{4}=(2+4+3+3)/4=3; x_{11}^{5} =(4+4+ 3+4)/4=3,75. The CI value is defined as $\lambda_{1j}^{k} = x_{1j}^{k} / N$, where N - sum of all ranks (N=10). According to the results listed in Table 1, it can be seen that the best result has ER_{11}^{1}, since $\bigwedge_{k=1}^{5} \lambda_{11}^{k} = \lambda_{11}^{k} = 0{,}18$.

Similarly, we will make calculations for ER_{1j}^{k}(j=2,3): for ER_{12}^{k} - x_{12}^{1} =(2+3+1+2)/4=2, x_{12}^{2}=(1+2+1+2)/4=1,5; x_{12}^{3}=(3+1+2+3)/4=2,25; x_{12}^{4} =(3+4+2+2)/4=2,75; x_{12}^{5}=(3+2+3+4)/4=3; для ER_{13}^{k} - x_{13}^{1}=(2+3+2+4)/4=2,75; x_{13}^{2}=(3+2+2+1)/4 =2; x_{13}^{3}=(2+3+1+1)/4=1,75; x_{13}^{4}=(3+4+3+4)/4= 3,5; x_{13}^{5} =(4+3+2 +4)/4=3,25.

Results and discussion. The calculation results (see Table 9) show that the best result for the rules ER_{12}, ER_{13}, have corresponding alternatives ER_{12}^{2}, ER_{13}^{3}.

The obtained data can be used as specific values in the construction of real rules in practical SEDA (Niu, et all, 2017- Ma, et all, 2019).

Ranks ER^k_{1j} and CI Table 9

ER^k_{1j}	j	k	Experts				λ^k_{1j}	λ^k_{1j}
			1	2	3	4		
ER^1_{11}	1	1	1	3	1	2	1,75	0,18
ER^2_{11}		2	2	1	3	2	2	0,2
ER^3_{11}		3	3	2	2	2	2,25	0,23
ER^4_{11}		4	2	4	3	3	3	0,3
ER^5_{11}		5	4	4	3	4	3,75	0,38
ER^1_{12}	2	1	2	3	1	2	2	0,2
ER^2_{12}		2	1	2	1	2	1,5	0,15
ER^3_{12}		3	3	1	2	3	2,25	0,23
ER^4_{12}		4	3	4	2	2	2,75	0,28
ER^5_{12}		5	3	2	3	4	3	0,3
ER^1_{13}	3	1	2	3	2	4	2,75	0,28
ER^2_{13}		2	3	2	2	1	2	0,2
ER^3_{13}		3	2	3	1	1	1,75	0,18
ER^4_{13}		4	3	4	3	4	3,5	0,35
ER^5_{13}		5	4	3	2	4	3,25	0,33

Thus the rule ER_{11} will take the form:

$ER_{11} = (t_{Tlog} \cong П, t_{Nlog} \cong BC, t_{TSlog} \cong B, t_{CPU} \cong B, t_{NEF} \cong B, t_{NEr} \cong$

$\cong ДМ) \rightarrow$ which can be verbally interpreted as follows: «If the current values of $t_{Tlog}, t_{Nlog}, t_{TSlog}, t_{CPU}, t_{NEF}, t_{NEr}, t_{RTPr/F}$ are closest to the values $П, BC, B, B, B, B, ДМ$ respectively, included in $T_{Tlog}, T_{Nlog}, T_{TSlog}, T_{CPU}, T_{NEF}, T_{NEr}, T_{RTPr/F}$, then the level of anomalous state generated by the attack of the intruder (in this case, a disinformer, cracker or hacker) will be LOW «.

Based on fuzzy parameters, we will form a set of DR ER_1 to identify such categories of the intruder as a disinformer, cracker, hacker and present it in the form of table 10, denoting the values of the parameters П - suspicious, Н - illegitimate, ВС - above average, В - high (large) , ОБ - very large, ОМ - very small, М – small:

Set of rules ER_i Table 10

Rule	t_{Tlog}	t_{Nlog}	t_{TSlog}	t_{CPU}	t_{NEF}	t_{Ner}	$t_{RTPr/F}$	Result
ER_{11}	П	ВС	В	В	В	В	ОМ	Н
ER_{12}	П	ВС	В	В	В	В	М	Н
ER_{13}	П	ВС	В	В	В	В	В	Н
ER_{14}	П	ВС	В	В	В	В	ОБ	Н
ER_{15}	П	ВС	В	В	ОБ	В	ОМ	БНВ
ER_{16}	П	ВС	В	В	ОБ	В	М	Н
ER_{17}	П	ВС	В	В	ОБ	В	В	Н
ER_{18}	П	ВС	В	В	ОБ	В	ОБ	БНВ
ER_{19}	П	ВС	ОБ	В	В	В	ОМ	БНВ
ER_{110}	П	ВС	ОБ	В	В	В	М	БНВ
ER_{111}	П	ВС	ОБ	В	В	В	В	БНВ
ER_{112}	П	ВС	ОБ	В	В	В	ОБ	БНВ
ER_{113}	П	ВС	ОБ	В	ОБ	В	ОМ	БВН
ER_{114}	П	ВС	ОБ	В	ОБ	В	М	БНВ
ER_{115}	П	ВС	ОБ	В	ОБ	В	В	БНВ
ER_{116}	П	ВС	ОБ	В	ОБ	В	ОБ	БВН
ER_{117}	П	В	В	В	В	В	ОМ	БНВ
ER_{118}	П	В	В	В	В	В	М	БНВ
ER_{119}	П	В	В	В	В	В	В	БНВ
ER_{120}	П	В	В	В	В	В	ОБ	БНВ
ER_{121}	П	В	В	В	ОБ	В	ОМ	БВН
ER_{122}	П	В	В	В	ОБ	В	М	БНВ
ER_{123}	П	В	В	В	ОБ	В	В	БНВ
ER_{124}	П	В	В	В	ОБ	В	ОБ	БВН
ER_{125}	П	В	ОБ	В	В	В	ОМ	БВН
ER_{126}	П	В	ОБ	В	В	В	М	БНВ
ER_{127}	П	В	ОБ	В	В	В	В	БНВ
ER_{128}	П	В	ОБ	В	В	В	ОБ	БВН
ER_{129}	П	В	ОБ	В	ОБ	В	ОМ	В
ER_{130}	П	В	ОБ	В	ОБ	В	М	БВН
ER_{131}	П	В	ОБ	В	ОБ	В	В	БВН
ER_{132}	П	В	ОБ	В	ОБ	В	ОБ	В
ER_{133}	Н	ВС	В	В	В	В	ОМ	БНВ
ER_{134}	Н	ВС	В	В	В	В	М	Н

ER_{135}	Н	ВС	В	В	В	В	В	Н
ER_{136}	Н	ВС	В	В	В	В	ОБ	БНВ
ER_{137}	Н	ВС	В	В	ОБ	В	ОМ	БВН
ER_{138}	Н	ВС	В	В	ОБ	В	М	БНВ
ER_{139}	Н	ВС	В	В	ОБ	В	В	БНВ
ER_{140}	Н	ВС	В	В	ОБ	В	ОБ	БВН
ER_{141}	Н	ВС	ОБ	В	В	В	ОМ	БВН
ER_{142}	Н	ВС	ОБ	В	В	В	М	БВН
ER_{143}	Н	ВС	ОБ	В	В	В	В	БВН
ER_{144}	Н	ВС	ОБ	В	В	В	ОБ	БВН
ER_{145}	Н	ВС	ОБ	В	ОБ	В	ОМ	В
ER_{146}	Н	ВС	ОБ	В	ОБ	В	М	БВН
ER_{147}	Н	ВС	ОБ	В	ОБ	В	В	БВН
ER_{148}	Н	ВС	ОБ	В	ОБ	В	ОБ	В
ER_{149}	Н	В	В	В	В	В	ОМ	БВН
ER_{150}	Н	В	В	В	В	В	М	БВН
ER_{151}	Н	В	В	В	В	В	В	БВН
ER_{152}	Н	В	В	В	В	В	ОБ	БВН
ER_{153}	Н	В	В	В	ОБ	В	ОМ	В
ER_{154}	Н	В	В	В	ОБ	В	М	БВН
ER_{155}	Н	В	В	В	ОБ	В	В	БВН
ER_{156}	Н	В	В	В	ОБ	В	ОБ	В
ER_{157}	Н	В	ОБ	В	В	В	ОМ	В
ER_{158}	Н	В	ОБ	В	В	В	М	БВН
ER_{159}	Н	В	ОБ	В	В	В	В	БВН
ER_{160}	Н	В	ОБ	В	В	В	ОБ	В
ER_{161}	Н	В	ОБ	В	ОБ	В	ОМ	К
ER_{162}	Н	В	ОБ	В	ОБ	В	М	В
ER_{163}	Н	В	ОБ	В	ОБ	В	В	В
ER_{164}	Н	В	ОБ	В	ОБ	В	ОБ	К

Based on fuzzy parameters, we will form sets of DRs to identify a spammer:

$$ER_2 = \{ER_{21} = (t_1 \cong C, t_{CPU} \cong B, t_{NEr} \cong B,$$
$$t_{RTPr/F} \cong ДМ) \rightarrow БНВ, ER_{22} = (t_1 \cong C,$$
$$t_{CPU} \cong B, t_{NEr} \cong B, t_{RTPr/F} \cong M) \rightarrow H,$$

$ER_{23} = (t_1 \cong B, t_{CPU} \cong B, t_{NEr} \cong B, t_{RTPr/F} \cong$
$ДМ) \rightarrow B, ER_{24} = (t_1 \cong B, t_{CPU} \cong B,$
$t_{NEr} \cong B, t_{RTPr/F} \cong M) \rightarrow БВН\}$

and ER_3 to detect and identify spam bots:

$ER_3 = \{ER_{31} = (t_1 \cong B, t_{CPU} \cong B, t_{NEr} \cong B,$
$t_{RTPr/F} \cong ДМ) \rightarrow B, ER_{32} = (t_1 \cong B, t_{CPU} \cong B,$
$t_{NEr} \cong B, t_{RTPr/F} \cong M) \rightarrow БВН\}.$

Based on fuzzy parameters, we will form a set of DRs for detecting and identifying a cracker bot and present it in the form of a table 11:

Set of rules ER_4 Table 11

Rule	t_{Tlog}	t_{Nlog}	t_{TSlog}	t_I	t_{CPU}	t_{NEF}	t_{Ner}	$t_{RTPr/F}$	Result
ER_{11}	П	ВС	В	В	В	В	В	ОМ	Н
ER_{12}	П	ВС	В	В	В	В	В	М	Н
ER_{13}	П	ВС	В	В	В	В	В	В	Н
ER_{14}	П	ВС	В	В	В	В	В	ОБ	Н
ER_{15}	П	ВС	В	В	В	ОБ	В	ОМ	БНВ
ER_{16}	П	ВС	В	В	В	ОБ	В	М	Н
ER_{17}	П	ВС	В	В	В	ОБ	В	В	Н
ER_{18}	П	ВС	В	В	В	ОБ	В	ОБ	БНВ
ER_{19}	П	ВС	ОБ	В	В	В	В	ОМ	БНВ
ER_{110}	П	ВС	ОБ	В	В	В	В	М	БНВ
ER_{111}	П	ВС	ОБ	В	В	В	В	В	БНВ
ER_{112}	П	ВС	ОБ	В	В	В	В	ОБ	БНВ
ER_{113}	П	ВС	ОБ	В	В	ОБ	В	ОМ	БВН
ER_{114}	П	ВС	ОБ	В	В	ОБ	В	М	БНВ
ER_{115}	П	ВС	ОБ	В	В	ОБ	В	В	БНВ
ER_{116}	П	ВС	ОБ	В	В	ОБ	В	ОБ	БВН
ER_{117}	П	В	В	В	В	В	В	ОМ	БНВ
ER_{118}	П	В	В	В	В	В	В	М	БНВ
ER_{119}	П	В	В	В	В	В	В	В	БНВ
ER_{120}	П	В	В	В	В	В	В	ОБ	БНВ
ER_{121}	П	В	В	В	В	ОБ	В	ОМ	БВН
ER_{122}	П	В	В	В	В	ОБ	В	М	БНВ
ER_{123}	П	В	В	В	В	ОБ	В	В	БНВ

ER_{124}	П	В	В	В	В	ОБ	В	ОБ	БВН
ER_{125}	П	В	ОБ	В	В	В	В	ОМ	БВН
ER_{126}	П	В	ОБ	В	В	В	В	М	БНВ
ER_{127}	П	В	ОБ	В	В	В	В	В	БНВ
ER_{128}	П	В	ОБ	В	В	В	В	ОБ	БВН
ER_{129}	П	В	ОБ	В	В	ОБ	В	ОМ	В
ER_{130}	П	В	ОБ	В	В	ОБ	В	М	БВН
ER_{131}	П	В	ОБ	В	В	ОБ	В	В	БВН
ER_{132}	П	В	ОБ	В	В	ОБ	В	ОБ	В
ER_{133}	Н	ВС	В	В	В	В	В	ОМ	БНВ
ER_{134}	Н	ВС	В	В	В	В	В	М	Н
ER_{135}	Н	ВС	В	В	В	В	В	В	Н
ER_{136}	Н	ВС	В	В	В	В	В	ОБ	БНВ
ER_{137}	Н	ВС	В	В	В	ОБ	В	ОМ	БВН
ER_{138}	Н	ВС	В	В	В	ОБ	В	М	БНВ
ER_{139}	Н	ВС	В	В	В	ОБ	В	В	БНВ
ER_{140}	Н	ВС	В	В	В	ОБ	В	ОБ	БВН
ER_{141}	Н	ВС	ОБ	В	В	В	В	ОМ	БВН
ER_{142}	Н	ВС	ОБ	В	В	В	В	М	БВН
ER_{143}	Н	ВС	ОБ	В	В	В	В	В	БВН
ER_{144}	Н	ВС	ОБ	В	В	В	В	ОБ	БВН
ER_{145}	Н	ВС	ОБ	В	В	ОБ	В	ОМ	В
ER_{146}	Н	ВС	ОБ	В	В	ОБ	В	М	БВН
ER_{147}	Н	ВС	ОБ	В	В	ОБ	В	В	БВН
ER_{148}	Н	ВС	ОБ	В	В	ОБ	В	ОБ	В
ER_{149}	Н	В	В	В	В	В	В	ОМ	БВН
ER_{150}	Н	В	В	В	В	В	В	М	БВН
ER_{151}	Н	В	В	В	В	В	В	В	БВН
ER_{152}	Н	В	В	В	В	В	В	ОБ	БВН
ER_{153}	Н	В	В	В	В	ОБ	В	ОМ	В
ER_{154}	Н	В	В	В	В	ОБ	В	М	БВН
ER_{155}	Н	В	В	В	В	ОБ	В	В	БВН
ER_{156}	Н	В	В	В	В	ОБ	В	ОБ	В
ER_{157}	Н	В	ОБ	В	В	В	В	ОМ	В
ER_{158}	Н	В	ОБ	В	В	В	В	М	БВН
ER_{159}	Н	В	ОБ	В	В	В	В	В	БВН

ER_{160}	H	B	ОБ	B	B	B	B	ОБ	B
ER_{161}	H	B	ОБ	B	B	ОБ	B	ОМ	К
ER_{162}	H	B	ОБ	B	B	ОБ	B	M	B
ER_{163}	H	B	ОБ	B	B	ОБ	B	B	B
ER_{164}	H	B	ОБ	B	B	ОБ	B	ОБ	К

At the second stage, after applying the DR, created on the basis of fuzzy parameters, the rules developed on the basis of clear parameters should be applied to verify the decision made and to carry out the final categorization of the intruder's personality. So they will look as follows:

$D = \{UID = 1, AtEF = PHP\ or\ Java\text{-}script,$

$UPr = 1, TrFin = 1, ModF = 1, TrFout = 0, KS = 1\}$ - for a disinformer,

$S = \{UID = 0, AtEF = PHP, UPr = 1, TrFin = 1,$

$ModF = 0, TrFout = 0, KS = 1\}$ - for a spammer,

$C = \{UID = 1, AtEF = .exe,\ .com, UPr = 1,$

$TrFin = 1, ModF = 1, TrFout = 1, KS = 1\}$ - for a cracker,

$H = \{UID = 1, AtEF = script, UPr = 1,$

$TrFin = 0, ModF = 1, TrFout = 1, KS = 1\}$ - for a hacker,

$SB = \{UID = 0, AtEF = PHP, UPr = 1,$

$TrFin = 1, ModF = 0, TrFout = 0, KS = 0\}$ - for a spam bot and

$B = \{UID = 0, AtEF = script, UPr = 1,$

$TrFin = 0, ModF = 1, TrFout = 1, KS = 0\}$ - for a hacker bot.

The obtained rules are used to develop PWIS based on expert methods and based on fuzzy logic (Khosravi et all, 2020).

Conclusions. Intruder identification in the cyberspace is not simple research task because cyberspace is complex and non-formalised space as well as security side faces with conditions of uncertainty. But there are some parameters in the cyberspace that can be monitored and analised. Intruder influences on these and the assessment process can explain the character of intruder action as well as his attack strategy. In their previous works, the authors described the parameters by which the intruder is identified – these are host and network parameters. Since the process of detecting and identifying an intruder takes place under conditions of uncertainty, and a number of certain parameters of systems for early detection of attacks are

fuzzy, the functioning of such a system should be based on fuzzy logic. This theory has chosen as the basis of research study.

In the work, based on the proposed parameters, using the MLTS, there were introduced LVs and were created models of the standards of parameters Tlog, Nlog, TSlog, I, CPU, Muse, NEF, NEr, RTPr/F. Also, for each LV, the FV were calculated and the graphs of their terms were plotted. Formed standards necessary for the formation of logical rules to ensure the functioning of the SEDA. The results obtained will be further used to create an IDS system based on honeypot technology (other technologies also can be used for attacks detection).

In addition, the proposed DR model using fuzzy logic for the first group of rules and conventional logic for the second group makes it possible to display the anomalous state by using the set "influence of the intruder-parameter", "influence of the intruder-set of logical-linguistic connections" and the universal model of parameter standards in IS, generated by the influence of an intruder of information security of a certain type. Based on this model, there were developed examples of rules to detect and identify the activities of a disinformer, spammer, cracker, hacker, spam bot and hacker bot, which can be used to improve existing or develop a new system for early detection of APT-attacks directed on the critical information infrastructure of the state or other important objects.

Information about authors:

Avkurova Zhadyra – Master of Technical Sciences, Senior Lecturer, Department of Artificial Intelligence Technology *https://www.scopus.com/authid/detail.uri? authorId=57226553845* ID *https://orcid.org/0000-0002-0706-6075*;

Sergiy Gnatyuk – Doctor of Technical Sciences, Professor, Faculty of Cybersecurity, Computer and Software Engineering, National Aviation University, Kyiv, Ukraine, *sergio.gnatyuk@gmail.com*, *https://www.scopus.com/authid/detail.uri?authorId=36184129600*, *https://orcid.org/0000-0003-4992-0564*;

Abduraimova Bayan – Candidate of technical sciences, associate professor ENU L.N.Gumilyov, abduraimova_bk@enu.kz, https://orcid.org/0000-0003-3913-1895;

Kydyralina Lazat – NAO "Shakarim University in Semey", PhD, acting associate professor, *lazat_75@mail.ru*, https://orcid.org/0000-0002-2836-0919.

REFERENCES

Avkurova Zh., Abduraimova B., Gnatyuk S., Gizun A. Analiz sovremennyh sistem obnaruzheniya atak na osnove tekhnologij virtual'noj primanki // №6(142) Vestnik KazNITU, s.654-659, noyabr', 2020. (in Kaz.).

Iashvili G., Avkurova Zh., Iavich M., Bauyrzhan M., Gagnidze A., Gnatyuk S. Content-Based Machine Learning Approach for Hardware Vulnerabilities Identification System, Lecture Notes on Data Engineering and Communications Technologies, Vol. 83, pp. 117-126, 2021. (in Eng.).

Zuzčák M. and Bujok P., "Causal analysis of attacks against honeypots based on properties of countries," in IET Information Security, vol. 13, no. 5, pp. 435-447, 9 2019, doi: 10.1049/iet-ifs.2018.5141. (in Eng.).

Zhang W., Zhang B., Zhou Y., He H. and Ding Z., "An IoT Honeynet Based on Multiport Honeypots for Capturing IoT Attacks," in IEEE Internet of Things Journal, vol. 7, no. 5, pp. 3991-3999, May 2020, doi: 10.1109/JIOT.2019.2956173. (in Eng.).

Golub V. Password protection [Electronic resource]: article / V. Golub // Relga. – 01.12. 2009. - No. 17 (197). - Access mode: *http://www.relga.ru/ Environ/WebObjects/tgu-www.woa/wa/Main?textid= 2516&level1= main&level2* (in Eng.).

Gizun A.I. Basic parameters for the identification of the violator of information security / A.I. Gizun, V.V. Volyanska, V.O. Риндюк, C.O. Hnatyuk // Information protection. - 2013. - Vol.15. - №1. - P. 66-75. (in Russ.).

Siddiqui S., Khan M.S., Ferens K. and Kinsner W., "Detecting advanced persistent threats using fractal dimension based machine learning classification", Proc. ACM Int. Workshop Secur. Privacy Anal. (IWSPA), pp. 64-69, 2016. (in Eng.).

Gornitska D.A. Determination of coefficients of importance for expert evaluation in the field of information security / D.A. Gornitska, Volyanska V.V., Korchenko A.O. // Information protection. - 2012. - №1 (54). - P. 108-121. (in Eng.).

Niu W., Zhang X., Yang G., Chen R. and Wang D., "Modeling attack process of advanced persistent threat using network evolution", IEICE Trans. Inf. Syst., vol. 100, no. 10, pp. 2275-2286, 2017. (in Eng.).

Ma Z., Li Q. and Meng X., "Discovering Suspicious APT Families Through a Large-Scale Domain Graph in Information-Centric IoT," in IEEE Access, vol. 7, pp. 13917-13926, 2019, doi: 10.1109/ACCESS.2019.2894509. (in Eng.).

Khosravi M. and Ladani B.T., "Alerts Correlation and Causal Analysis for APT Based Cyber Attack Detection," in IEEE Access, vol. 8, pp. 162642-162656, 2020, doi: 10.1109/ACCESS.2020.3021499. (in Eng.).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 52-70
https://doi.org/10.32014/2022.2518-1726.139
УДК 372.851.02., 372.800.4.02

М.А. Болатбек[1], К.Б. Багитова[2*], Ш.Ж. Мусиралиева[2]

[1]Пассау университеті, Германия, Пассау;
[2]Әл-Фараби атындағы Қазақ ұлттық университеті, Қазақстан, Алматы.
E-mail: *kbbagitova@gmail.com*

КИБЕРҚАУІПСІЗДІК МӘСЕЛЕЛЕРІН ТАБИҒИ ТІЛДІ ӨҢДЕУ ӘДІСТЕРІ АРҚЫЛЫ ШЕШУ ТАҚЫРЫБЫНА ЖҮЙЕЛІК ШОЛУ

Аннотация. Қазіргі таңда ақпараттық-коммуникациялық Интернет желісі адамзат өмірінің ажырамас бөлігіне айналды. Адамдардың жұмысы, жеке өмірі және қаржысы Интернет, мобильді компьютерлер мен электрондық бұқаралық ақпарат құралдары әлеміне бет бұра бас-тады. Адамдар «Твиттер», «ВКонтакте», «Facebook» және т.с.с. әлеу-меттік желілерді жаһандық байланыс орнату, пікір алмасу, білім алу және т.б. мақсаттарда пайдалануда. Өкінішке қарай, бұл кең таралған құбылыс бізді зиянды шабуылдарға, жеке өмірге қол сұғуға, алаяқтыққа және басқа да осындай қиыншылықтарды бұрынғыдан да көбейтеді.

Жеке пайдаланушылардың ғана емес, сонымен қатар ақпараттық ұйымдардың да бүкіл әлемдік кеңістікке белсенді қатысуы ұлттық қауіпсіздікті қамтамасыз ету бойынша ақпараттық-коммуникациялық технологиялар дамуының қазіргі тенденцияларына сәйкес келетін іс-шараларды ұйымдастыру қажеттілігін анықтайды. Сондықтан кибер-қауіпсіздік қауіпсіз және реттелген сандық әлемнің маңызды бөлігі болып табылады.

Google, Facebook және Twitter алпауыттары интернеттегі террористік мазмұнды жылдам анықтап, жою үшін жасанды интеллект (AI) технологиясын қолдануға уәде берді. IBM-де жоғарыда аталған әлеуметтік желілердегі барлық деректерді талдай алатын Watson әзірлемесі бар. Ресейде Платонның IT-авторы әлеуметтік желілерді

бақылау және қауіп-қатерлерді болжау жүйесін құруда. Германия үкіметі террористік актілерден кейін Интернеттегі террористермен күресу үшін ZITiS атты жаңа киберқауіпсіздік бөлімшесі құрылғанын жариялады. Мұндай жүйелер әзірге Қазақстанда жоқ. Осы себепті Интернеттегі веб-ресурстарға талдау ақпараттық қауіпсіздікті қамтамасыз етуші ұйымдар үшін аса өзекті болып табылады.

Түйін сөздер: киберқауіпсіздік, табиғи тілді өңдеу, әлеуметтік желі, интернет, қауіпсіздік, машиналық оқыту, терең оқыту, мәтінді жіктеу.

М.А. Болатбек[1], К.Б. Багитова[2*], Ш.Ж. Мусиралиева[2]

[1] Университет Пассау, Германия, Пассау;
[2]Казахский национальный университет им. аль-Фараби,
Казахстан, Алматы.
E-mail: *kbbagitova@gmail.com*

СИСТЕМАТИЧЕСКИЙ ОБЗОР ТЕМЫ РЕШЕНИЯ ЗАДАЧ КИБЕРБЕЗОПАСНОСТИ С ПОМОЩЬЮ МЕТОДОВ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

Аннотация. В настоящее время информационно-коммуникационная сеть Интернет стала неотъемлемой частью жизни человечества. Работа, личная жизнь и финансы людей стали уходить в мир Интернета, мобильных компьютеров и электронных средств массовой информации. Люди используют социальные сети в «Твиттер», «ВКонтакте», «Facebook» и т.д. для глобального общения, обмена мнениями, получения знаний и т.д. К сожалению, это распространенное явление делает нас еще более уязвимыми для вредоносных атак, вторжений в личную жизнь, мошенничества и других подобных неприятностей.

Активное участие не только отдельных пользователей, но и информационных организаций во всем мировом пространстве определяет необходимость организации мероприятий по обеспечению национальной безопасности, соответствующих современным тенденциям развития информационно-коммуникационных технологий. Поэтому кибербезопасность является важной частью безопасного и упорядоченного цифрового мира.

Гиганты Google, Facebook и Twitter пообещали использовать технологию искусственного интеллекта (ИИ) для быстрого обнаружения и

уничтожения террористического контента в Интернете. У IBM есть разработка Watson, которая может анализировать все данные из вышеупомянутых социальных сетей. В России IT-автор Платона создает систему мониторинга социальных сетей и прогнозирования рисков. Правительство Германии объявило о создании нового подразделения кибербезопасности под названием ZITiS для борьбы с террористами в Интернете после террористических актов. Таких систем в Казахстане пока нет. По этой причине анализ веб-ресурсов в Интернете является наиболее актуальным для организаций, обеспечивающих информационную безопасность.

Ключевые слова: кибербезопасность, обработка естественного языка, социальная сеть, Интернет, безопасность, машинное обучение, глубокое обучение, классификация текста.

M. Bolatbek[1], K. Bagitova[2*], Sh. Musiralieva[2]

[1]University of Passau, Germany, Passau;
[2]Al-FarabiKazakh National University, Kazakhstan, Almaty.
E-mail: *kbbagitova@gmail.com*

A SYSTEMATIC REVIEW ON CYBERSECURITY ISSUES USING NATURAL LANGUAGE PROCESSING TECHNIQUES

Abstract. Currently, the Internet information and communication network has become an integral part of human life. Work, personal life and finances of people began to go into the world of the Internet, mobile computers and electronic media. People use social networks Twitter, VKontakte, Facebook, etc. for global communication, exchange of opinions, gaining knowledge, etc. Unfortunately, this common phenomenon makes us even more vulnerable to malicious attacks, invasions of privacy, fraud and other similar troubles.

The active participation not only of individual users, but also of information organizations throughout the world space determines the need to organize measures to ensure national security that correspond to modern trends in the development of information and communication technologies. Therefore, cybersecurity is an important part of a secure and orderly digital world.

The giants Google, Facebook and Twitter have promised to use artificial intelligence (AI) technology to quickly detect and destroy terrorist content

on the Internet. IBM has a Watson development that can analyze all the data from the aforementioned social networks. In Russia, Plato's IT author creates a system for monitoring social networks and forecasting risks. The German government has announced the creation of a new cybersecurity unit called ZITiS to combat terrorists on the Internet after terrorist attacks. There are no such systems in Kazakhstan yet. For this reason, the analysis of web resources on the Internet is the most relevant for organizations providing information security.

Key words: cybersecurity, natural language processing, social network, Internet, security, machine learning, deep learning, text classification.

Кіріспе. Компьютерлік қауіпсіздік – есептеу құрылғыларын (компьютерлерді, смартфондарды және басқаларды), сондай-ақ компьютерлік желілерді (Интернетті қоса алғанда, жеке және жалпыға қолжетімді желілерді) қорғау үшін қолданылатын қауіпсіздік шаралары. Жүйелік әкімшілердің қызметі сандық жабдықтар, ақпараттық өріс және қызметтер кездейсоқ немесе рұқсатсыз кіруден, деректерді өзгертуден немесе жоюдан қорғалған және дамыған қауымдастықтағы компьютерлік жүйелерге тәуелділіктің артуына байланысты маңызды болып табылатын барлық процестер мен механизмдерді қамтиды.

Киберқауіпсіздік – деректердің құпиялылығын, тұтастығын және қол жетімділігін қамтамасыз ету үшін қауіпсіздік шараларын қолдану процесі. Жүйелік әкімші компьютерлердің, серверлердің жергілікті желісінің деректерін қоса алғанда, активтерді қорғауды қамтамасыз етеді. Сонымен қатар, ғимараттар мен ең бастысы қызметкерлер тікелей қорғауға алынады. Киберқауіпсіздікті қамтамасыз етудің мақсаты деректерді қорғау болып табылады. Деректердің қауіпсіздігін қамтамасыз ету мақсатында қарсы шаралар да қолданылуы мүмкін. Бұл шаралардың қатарына кіруді бақылау, қызметкерлерді оқыту, аудит және есеп беру, ықтимал тәуекелдерді бағалау, ену тестілері және авторизация талаптары кіреді (бірақ олармен шектелмейді).

Киберқауіпсіздік-бұл электрондық деректер мен ақпаратты қорғау. Бұл компьютерлер, ұялы телефондар, серверлер және желілер сияқты құрылғылардағы электрондық жүйелерді зиянды шабуылдардан қорғау.

Қазір кейбір елдерде киберқауіпсіздікті мектеп қабырғасынан бастап оқыту жоспарлануда. Мысалы, Ұлыбританияда оқушыларға Киберқауіпсіздік бойынша сабақтар ұсынылады, онда олар британдық компаниялар мен ұйымдардың хакерлердің желілік шабуылдарынан қауіпсіздігін қамтамасыз ету дағдыларын үйренеді. Оқушылармен

Киберқауіпсіздіктің нақты мәселелері және оларды шешу тәжірибесі қарастырылады. Бағдарлама 14-18 жас аралығындағы оқушыларға бағытталған.

Қиберқауіпсіздікке қатысты бірқатар деректерге тоқталатын болсақ:

1. Киберқауіпсіздіктің 95 пайызы адамның қателіктерінен болады. (Дүниежүзілік экономикалық форум).

2. Әлемдік ақпараттық қауіпсіздік нарығы 2028 жылы 366,1 миллиард долларға жетеді деп болжануда. (Fortune бизнес-аналитикасы).

3. 2020 жылы АҚШ кибер шабуылдардың 46 пайызын нысанаға алды, бұлкез-келген елге қарағанда ек іесе көп. (Microsoft Корпорациясы).

4. Бизнес жетекшілерінің 68 пайызы Киберқауіпсіздік бойынша тәуекелдер артып келеді деп санайды. (Акцент).

5. Орташаалғанда, Компания қалталарының тек беспайызы дұрыс қорғалған. (Варонис).

6. Деректердің ағып кетуі 22 жылы 2021 миллиард жазбаны анықтады. (Тәуекелге негізделген қауіпсіздік).

7. 2021 жылы бұзушылықтардың шамамен 70 пайызы қаржылық тұрғыдан негізделген, ал бес пайызданазы тыңшылыққа негізделген. (Verizon).

8. 2021 жылы бұзушылықтардың шамамен 40 пайызы фишингке, 11 пайызға жуығы зиянды бағдарламаларға және 22 пайызға жуығы бұзылуларға байланысты болды. (Verizon).

9. 2021 жылы 1862 деректердің ағуы тіркелді, бұл 2017 жылғы рекордтан 1506 ағып кетуден асып түсті. (CNET).

10. Зияндыэлектрондық пошта тіркемелерінің ең көп таралғантүрлері .docжәне -37 пайызы; келесі үлкені - .exe (19,5 пайыз). (Symantec).

Киберқауіпсіздік өте маңызды, себебі ол желідегі мәліметтерді оны ұрлағысы келетін және оны зиян келтіру үшін пайдаланғысы келетін кибершабуылдардан қорғауға қатысты процестерді қамтиды. Олардың қатарында құпия деректер, мемлекеттік және салалық ақпарат, жеке ақпарат, жеке басын анықтауға мүмкіндік беретін ақпарат, зияткерлік меншік және қауіпсіз медициналық ақпарат болуы мүмкін. Аталған деректерді қорғауға арналған алдыңғы қатарлы бағдарламалар мен киберқорғаныс механизмдерінің болуы өте маңызды. Қоғамның әр мүшесі ауруханалар мен басқа да денсаулық сақтау мекемелері, қаржылық қызметтер бағдарламалары және электр станциялары сияқты маңызды инфрақұрылымдардың қолданушысы болып табылады. Мысалы, қолданбаға кірген кезде немесе сандық денсаулық сақтау жүйелерінде құпия деректерді толтырған кезде аталған жүйелер,

желілер мен инфрақұрылымдар тиісті қорғанысқа ие болмаса, ол жердегі деректер теріс мақсатта пайдаланылуы мүмкін.

Жеке деңгейдегі кибершабуылдар адамның жеке басына қажетті ақпаратты ұрлауға және бопсалауға әкелуі мүмкін, бұл адамның өміріне айтарлықтай зиян келтіруі мүмкін.

Киберқауіпсіздіктің маңыздылығын айқындайтын себептерге тоқталып өтетін болсақ, ең алдымен жылдам кең жолақты байланыс, жетілдірілген гаджеттер және бұлтты есептеу сияқты технологияның қарқынды дамуы қосылған құрылғылардың көбеюіне әкелді. Бұл қараңғы Интернетте әр түрлі киберқылмыс түрлерін орындауға керемет мүмкіндік туғызды. Келесі себеп - технологияны пайдаланушылардың осалдығы. Қазіргі уақытта адамдардың көпшілігінің ақпараттық және коммуникациялық технологияларға көбірек сенім артатындығы киберқылмыскерлер үшін қылмыс жасау мүмкіндігінің тез өсіп келе жатқанын білдіреді. Бұлтты сақтаудың кеңеюі және әлеуметтік медианың өсуі сияқты факторлар көптеген адамдарды кибершабуылдарға осал етті. Бұл киберқауіпсіздікті бұрынғыдан да маңызды етеді. Сонымен қатар, банк деректемелері мен құпиясөздер сияқты құпия ақпаратты бұлтта сақтауға болады, бұл оның ұрлану қаупін арттырады. Сондай-ақ,әлеуметтік медианың өсуі де жеке деректердің алаяқтық жағдайда пайдаланылуының көбеюіне әкелді.

Соңғы зерттеулерге сәйкес, өткен жылы ұйым үшін киберқылмыстың орташа құны шамамен 13 миллион долларды құрады. Зерттеулер сонымен қатар қаржылық ақпарат, медициналық жазбалар, коммерциялық құпия, жеке мәліметтер және зияткерлік меншік сияқты ақпараттың күрт өсуін анықтады.

Компьютерлік вирустар өте тез таралуы мүмкін. Егер олар басқарылмаса, бұл бизнес үшін үлкен қиындықтар тудыруы мүмкін. Компьютерлік вирустар файлдар мен жүйелерді зақымдауы мүмкін. Сондықтан киберқауіпсіздікке байыпты қарау өте маңызды, өйткені бұл компьютерлік жүйені вирустардан құтқара алады.

Технологияның өсуі мен дамуы қара торды артта қалдырмады. Қараңғы веб-сайт – бұл тек мамандандырылған веб-шолғыштар арқылы қол жеткізуге болатын интернет-сайттардың құпия ынтымақтастығы. Ол негізінен интернет қызметін жасыру және пайдаланушылардың анонимділігі мен құпиялылығын сақтау үшін қолданылады.Қараңғы вебті заңды түрде қолдануға болады, бірақ ол көптеген заңсыз операциялардың орны ретінде де танымал. Есірткі мен адам саудасы, қаруды заңсыз тарату, бағдарламалық жасақтаманы тарату, заңсыз

аукциондар, қарақшылық және басқа да көптеген заңсыз әрекеттер қара торды қолданумен байланысты екені белгілі.

Киберқауіпсіздік өте маңызды, өйткені ол ұйымдарды ықтимал киберқауіптерден қорғайды. Технологияның дамуы көптеген адамдарды бұзу, деректерді ұрлау және бүлдіру, сондай-ақ өнеркәсіптік тыңшылық сияқты киберқылмыскерлердің әрекеттеріне осал етті. Киберқылмыс деңгейі өсуде, сондықтан киберқауіпсіздік болмаса, құпия ақпаратты, ақшаны немесе беделді жоғалту ықтималдығы артады.

Материалдар мен әдістер. Табиғи тілдегі мәтіндерді өңдеу (Natural Language Processing, NLP) – жасанды интеллект пен математикалық лингвистиканың жалпы бағыты. Ол табиғи тілдердегі мәтіндерді компьютерлік талдау және синтездеу мәселелерін зерттейді.

NLP-ды киберқауіпсіздік есептері үшін қолданудың артықшылықтарына келетін болсақ,

1. Терең оқыту алгоритмдері кез-келген ақпараттық корпустың құрылымдалмаған сипатын ескере отырып, Киберқауіпсіздік бойынша NLP деңгейінің жоғарылауының негізгі қозғаушы күші болып табылады.

2. NLP лексикалық талдауы үшін терең оқытуды қолдану екілік файлдар немесе бастапқы кодтар сияқты стандартты емес тілдердегі жағдайларды талдауға мүмкіндік береді.

3. NLP – бұл анықтау, тергеу және жауап беруді үйлестірудің өте пайдалы құралы, оның талдау және ақпарат алу мүмкіндіктерін ескере отырып тиімді пайдалануға болады.

Компьютерлер мен адам (табиғи) тілдерінің өзара әрекеттесуі тілді өңдеуге негізделгенін ескере отырып, тілдік деректердің үлкен көлемін өңдеу мен талдауда компьютерлердің біліктілігін арттыру маңызды бола түсуде. Осылайша, адам мен машинаның өзара әрекеттесуіндегі табиғи тілді өңдеудің рөлі, демек, киберқауіпсіздік әлемінде қарқын ала бастайды. Сонымен, NLP-ның киберқауіпсіздіктегі рөлі мен қолданылуы қандай болуы мүмкін?

(Spear-) фишинг алаяқтар алаяқтық құрбандарына арнайы электрондық пошталар, мәтіндік хабарламалар немесе телефон қоңырауларын жіберген кезде пайда болады. Негізгі ресурс ретінде әлеуметтік медианың пайда болуымен фишерлер одан да тиімді бола бастайды және NLP құралдары осы қосымшаларды қолдануға өте ыңғайлы. NLP ақпаратын жинау міндеттерін қолдана отырып, шабуылдаушылар әлеуметтік желілер мен басқа көздер арқылы жеке ақпаратты жинауды автоматтандырады, бұл шабуылды анықтауды қиындатады. Алайда, қорғаушылар осы шабуылдарға қарсы tdd сияқты құралдарды қолдана

алады. NLP-ді ұйымға қарсы фишингтік науқандарды анықтау үшін қолдануға болады.

Мүмкін болатын веб-шабуылдар мен қорғаныс механизмдерінің кең спектрі уақыт қатарлары немесе HTTP сұраулары / жауаптары болсын, терең оқыту мен NLP-ді қолдана алады. Шабуылдаушылар ақпарат жинау әдістерін немесе бәсекеге қабілетті машиналық оқытуды қолдана алады, ал қорғаушылар seq2seq atoencoders және басқа веб-қауіпсіздік модельдері сияқты NLP қосымшаларын қолдана отырып, өз позицияларын нығайта алады. Үнемі өсіп келе жатқан техникалық проблемаларға қарамастан, WAF және аномалияларды анықтау саласындағы нарық көшбасшылары өз өнімдеріне осындай мүмкіндіктерді енгізуді қарастыруда деп болжай аламыз.

Зиянды бағдарламалар мен кодтарды талдау туралы сөз болғанда, Endgame деректерді өңдеу мамандары зиянды кодты жақсы анықтау және түсіну үшін NLP-дің озық әдістерін қолданады. Олар зиянды бағдарламаларды талдауға арналған malicious Language Processing платформасын жасады. Оның мақсаты-NLP-ді жақсы кодта жасырылған зиянды кодты анықтауды автоматтандыру және жеделдету арқылы іске қосу. Лексикалық талдау тұжырымдамасын қолдана отырып, олар зиянды екілік файлдарды үлкен мәтін ретінде қарастырады. Осылайша, машиналар кодты оны орындамай-ақ "түсіне" алды. Сол сияқты, осалдықты экстраполяциялау арқылы осалдықты бағалау үшін NLP әдістерін кеңейтуге болады.

NLP көмегімен threat intelligence өнімдері бірнеше тілдегі сөздер мен техникалық деректердің мағынасын оқып, түсініп қана қоймайды, сонымен қатар заңдылықтарды анықтау үшін миллиардтаған мәліметтер нүктелерін қолданады.

NLP негізіндегі онтология IT-тәуекелдерді басқару мен кибер тұрақтылықты автоматтандыру мен біріктіруді қолдайды. Тәуекелдерді талдау және бағалау мәтіндік ақпаратты да қамтитынын ескере отырып, NLP пайдаланатын IT-тәуекелдерді басқаруға арналған өнімдер әртүрлі құрылымдар мен әдіснамалардың талаптарын IT-ортадан жинақталған деректермен салыстыра алады. Бұл өнімнің нормативтік және құқықтық талаптарға сәйкестігін қамтамасыз етуге мүмкіндік береді және құқық қорғау органдарымен қарым-қатынасты жеңілдетеді. Сонымен қатар, NLP контент-аналитикасының мүмкіндіктері нормативтік талаптардың өзгеруін тиімді бақылауға және талаптарға байланысты шығындарды бағалауды қолдауға мүмкіндік береді. NLP-ны қауіп-қатер модельдерін

жақсарту арқылы киберқауіпсіздік моделінің қаупін азайту үшін пайдалануға болады.

Нәтижелер. Электрондық коммерция және электрондық төлем жүйелері саласындағы соңғы жетістіктер несие карталарына қатысты қаржылық алаяқтық жағдайларының көбеюіне әкелді. Сондықтан несие картасындағы алаяқтықты анықтайтын құралдарды енгізу өте маңызды. Бұл мақалада белгілерді таңдау үшін генетикалық алгоритмді (GA) қолдана отырып, машиналық оқыту (ML) негізінде несие карталарын алаяқтықты анықтау механизмін ұсынған. Оңтайландырылған функцияларды таңдағаннан кейін, ұсынылған анықтау механизмі келесі машиналық оқыту жіктеуіштерін қолданады: шешім ағашы (DT), кездейсоқ орман (RF), логистикалық регрессия (LR), жасанды нейрондық желі (ANN) және аңғал Байес (NB). Тиімділікті тексеру үшін несие карталарының алаяқтықтарын анықтаудың ұсынылған механизмі еуропалық карта ұстаушыларынан алынған мәліметтер жиынтығын қолдана отырып бағаланады. Нәтиже ұсынылған тәсіл қолданыстағы жүйелерден асып түсетінін көрсетті (Ilebari etal., 2022:5).

Бұл мақалада (Alshehri et al., 2022:2) пайдаланушының мінез-құлқын талдаумен бірге машиналық оқытуды қолдана отырып, кибершабуылдарды анықтауға арналған жаңа платформа ұсынылған. Жақтаушы пайдаланушының мінез-құлқын оның әрекеттерін білдіретін оқиғалар тізбегі ретінде модельдейді. Содан кейін ұсынылған тізбектер жеке пайдаланушылардың ерекше мінез-құлқын анықтайтын белгілерді алу үшін нейрондық желінің қайталанатын моделіне орналастырылады. Осылайша, модель желідегі пайдаланушының мінез-құлқын қалыптастыру үшін тұрақты мінез-құлық жиілігін тани алады. Кейінгі процедура қайталанатын нейрондық желі белгісіз мінез-құлықты әдеттегі немесе тұрақты емес мінез-құлық ретінде жіктеу арқылы қалыптан тыс мінез-құлықты анықтайды. Ұсынылған құрылымның маңыздылығы кибершабуылдардың көбеюіне байланысты. Әдетте, ішкі шабуылдарды анықтау әлдеқайда қиын міндет, өйткені қауіпсіздік хаттамалары желідегі сенімді ресурстардан тұрады, соның ішінде пайдаланушылардан шабуылдарды тану қиынға соғады. Осылайша, пайдаланушының мінез-құлқын анықтауға болады және нәтижесінде қарапайым шаблондар әдеттегі желілік жұмыс процесін көрсететін терең заңдылықтарды тануға үйренеді. Эксперименттік нәтижелербұл тәсіл басқа тәсілдермен салыстырғанда жақсы нәтиже көрсеткенін және RNN-LSTM 1 көмегімен AUC 0,97 қол жеткізілгенін көрсетеді.

Заттар интернетінің (IoT) және киберфизикалық жүйелердің (CPS)

кеңеюін ескере отырып, киберқауіпсіздіктің ықтимал мәселелерін тиімді анықтау мен басқаруды дамыту ғана емес, сонымен қатар заттар интернетінің қауіпсіздігін қамтамасыз ету стандарттарына негізделген тиімді және бейімделгіш басқарумен байланысты мәселелерді шешу маңызды. Бұл зерттеу қолданыстағы стандарттарға кең және сыни зерттеулер жүргізеді және киберфизикалық желілерді кеңінен қолдануға қолдау көрсету үшін назар аудару керек бағыттарды анықтайды (Dong et al., 2022:7).

Фишинг – бұл сандық коммуникациядағы сенімді ұйым ретінде көрінетін құпия ақпаратты алуға бағытталатын алаяқтық әрекет. Бұл кибершабуылдың бір түрі, ол көбінесе сәтті болады, өйткені пайдаланушылар өздерінің осал тұстарын білмейді немесе тәуекелдерді түсіне алмайды. Бұл мақалада (Desolda et al., 2021:2) адам факторы мен фишинг саласындағы ең маңызды зерттеу жұмыстарының «бейнесін» салу мақсатында жүргізілген әдебиеттерге жүйелі шолу берілген. Әдебиеттерді жүйелі шолуда қарастырылған зерттеу мәселелеріне сәйкес алынған жарияланымдарды талдау фишингтік шабуылдардан қорғану үшін адам факторын қалай ескеру керектігін түсінуге көмектеседі.

Қауіп-қатер оқиғалары мен ымыраға келу индикаторларын бөлісу кибершабуылдарға қарсы тиімді қарсы шараларға қатысты тез және сыни шешім қабылдауға мүмкіндік береді. Алайда, қауіп-қатер туралы ақпарат алмасудың қолданыстағы шешімдері машинаны оқыту әдістерін (ML) қолдана отырып, қауіпті анықтау жүйелері (атап айтқанда, интрузияны анықтау жүйелері (IDS)) арасында ақпарат пен білімді оңай алмасуға мүмкіндік бермейді. Сонымен қатар, ML алгоритмдері үшін сенімді кірістерді жинаудың маңызды құрамдас бөлігі болып табылатын сарапшымен өзара әрекеттесу нашар қолдау табады. Осы мәселелердің барлығын шешу үшін ORISHA, қауіп-қатерді анықтау жүйелері мен ақпараттың басқа компоненттері арасында ынтымақтастық орнатуға мүмкіндік беретін ұйымдасқан ақпарат алмасу және ақпараттандыру платформасы ұсынылады. ORISHA-ға әр түрлі ұйымдарға тиесілі бірнеше қауіпті анықтау деңгейлерімен байланыс орнатуға мүмкіндік беретін зиянды бағдарламалар туралы ақпарат алмасу платформасының өзара байланысты даналары желісіне негізделген таратылған қауіп-қатерді талдау платформасы қолдау көрсетеді. Белгілі шабуылдарды анықтау сынағында жүргізілген эксперимент ұсынылған архитектураның дұрыстығын көрсетеді (Guarascio et al., 2022:6).

Соңғы жылдары қол жеткізілген жылдам технологиялық прогрестің арқасында көптеген адамдар өздерінің өмір салтын дәстүрлі бизнес тәсілдерінен электронды ресурстарға ауыстыруда. Аталған үдеріс берілген мақалада "қаскүнемдер" деп аталатын киберқылмыскерлердің назарын аударды (және әлі де жалғасуда), олар Интернет құрылымын фишинг сияқты киберқылмыскерлерді пайдаланушыларды жеке мәліметтерді, соның ішінде жеке ақпаратты, банктік және несие карталарын ашуға алдау үшін қолданады. Сенімді ұйымдардың заңды веб-сайттарының көшірмелері арқылы деректер, идентификаторлар, парольдер және маңызды ақпараттар ұрланады. Қазіргі таңда орын алған COVID-19 пандемиясы бұрын-соңды болмаған жағдай. Нәтижесінде, көптеген адамдар осы қауіпті жағдай туралы сенімді ақпарат жинауға тырысып, кибершабуылдарға осал болу үстінде. Өкінішке орай, осы жағдайды пайдаланып, пандемияға байланысты нақты шабуылдардың саны күрт өсті. Осы себепті корпорациялар мен киберқауіпсіздік зерттеушілері осы өсіп келе жатқан мәселені шешу үшін үнемі тиімді және инновациялық шешімдерді әзірлеуі керек. Қара тізімдерді, визуалды эффектілерді, эвристиканы және басқа да қорғаныс шешімдерін қолдану сияқты фишингпен күресудің бірнеше тәсілдері қазірдің өзінде қолданылып жатқанына қарамастан, олар фишинг шабуылдарының алдын алуға тиімді бола алмайды. Берілген мақалада авторлар COVID-19-мен байланысты домендік атауларды зиянды немесе заңды деп жіктеу үшін шектеулі функцияларды қолданатын машиналық оқыту модельдерін ұсынады. Алынған алғашқы нәтижелер домендік атаулардан алынған лексикалық белгілердің аз жиынтығы модельдерге жоғары балл алуға мүмкіндік беретіндігін, сонымен қатар, функция ретінде қосалқы домендер деңгейінің саны болжамдарға үлкен әсер етуі мүмкін екендігін көрсетеді (Mvula et al., 2022:3).

Киберқауіптердің үдемелі нашарлауы жағдайында олар туралы ақпаратты (CTI) ашық бастапқы қауіп-қатер туралы ақпаратты жариялау платформаларынан (OSTIPs) жинау ақпараттық қауіпсіздік қызметкерлеріне қоғамдық пікірді нақты анықтауға, төтенше жағдайларды жеңуге және тіпті қазіргі заманғы тұрақты қауіптерге қарсы тұруға көмектеседі. Алайда, жиіұсынылатын ақпарат көлемінің тез өсуіне байланысты CTI қолмен жинау тиімсіз болып шықты. OSTIPs-те жарияланған мақалалар құрылымданбаған, бұл CTI жазбаларын тек табиғи тілдерді өңдеу әдістерімен (NLP) автоматты түрде жинаудың шұғыл қажеттілігіне әкеледі. Осы шектеулерді жою үшін, берілген мақалада NLP әдісін, машиналық оқыту әдісін және киберқауіпсіздік

қауіптері туралы білімді біріктіретін көп типті кеңестерге (GCO) негізделген CTI жазбаларын құрудың автоматты тәсілі ұсынылған. Эксперимент нәтижелері GCO ұсынған мақалаларды жіктеудің және киберқауіпсіздік туралы мәліметтерді (CSIs) 93%-дан асатын дәлдікпенанықтайтынын көрсетеді, нәтижесінде Neo4j негізіндегі CTI деректер базасында жасалған жазбалар зиянды қауіптер тобын анықтауға көмектеседі (Sun et al., 2021:18).

Киберқауіпсіздік бойынша сарапшылар өз жұмысында NVD сияқты мәліметтер базасында сақталған білімге сүйенеді, бірақ бұл қауіптер мен осалдықтар туралы жалғыз ақпарат көзі емес. Бұл ақпараттың көп бөлігі әлеуметтік медиа арналары арқылы келеді. Бұл мақалада авторлар қауіпсіздік мамандары мен қарапайым пайдаланушылар онтологиялық көзқараста әртүрлі білім көздерін біріктіру арқылы семантикалық желі технологияларынан пайда көре алады деп мәлімдейді. Олар осалдықтардың онтологиясына негізделген, бірақ NLP құралдарымен толықтырылған, әлеуметтік желілерде киберқауіпсіздікке қатысты ақпаратты анықтауға және әртүрлі деректер көздеріне сұраныстарды іске қосуға арналған жүйені ұсынады. Биомедициналық салада дәлелденген киберқауіпсіздікті қамтамасыз ету үшін семантикалық желі технологиясының трансформациялық күші бағаланып, талқыланады (Aranovich et al., 2021:8).

Осалдықтардың ауырлығын тез және дәл бағалау және кибершабуылдарға қарсы шаралардың басымдықтарын анықтау үшін осалдықтар мен шабуылдар туралы ақпарат жинау қажет. Жалпы осалдықтар – оқиғаларды тізімдейтін сөздік, жалпы шабуыл үлгілерін тізімдеу және жіктеу-шабуыл үлгілерінің сөздігі. Шабуыл үлгілерін жалпы осалдықтарға тікелей сәйкестендіру және жіктеу қиын, өйткені олар әрқашан тікелей байланысты бола бермейді. Бұл жұмыста сөздіктер арасындағы ортақ байланыстарды тікелей іздеу тәсілін ұсынады. Содан кейін ұқсастық шаралары мен танымал алгоритмдердің жиынтығы болып табылатын бірнеше шаблондар, мысалы, терминдердің жиілігі – құжаттың кері жиілігі, әмбебап сөйлем кодтаушысы және BERT сөйлемдері ұсынылған тәсілді қолдана отырып эксперименталды түрде бағаланады. Жүргізілген эксперименттер term frequency–inverse document frequency алгоритмі ең жақсы жалпы өнімділікті қамтамасыз ететіндігін растайды (Kanakogi et al., 2022:6).

Адамның қарым-қатынасы олардың өзара әрекеттесуінің негізіндегі эмоцияларға байланысты. Әлеуметтік медианы қолданудың өсуі мәтіндік деректердің өзара байланысын интернеттегі өзара әрекеттесу арқылы

анықтауға болады. Мұндай өзара әрекеттесулер мәтіндік хабарламалар, электрондық пошталар және әлеуметтік медиа хабарламалары түрінде қол жетімді мәтіндік деректердің көптігіне әкеледі. Адамның қарым-қатынасын анықтау және талдау киберқауіпсіздіктен бастап қоғамдық денсаулыққа дейінгі көптеген қосымшалар үшін пайдалы. Бұл мақалада авторлар RIEA (эмоцияны талдау арқылы қарым-қатынасты анықтау) деп аталатын әдісті ұсынады, олардың арасындағы әңгімені талдау арқылы бірнеше зияткерлік агенттер арасындағы қатынасты анықтауға болады. Берілген жұмыстың мақсаты – эмоцияларды шығару және оларды қатынастар жиынтығына көрсету және уақыт өте келе қатынастардың қалай өзгеретінін талдау үшін когнитивті психология және табиғи тілді өңдеу (NLP) ұғымдарын біріктіру. Авторлар көптеген әңгімелерді талдау үшін психологиялық модельдерді қолданады және эмоциялар мен қатынастардың сәйкестігін анықтау үшін машиналық оқыту әдістерін қолданады. Олар жіктеу үшін ең жақсы масштабтау әдісін қолдана отырып, төрт түрлі ассоциативті сыныпты және төрт тіркеме стилін қолданады. Алынған нәтижелер RIEA тұлғааралық қатынастарды 85% дәлдікпен дұрыс анықтай алатындығын көрсетеді. Бағалау RIEA сөйлесулерден тұлғааралық қатынастарды дәл анықтай алатындығын және күрделі қатынастарды анықтау үшін кеңейтілуі мүмкін екенін көрсетеді. Бұл зерттеу сонымен қатар эмоционалды мінез-құлықтағы өзгерістердің уақыт өте келе қарым-қатынастың дамуына әсерін көрсетеді (Qamar et el., 2021:7).

Әлеуметтік медиа қылмыстарды жасау және анықтау үшін қолданылады. Автоматтандырылған әдістерді қолдана отырып, қылмысты ашуды да кеңейтуге болады. Қылмыскерлердің көптеген адамдарды қамту қабілеті бұл аймақты жиі зерттеу тақырыбына айналдырды, сондықтан әлеуметтік платформаларда жасалған нақты қылмыстарды қарастыратын бірнеше сауалнамалар жүргізілді. Осы уақытқа дейін әлеуметтік желілердегі қылмыстардың барлық түрлерін, олардың ұқсастықтарын, сондай-ақ олардың ашылуын қарастыратын шолу мақаласы болған жоқ. Қылмыстар мен оларды ашу әдістерінің ұқсастығын көрсету домендер арасында әдістер мен деректерді жіберуге мүмкіндік береді. Осылайша, зерттеудің берілген мақсаты – әлеуметтік желілерде жасалған қылмыстарды құжаттау және олардың ұқсастықтарын қылмыс таксономиясы арқылы көрсету. Сонымен қатар, бұл сауалнама жалпыға қол жетімді деректер жиынтығын құжаттайды (Drury et al., 2022:4).

Фишинг жыл сайын миллиардтаған доллар шығындарға алып

келеді және интернет-экономикаға үлкен қауіп төндіреді. Фишингтік шабуылдар қазіргі уақытта көбінесе электрондық пошта арқылы жүзеге асырылады. Фишингтік электрондық поштаны анықтаудың қазіргі зерттеу тенденциясын жақсы түсіну үшін бірнеше зерттеу жұмыстары жүргізілді. Алайда, бұл мәселені әр түрлі тұрғыдан бағалау маңызды. Тек бірнеше баламаларды зерттей отырып, жіктеу және оқыту мақсаттары үшін NLP әдістерін қолдануға жарық берген бір жұмысты қоспағанда, бірде-бір сауалнама фишингті анықтау үшін табиғи тілдерді өңдеу әдістерін (NLP) қолдануды ешқашан жан-жақты зерттемеген. Бұл олқылықтың орнын толтыру үшін берілген зерттеудің мақсаты фишингтік электрондық пошталарды анықтау үшін NLP қолдану бойынша зерттеулерді жүйелі түрде шолу және қорытындылау болып табылады. Алдын-ала анықталған критерийлер негізінде 2006 жылдан 2022 жылға дейін жарияланған 100 ғылыми мақала іріктеліп, талданды. Авторлар NLP көмегімен фишинг хаттарын анықтау бойынша негізгі зерттеу салаларын, фишингті анықтау үшін электрондық пошталарда қолданылатын машиналарды оқыту алгоритмдерін, фишинг хаттарындағы мәтіндік функцияларды, фишинг хаттарында қолданылған мәліметтер жиынтығы мен ресурстарды және бағалау критерийлерін зерттеген. Көптеген жіктеу алгоритмдерінің ішінде фишинг хаттарын анықтау үшін тірек векторлық машиналар (SVM) кеңінен қолданылады. NLP-тің ең көп қолданылатын әдістері-TF-IDF және сөздерді ендіру. Сонымен қатар, фишингтік электрондық поштаны анықтау әдістерін салыстырмалы талдау үшін ең көп қолданылатын мәліметтер жиынтығы-Nazario phishing corpus. Ұсынылған жұмыстарды талдау NLP әдістерін қолдана отырып, араб тіліндегі фишингтік электрондық пошталарда көп жұмыс жүргізілмегенін көрсетті (Salloum et al., 2022:22).

SMS арқылы фишингтік алаяқтық смартфондардың кеңінен қолданылуына және мобильді интернет технологиясының қол жетімділігіне байланысты жиі кездеседі. Құрылымданбаған қысқа мәтіндерді талдау арқылы фишингтік SMS-хабарламаларды анықтау жасанды интеллектке негізделген киберқауіпсіздік саласындағы күрделі міндет болып табылады. Табиғи тілді өңдеумен біріктірілген машиналық оқытуға негізделген әдістер фишинг пен заңды SMS хабарламаларының арасындағы айырмашылықты анықтауда үлкен әлеуетке ие. Бұл мақалада авторлар анықтамалық мәліметтер базасында бірнеше заманауи машиналарды оқыту алгоритмдерімен тәжірибе жасады. Сонымен қатар, фишингті анықтаудың автоматты стратегиясын құру

үшін NLP негізінде белгілерді алу және белгілерді таңдау кезеңдері кіреді. Белгілерді алу және таңдаудан кейін қолданылған кезде тірек векторларының машиналық жіктеуіші заңды SMS үшін F1 99,08% және дәлдігі 98,39% құрады. Тексерілген әдістердің тиімділігі бақылау жиынтығындағы танымал бағалау көрсеткіштерінің көмегімен бағаланған (Ulfath et al., 2021:8).

Кибернетикалық осалдықтарды анықтаудың дәстүрлі әдістерінің кемшіліктері жаңа қауіптерді анықтауға, оларды ортақ осалдықтардың (CVE) жазбаларына тіркеуге және оларды жалпы осалдықтарды бағалау жүйесі (CVSS) арқылы бағалауға қажетті уақытпен байланысты. Бұл проблемаларды әлеуметтік медиа мен ашық бастапқы деректерге негізделген осалдықтарды ерте анықтау жүйелері арқылы жеңілдетуге болады.Бұл жұмыста киберқауіпсіздік туралы жаңалықтардағы кибернетикалық осалдықтар Open Source Intelligence (OSINT) көмегімен ерте кибернетикалық қауіптерді автоматты түрде анықтау жүйесінің бөлігі ретінде анықтауға бағытталған модель ұсынылады. Машиналарды оқытудың үш моделі киберқауіпсіздік туралы мақалаларды тиісті (яғни қауіпсіздікке жаңа қауіп төндіретін) немесе маңызды емес деп жіктеудің берік негізін құру үшін 1000 таңбаланған жаңа мақалалар жиынтығында оқытылды: тірек векторлық машинасы, аңғал Байес классификаторы және BERT моделі. BERT моделі сынақ жиынтығында 88,45% орташа дәлдікпен ең жақсы өнімділікті көрсетті. Алынған тәжірибелер табиғи тілдерді өңдеу модельдері (NLP) киберқауіпсіздік туралы жаңалықтар мақалаларынан тиісті ақпаратты алу үшін осалдықты ерте анықтау жүйелері үшін қолайлы таңдау болып табылады деген қорытындыға әкеледі (Iorga et al., 2020:5).

Бұл жұмыста киберқауіпсіздік оқиғалары туралы ақпаратты алатын және киберқауіпсіздік туралы мәліметтерді графикке интеграциялаудың түпкі мақсаты бар семантикалық модельді толтыратын жүйені ұсынылады. Ол 2017-2019 жылдардағы ағылшын тіліндегі 1000 жаңалықтан тұратын жаңа корпуста оқытылды, олар оқиғаларға негізделген егжей-тегжейлі аннотациялармен белгіленген және кибершабуылдарды да, осалдықтармен байланысты оқиғаларды да қамтиды. Ұсынылатын модель оқиғаның бес түрін, олардың семантикалық рөлдерін және оқиғаға қатысты дәлелдердің 20 түрін (мысалы, файл, құрылғы, бағдарламалық жасақтама, ақша) анықтайды. CASIE терең нейрондық желілерге назар аудара отырып, әртүрлі тәсілдерді қолданады және бай лингвистикалық мүмкіндіктер мен сөздерді ендіруді қамтуы мүмкін. Авторлар оқиғаларды анықтаудың әр компонентімен

тәжірибе жүргізген және нәтижелер әрбір ішкі жүйенің жақсы жұмыс істейтінін көрсетеді (Satyapanich et al., 2020:2).

Хакерлік форумдар және басқа да әлеуметтік платформалар киберқауіпсіздік қауіптері туралы маңызды ақпаратты қамтуы мүмкін. Бірақ осы көздерден тиісті қауіп-қатер туралы ақпаратты алу үшін қолмен талдауды қолдану көп уақытты қажет ететін және қателікке бейім процесс болып табылады, ол айтарлықтай ресурстарды бөлуді қажет етеді. Бұл мақалада авторлар хакерлік форумдардағы қауіп-қатер туралы ақпаратты тез табу үшін машиналық оқыту әдістерінің әлеуетін зерттейді. Нақты хакерлік форумнан алынған мәтіндік деректерді қолдана отырып, авторлар мәтінді конвульсиялық нейрондық желі әдістерімен жіктеудің тиімділігін машиналық оқытудың дәстүрлі тәсілдерімен салыстырады.Олар машиналық оқытудың дәстүрлі әдістері, мысалы, тірек векторлар машинасы, конвульсиялық нейрондық желілердің алгоритмдерімен салыстырылатын өнімділіктің жоғары деңгейін қамтамасыз ете алатындығын анықтады (Deliu et al., 2017:8).

Талқылау. Қазіргі уақытта Интернеттегі киберқауіпсіздік туралы мәліметтер тез өсуде, бірақ олардың көпшілігі қауіпсіздікті талдау үшін уақытында түсіну қиын және автоматтандырылған қауіпсіздік жүйелерін тікелей пайдалануға жарамсыз құрылымданбаған мәтіндік мәліметтер болып табылады. Киберқауіпсіздік туралы ақпаратты құрылымданбаған мәтіндік көздерден құрылымдық көріністерге нақты уақыт режимінде автоматты түрде ауыстыру киберқауіпсіздік талдаушыларына кибернетикалық жағдайды жақсы түсінуге көмектеседі. Аталған нысандарды тану (NER) құрылымданбаған деректерді құрылымдық деректерге айналдыра алады. Жақында Transformers (BERT) ұсынған Bidirectional Encoder Representations деп аталатын тілдік бейнелеу моделі NLP-нің әртүрлі тапсырмаларында айтарлықтай жақсартуларға қол жеткізді. Бұл мақалада авторлар Bert және оның бүкіл әлемдегі BERT жетілдірілген нұсқасын (Bert wwm) NER киберқауіпсіздік есебіне қолданады. Олар BERT моделін BiLSTM-CRF архитектурасымен біріктіреді және эксперимент көрсеткендей, ұсынылатын әдіс F1 дәлдігі, Recallбағалауы бойынша қазіргі қолданыстағы модельге қарағанда жоғары өнімділікті қамтамасыз ететінін көрсетті (Zhou et al., 2021: 2).

Трафикті анықтау соңғы жылдары интрузияны анықтау жүйелерінде (IDS) маңызды рөл атқаруда. Бұл мақалада табиғи тілді өңдеу арқылы (NLP) пакеттік деңгейдегі трафикті анықтаудың жаңа тәсілі ұсынылған, ол ендірудің үлгісі ретінде қарапайым кон-

трастты ұсыныс қосымшаларын (SimCSE) қолданады. Жаңа тәсіл шикі пакеттік мәліметтерге негізделген трафиктің ерекшеліктерін зерттеуге мүмкіндік береді. Ұсынылатын әдісті бағалау үшін екі белгілі мәліметтер жиынтығымен тәжірибелер жүргізілді. Зиянды әрекетті анықтау үшін ұсынылатынмодель USTC-TFC2016 деректер жиынтығында 99,99% дәлдікке қол жеткізді, ал виртуалды жеке желі (VPN) қызметін анықтау үшін ұсынылатын модель ISCXVPN2016 деректер жиынтығында 99,98% дәлдікке қол жеткізді. Сонымен қатар, алынған модель нөлдік күндік шабуылдарды анықтау үшін тиімді екендігі анықталды, бұл модельдің бұрын байқалмаған шабуылдарды анықтау қабілетін көрсетеді. Тәжірибелер көрсеткендей, ұсынылатын тәсіл желілік трафикті тиімді анықтай алады және көптеген басқа заманауи әдістерден асып түседі (Bar et al., 2022:5).

Біреуді өз білімімен бөлісуге сендіру әдісі әлеуметтік инженерия ретінде белгілі. Әлеуметтік инженерлер адамдардың құнды ақпарат алмасудың салдарын білмеуіне, сондай-ақ олардың жүйелері мен ақпараттық технологиялар инфрақұрылымын қауіпсіздік шабуылдарынан қорғау туралы білімінің болмауына сүйенеді. Бұл шабуылдарды ұйым қызметкерлері үшінші тарап агенттігі арқылы жүзеге асыра алады. Олар қаржылық пайда немесе кек алу үшін ұйым ережелерін бұзады. Зиянкес зардап шеккендердің құпия ақпаратын жинау үшін әртүрлі тактикаларды қолданады, бұл – әлеуметтік инженерияға шабуыл жасау әдісі. Құпия ақпаратты заңсыз алу процесі қылмыстық әрекет болып табылады. Берілген зерттеу табиғи тіл процесі (NLP) арқылы белгісіз көзден немесе URL мекен-жайынан алынған хабарламаның спам немесе заңды екенін анықтау үшін құрылым ұсынады. COVID-19 кезінде көптеген адамдар интернетті күнделікті іс-әрекеттері үшін ол жердегі қауіпсіздік қатерлерін білместен қолдана бастады. Бұл шабуылдаушылардың осы құрбандарды нысанаға алуына және шабуылдарын тиімді түрде орындауына түрткі болды. SEA-бұл киберқауіпсіздікке қарсы шабуылдың бір түрі, ол бақылау жүйелерін бұзу үшін адамдардың табиғи қызығушылығын пайдаланады және сәттіліктің жоғары пайызына ие. Берілген зерттеудің мақсаты – COVID-19 пандемиясының әлеуметтік инженерлік шабуылдардың кеңеюіне қалай жол ашқаны, сондай-ақ осы шабуылдарды анықтау мен жеңілдетудің әртүрлі әдістері туралы егжей-тегжейлі зерттеу (Shalke et al., 2022:4).

Қорытынды. Бұл мақалада қазіргі таңда кейбір қауіпсіздік қатерлері мен кибершабуылдарға байланысты өте маңызды болып табылатын

киберқауіпсіздік мәселесі көтеріледі. Кибер қауіпсіздікке анықтама беріліп, қазіргі таңда жиі кездесетін шабуыл түрлері келтіріледі. Сонымен қатар, табиғи тілді өңдеу әдістері арқылы киберқауіпсіздіктің кей мәселелерін шешуге болатыдығы көрсетіліп, ағымдағы жұмыстарға кең әдеби шолу жасалған. Ғылыми жұмыс № AP06851248 «Мәтіндегі экстремистік бағытты анықтау үшін веб-ресурстардағы семантикалық талдау модельдерін, алгоритмдерін құрастыру және кибер-криминалистика құрал-жабдықтарын әзірлеу» атты жобаның аясында орындалды.

Information about the authors:

Bolatbek M.A. – University of Passau, researcher, Passau, Germany. Senior Lecturer, Department of Information Systems, Al-Farabi Kazakh National University, PhD, Almaty, Kazakhstan, E-mail: *bolatbek.milana@gmail.com*, *https://orcid.org/0000-0002-2153-180X*;

Bagitova K.B. – Doctoral student of the Department of Information Systems of the Al-Farabi Kazakh National University, Almaty, Kazakhstan. E-mail: *kbbagitova@gmail.com*, *https://orcid.org/0000-0003-1587-1995;*

Mussiraliyeva Sh.Zh. – Candidate of Physical and Mathematical Sciences, Head of the Department of Information Systems of the Al-Farabi Kazakh National University, Almaty, Kazakhstan. E-mail: *mussiraliyevash@gmail.com*, *https://orcid.org/0000-0001-5794-3649*.

REFERENCES:

Alshehri A., Khan N., Alowayr A. and Yahya M. Alghamdi, “Cyberattack detection framework using machine learning and user behavior analytics,” Computer Systems Science and Engineering, vol. 44, no.2, pp. 1679–1689, 2023.

Aranovich R., Wu M., Yu D., Katsy K., Ahmadnia B., Bishop M., Filkov V. and Sagae K. 2021. Beyond NVD: Cybersecurity meets the Semantic Web. In New Security Paradigms Workshop (NSPW ‘21). Association for Computing Machinery, New York, NY, USA, 59–69. https://doi.org/10.1145/3498891.3501259.

Bar R. and HajajC., “SimCSE for Encrypted Traffic Detection and Zero-Day Attack Detection,” in IEEE Access, vol. 10, pp. 56952-56960, 2022, doi: 10.1109/ACCESS.2022.3177272.

Deliu I., Leichter C., Franke K.. (2017). Extracting cyber threat intelligence from hacker forums: Support vector machines versus convolutional neural networks. 3648-3656. 10.1109/BigData.2017.8258359.

Desolda G.S., Ferro L., Marrella A., Catarci T. and Francesca Costabile M.. 2021. Human Factors in Phishing Attacks: A Systematic Literature Review. ACM Comput. Surv. 54, 8, Article 173 (November 2022), 35 pages. https://doi.org/10.1145/3469886.

Dong S., Cao J., Flynn D. et al. Cybersecurity in smart local energy systems: requirements, challenges, and standards. Energy Inform 5, 9 (2022). https://doi.org/10.1186/s42162-022-00195-7.

Drury B., Drury S.M., Arafatur Rahman Md, Ullah I., A social network of crime: A review of the use of social networks for crime and the detection of crime, Online Social Networks and Media,Volume 30,2022,100211,ISSN 2468-6964, https://doi.org/10.1016/j.osnem.2022.100211.

Guarascio M., Cassavia N., Sergio Pisani F., Manco G., Boosting Cyber-Threat Intelligence via Collaborative Intrusion Detection, Future Generation Computer Systems, Volume 135, 2022, Pages 30-43, ISSN 0167-739X,https://doi.org/10.1016/j.future.2022.04.028.

Ileberi E., Sun Y. & Wang Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. J Big Data 9, 24 (2022). https://doi.org/10.1186/s40537-022-00573-8.

Iorga D., Dragos-Georgian C., Octavian G., Cristian S., Mihai D., Razvan R. (2020). Early Detection of Vulnerabilities from News Websites using Machine Learning Models. 1-6. 10.1109/RoEduNet51892.2020.9324852.

Kanakogi K., Hironori W., Yoshiaki F., Shinpei O., Takao O., Takehisa K., Hideyuki K., Atsuo H. and Nobukazu Y. 2022. "Comparative Evaluation of NLP-Based Approaches for Linking CAPEC Attack Patterns from CVE Vulnerability Information" Applied Sciences 12, no. 7: 3400. https://doi.org/10.3390/app12073400.

Mvula P.K., Branco P., Jourdan G., Viktor H.L., COVID-19 malicious domain names classification, Expert Systems with Applications, Volume 204, 2022, 117553, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.117553.

Qamar S., Mujtaba H., Majeed H. et al. Relationship Identification Between Conversational Agents Using Emotion Analysis. Cogn Comput 13, 673–687 (2021). https://doi.org/10.1007/s12559-020-09806-5.

Salloum S., Gaber T., Vadera S. and Shaalan K., "A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques," in IEEE Access, vol. 10, pp. 65703-65727, 2022, doi: 10.1109/ACCESS.2022.3183083.

Satyapanich, Taneeya, Francis Ferraro and Tim Finin. "CASIE: Extracting Cybersecurity Event Information from Text." AAAI (2020).

Shalke C.J. and Achary R., "Social Engineering Attack and Scam Detection using Advanced Natural Langugae Processing Algorithm," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1749-1754, doi: 10.1109/ICOEI53556.2022.9776697.

Sun, Tianfang, Pin Yang, Mengming Li, and Shan Liao. 2021. "An Automatic Generation Approach of the Cyber Threat Intelligence Records Based on Multi-Source Information Fusion" Future Internet 13, no. 2: 40. https://doi.org/10.3390/fi13020040.

Ulfath Rubaiath E & Sarker, Iqbal & Chowdhury, Mohammad & Hammoudeh, Mohammad. (2021). Detecting Smishing-Attacks Using Feature Extraction and Classification Techniques. 10.1007/978-981-16-6636-0_51.

Zhou S., Liu J., Zhong X. and Zhao W., "Named Entity Recognition Using BERT with Whole World Masking in Cybersecurity Domain," 2021 IEEE 6th International Conference on Big Data Analytics (ICBDA), 2021, pp. 316-320, doi: 10.1109/ICBDA51983.2021.9403180.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF
THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 71-90
https://doi.org/10.32014/2022.2518-1726.140
УДК 51.74

А.К. Жумадиллаева[1], М.Д. Кабибуллин[1*], Б.Б. Оразбаев[1], К.Н. Оразбаева[2], Ж.Н. Тулеуов[3]

[1]Евразийский национальный университет им. Л.Н. Гумилева, Казахстан, Астана;
[2] Esil University, Казахстан, Астана;
[3]Атырауский нефтеперерабатывающий завод, Казахстан, Атырау.
E-mail: madyar_kabibullin@mail.ru

ОПТИМИЗАЦИЯ РЕЖИМОВ РАБОТЫ РЕАКТОРОВ РИФОРМИНГА УСТАНОВКИ КАТАЛИТИЧЕСКОГО РИФОРМИНГА НА ОСНОВЕ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

Аннотация. Исследованы проблемы оптимизации режимов работы реакторов риформинга, установки каталитического риформинга ЛГ-35-11/300-95, функционирующей на Атырауском нефтеперерабатывающем заводе в условиях неопределенности из-за случайного характера и нечеткости исходной информации, и предлагаются подходы к их решению. Оптимизацию режимов работы исследуемых объектов предлагается провести на основе компьютерного моделирования, для чего создается интеллектуализированная система оптимизации с использованием моделей, построенных на основе нечеткой информации в виде знания, опыта и интуиции экспертов предметной области. Для решения проблем дефицита исходной количественной информации и нечеткости доступной информации при разработке математических моделей предлагается системно использовать статистические методы, методы экспертной оценки, статистические методы и математического аппарата теорий нечетких множеств. Такой подход за счет эффекта синергизма и эмерджентности системы методов позволяет решить проблему неопределенности и построить эффективные модели сложных, нечетко описываемых объектов.

В работе на основе предлагаемого подхода и с использованием доступной информации статистического и нечеткого характера построены математические модели реакторов риформинга установки каталитического риформинга ЛГ-35-11/300-95 Атырауского нефтеперерабатывающего завода. Объем производимого катализата с реакторов риформинга определяется на основе модели в виде множественного регрессионного уравнения, построенного на экспериментально-статистических данных. Так как октановое число катализата (качество целевой продукции блока риформинга) характеризуется нечеткостью, для его оценки использована нечеткая информация в виде формализованных знаний и опыта экспертов, оценивающих качества катализата и построена нечеткая модель. Создается структура интеллектуализированной системы оптимизации режимов работы реакторов риформинга на основе компьютерного моделирования и описывается основной интерфейс этой системы, предназначенной для оптимизации режимов работы реакторов риформинга на основе компьютерного моделирования.

Ключевые слова: компьютерное моделирование, оптимизация, нечеткие модели, системный подход, пакет моделей, реакторы риформинга, катализат.

А.К. Жумадиллаева[1], М.Д. Кабибуллин[1*], Б.Б. Оразбаев[1], К.Н. Оразбаева[2], Ж.Н. Тулеуов[3]

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Казақстан, Астана;
[2]Esil University, Казақстан, Астана;
[3]Атырау мұнай өңдеу зауыты, Казақстан, Атырау.
E-mail: madyar_kabibullin@mail.ru

КАТАЛИТИКАЛЫҚ РИФОРМИНГ ҚОНДЫРҒЫСЫ РИФОРМИНГТЕУ РЕАКТОРЛАРЫ ЖҰМЫС РЕЖИМДЕРІН КОМПЬЮТЕРЛІК МОДЕЛЬДЕУ НЕГІЗІНДЕ ОПТИМИЗАЦИЯЛАУ

Аннотация. Бастапқы ақпараттың кездейсоқ және айқын еместігіне байланысты туындайтын анықсыздық жағдайында Атырау мұнай өңдеу зауытында жұмыс жасайтын ЛГ-35-11/300-95 каталитикалық

риформинг қондырғысының риформингтеу реакторларының жұмыс режимдерін оптимизациялау мәселелері зерттеліп, оларды шешу тәсілдемелері ұсынылады. Зерттелетін нысандардың жұмыс режимдерін компьютерлік модельдеу негізінде оптимизациялау үшін пән облысы мамандарының, яғни эксперттердің білімі, тәжірибесі және интуициясы түріндегі анық емес ақпарат негізінде құрастырылған модельдер негізінде интеллектуалдандырылған оптимизациялау жүйесі жасалады. Математикалық модельдерді құруда бастапқы сандық ақпараттың тапшылығы және қолда бар ақпараттың айқын еместігі мәселелерін шешу үшін статистикалық әдістерді, эксперттік бағалау тәсілдерін, статистикалық тәсілдерді және айқын емес жиындар теорияларының математикалық аппаратын жүйелі түрде қолдану ұсынылады. Мұндай тәсілдеме синергизм әсері мен жүйенің эмерджендік қасиеті есебінен анықсыздық мәселесін шешуге және күрделі, анық емес сипатталған объектілердің тиімді модельдерін құруға мүмкіндік береді.

Ұсынылған тәсілдер негізінде статистикалық және айқын емес сипаттағы қол жетімді ақпаратты пайдалана отырып, Атырау мұнай өңдеу зауытының ЛГ-35-11/300-95 каталитикалық риформинг қондырғысының риформингтеу реакторларының математикалық модельдері тұрғызылады. Реформингтеу реакторларынан өндірілетін катализаттың көлемі тәжірибелік-статистикалық мәліметтерге негізделген көпреттік регрессия теңдеуі түріндегі модель көмегімен анықталады. Катализатордың октандық саны (блогының мақсатты өнімінің сапасы) айқын еместікпен сипатталатындықтан, оны бағалау үшін катализат сапасын бағалайтын эксперттердің формализацияланған білімі мен тәжірибесі түрінде айқын емес ақпарат негізінде айқын емес модель құрылады. Компьютерлік модельдеу негізінде риформингтеу реакторлары жұмыс режимдерін оптимизациялау интеллектуалдандырылған жүйесінің құрылымы құрылды және осы жүйеде компьютерлік модельдеу негізінде риформингтеу реакторлардың жұмыс режимдерін оптимизациялауға арналған негізгі интерфейсі сипатталған.

Түйін сөздер: компьютерлік модельдеу, оптимизациялау, айқын емес модельдер, жүйелік тәсілдеме, модельдер пакеті, риформингтеу реакторлары, катализат.

A. Zhumadillayeva[1], M. Kabibullin[1*], B. Orazbayev[1], K. Orazbayeva[2], Zh. Tuleuov[3]

[1]L.N. Gumilyov Eurasian National University, Kazakhsta, Astana;
[2]Esil University, Kazakhstan, Astana;
[3]Atyrau Oil Refinery, Kazakhstan, Atyrau.
E-mail: madyar_kabibullin@mail.ru

OPTIMIZATION OF THE OPERATING MODES OF THE REFORMING REACTORS OF THE CATALYTIC REFORMING UNIT BASED ON COMPUTER MODELING

Abstract. The problems of optimizing the operating modes of the reforming reactors of the LG-35-11/300-95 catalytic reforming unit operating at the Atyrau Oil Refinery under uncertainty due to the random nature and fuzziness of the initial information are investigated, and approaches to their solution are proposed. It is proposed to optimize the operating modes of the objects under study on the basis of computer simulation, for which an intellectualized optimization system is created using models built on the basis of fuzzy information in the form of knowledge, experience and intuition of subject matter experts. To solve the problem of the lack of initial quantitative information and the fuzziness of available information in the development of mathematical models, it is proposed to systematically use statistical methods, peer review methods, statistical methods and the mathematical apparatus of fuzzy set theories. This approach, due to the effect of synergy and the emergence of a system of methods, allows us to solve the problem of uncertainty and build effective models of complex, indistinctly described objects.

Based on the proposed approach and using the available information of a statistical and fuzzy nature, mathematical models of the reforming reactors of the catalytic reforming unit LG-35-11/300-95 of the Atyrau oil refinery were developed. The volume of produced catalyzate from reforming reactors is determined on the basis of a model in the form of a multiple regression equation built on experimental and statistical data. Since the octane number of the catalyzate (the quality of the target product of the reformer unit) is characterized by fuzziness, fuzzy information was used to evaluate it, in the form of formalized knowledge and experience of experts evaluating the quality of the catalyzate, and a fuzzy model was developed. The structure of an intellectualized system for optimizing the operating modes of reforming

reactors based on computer simulation is created and the main interface of this system is described, which is designed to optimize the operating modes of reforming reactors based on computer simulation.

Key words: computer modeling, optimization, fuzzy models, systems approach, model package, reforming reactors, catalyzate.

Введение. Задачи оптимизации режимов работы химико-технологических систем (ХТС) и их основных агрегатов нефтеперерабатывающих, химических, а также других производств на основе их математических моделей относятся к весьма актуальной научно, практической задачей. На практике решение задачи оптимизации режимов работы ХТС усложняются тем, что эти системы характеризуются сложностью, множеством взамосвязанных параметров, многокритериальностью, которые как правило являются проиворечивыми, а также дефицитом и нечеткостью исходной информации (Кашин, 2011:135; Orazbayev et all, 2020). Все это намного усложняет процесс разработки математических моделей, необходимые для подготовки и принятия решений в условиях противоречивых критериев по выбору оптимальных режимов работы режимов ХТС.

Оптимальные режимы работы ХТС, которые определяются путем принятия решений эффективных решений на основе математических моделей системы позволяют повышать эффективность проимзводства (Засканов и др., 2013:175), (Липин, 2018:135). К наиболее эффективным средством многокритеривальной оптимизации технологических параметров сложных ХТС нефтепереработки и оптимального управления режимами их работы является различные автоматизированные системы на базе современных математических методов и компьютерной технологии (Муленко, 2015:73). В данной статье исследуются основные проблемы создания одного из перспективного вида таких систем, называемые интеллектуализированными системами компьютерного моделирования и оптимизации режимов работы ХТС (ИСКМО), которые предназначенны для поддержки принятия решений при управлении режимами работы ХТС с использованием методов искусственного интеллекта (Sansyzbay et all, 2020), (Isakov, 2018).

ИСКМО представляет собой информационно-вычислительную систему на базе современных компьютеров, функционирующая с участием человека и имеющая в своей обеспечивающей части систему моделей ХТС, эвристических алгоритмов принятия решений. Такие системы способны оказать интеллектуальную поддержку в

процессах подготовки и принятия решений по управлению объектами и функционируют с участием лица, принимающего решение (ЛПР) (Макаров и др., 2012).

Объектом исследования данной работы является реакторы риформинга ХТС установки каталитического риформинга ЛГ-35-11/300-95 Атырауского нефтеперерабатывающего завода (НПЗ). Установка каталитического риформинга ЛГ-35-11/300-95 предназначена для проведения процессов гидроочистки и риформинга бензиновых фракций (прямогонного бензина) и производства высокооктанового компонента автомобильных бензинов и технического водорода, который является сырьем нефтехимии. Технологический процесс каталитического риформинга прямогонного бензина, который протекает на реакторах риформинга установки ЛГ-35-11/300-95 является одним из важнейших процессов современной нефтепереработки и нефтехимии (Прокопюк и др., 2018). В этой связи в данной работе исследуются вопросы создания ИСКМО для выбора оптимального режима работы реакторов риформинга при управлении ими и предлагаются подходы к их решению.

Реакторный блок, в котором протекает процессы риформинга состоит из 4-х реакторов, которые соединены между собой последовательно (Р-2, Р-3, Р-4, Р-4, а) и параллельно (Р-4 и Р-4, а). Структура реакторов Р-2, Р-3, Р-4, Р-4, а и печи риформинга П-1 блока риформинга установки ЛГ-35-11/300-95 Атырауского НПЗ представлена на рисунке 1.

Как видно из приведенного рисунка 1, сырье риформинга (гидрогенизат с блока гидроочистки), подогретое в соответствующих секциях многосекционный печи П-1 последовательно проходят процесс риформинга в реакторах Р-2, Р-3, Р-4, Р-4, а. Последняя стадия процесса риформинга протекает в параллельно соединенных реакторах Р-4 и Р-4, а.

В настоящее исследование и решение проблем повышения экономической эффективности и экологической безопасности ХТС нефтепереработки на основе научно обоснованных методов с использованием современных математических методов, средств информационных технологий, например ИСКМО является актуальной задачей науки и практики. Рассмотрим результаты литературного анализа по теме исследования.

В работах (Засканов, 2013), (Carreno, 2014), (Абдулов, 2018), (Biegler et all, 2016) исследованы проблемы повышения эффективности и качества применяемых решений, на основе методов математического

моделирования, многокритериальной оптимизации и теорий принятия решений. Известны работы, в которых исследованы основные вопросы проектирования и создания различных автоматизированных систем в т. ч. систем поддержки принятия решений (Муленко, 2015), (Попов, 2019). Однако на практике функционируют многие ХТС нефтеперерабатывающего производства, которые характеризуются нечеткостью исходной информации и принятия решений по управлению ими на основе традиционных математических методов, не дает необходимых результатов или невозможны (Чернов, 2019), (Chen и др., 2018). Кроме того, постановка и решение задач принятия решений по управлению режимами работы таких количественно трудно описываемых ХТС усложняются еще из-за их сложности и многокритериальности объекта управления.

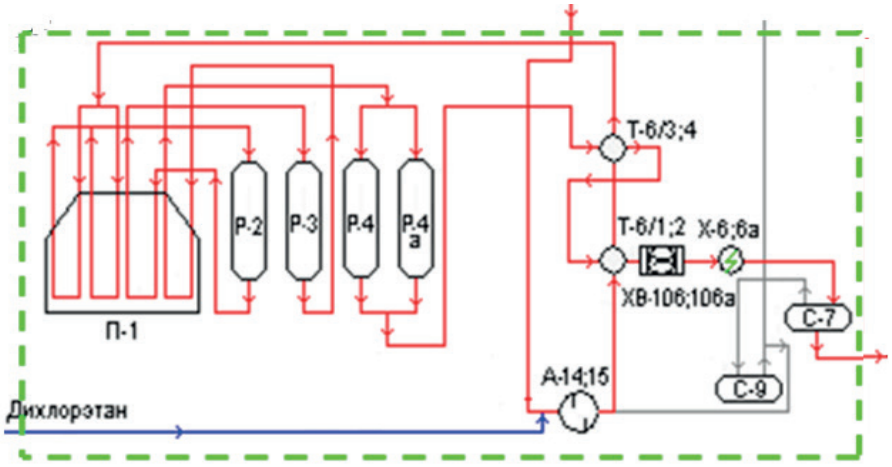


Рисунок 1. Структура реакторов и печи риформинга установки ЛГ-35-11/300-95

Результаты анализа источников, посвященных к решению вопросов моделирования, оптимизации и принятия решений сложными 6 нечетко описываемыми ХТС показывают, что полностью не исследованы и описаны концепции и принципы их моделирования и управления. В связи с этим исследование и решение проблем разработки моделей, формализация и решения задачи принятия решений по управлению сложными нечетко описываемыми ХТС на основе создания ИСКМО является весьма актуальной задачей науки и практики. Основная идея предлагаемой работы заключается в решение проблем дефицита и нечеткости исходной информации при разработке моделей и опти-

мизации режимов работы ХТС установки каталитического риформинга за счет использования знания, опыта и интуиции ЛПР, специалистов-экспертов. При этом для формализации и использования нечеткой и других видов информации используется методы экспертной оценки, статистические методы и математический аппарат теорий нечетких множеств.

В настоящее время производства высококачественных автомобильных бензинов без добавления присадок на установках каталитического риформинга и крекинга является весьма актуальной задачей нефтепереработки и экономики Казахстана. Данные технологии повышение качества моторных, хотя более дорогой, но отвечают современным требованиям экологических стандартов и нормативов по производству экологически безопасных топлив. В качестве самого эффективного подхода к решению рассмотренной задачи можно отметить применение методов математического моделирования и оптимизации, которые позволяют оптимально управлять режимами работы установки каталитического риформинга (Sharikov и др., 2013).

В процессе разработки математических моделей и алгоритмов оптимизации ХТС как ЛГ-35-11/300-95 Атырауского НПЗ, которая функционирует длительное время возникают проблемы неопределенности из-за дефицита, нечеткости исходной информации (Fayaz et all, 2018). Для разработки моделей, алгоритмов оптимизации и управления режимами работы таких ХТС, необходимо использовать доступную информацию различного характера, в т.ч. нечеткую информацию от ЛПР и экспертов, характеризующую работу системы.

Постановка задачи и методы исследования. Целью исследования является многокритериальная оптимизация режимов работы реакторов риформинга установки каталитического риформинга ЛГ-35-11/300-95 на основе компьютерного моделирования на основе построенных моделей реакторов в условиях дефицита и нечеткости исходной информации.

Для достижения сформулированной цели исследования решаются следующие основные задачи, являющиеся задачами исследования:

– на основе экспериментально-статистических данных и нечеткой информации от специалистов-экспертов предметной области разработка моделей реакторов риформинга установки каталитического риформинга ЛГ-35-11/300-95 Атырауского НПЗ;

– построение структуры интеллектуализированной системы компьютерного моделирования и оптимизации режимов работы реакторов установки каталитического риформинга;

– создание блока пакета моделей реакторов риформинга интеллектуализированной системы для оптимизации режимов работы реакторов риформинга на основе компьютерного моделирования работы объекта исследования.

Для разработки математических моделей реакторов установки каталитического риформинга ЛГ-35-11/300-95 на основе статистической и нечеткой информации используются методы системного анализа (Reverberi et all, 2016); методы построения статистических моделей (Карманов,2017), (Motlatsi et all, 2021) методы экспертной оценки и аппарат теорий нечетких множеств (Гуцыкова, 2017), (Рыжов, 2017), а также гибридный метод разработки математических моделей на основе информации различного характера (Orazbayev и др., 2018).

Результаты исследования. Модели реакторов установки каталитического риформинга ЛГ-35-11/300-95 Атырауского НПЗ. На основе экспериментально-статистических данных и нечеткой информации от экспертов идентифицированы структуры моделей, позволяющие определенить объема катализата с выхода реакторов риформинга Р-4, Р-4а и качества катализата в зависимости от входных, режимных параметров:

$$y_1^{P-4,4a} = a_0 + \sum_{i=1}^{5} a_i x_i + \sum_{i=1}^{5}\sum_{k=i}^{5} a_{ik} x_i x_k, \qquad (1)$$

$$\widetilde{y}_2 = \widetilde{a}_0 + \sum_{i=1}^{5} \widetilde{a}_i x_i + \sum_{i=1}^{5}\sum_{k=i}^{5} \widetilde{a}_{ik} x_i x_k, \qquad (2)$$

где $y_1^{R-4,R-4a}$ – объем катализата, с реакторов риформинга Р-4, Р-4а; $\widetilde{y}_2$ – нечеткое значение октанового числа катализата, оцениваемые экспертами в лаборатории; $x_i, i=\overline{1,5}$ – входные, режимные параметры реакторов риформинга, соответственно: x_1 – расход сырья (гидрогенизата; x_2 – объемная скорость в реакторах риформинга; x_3 и x_4 – температура и давление в реакторах Р-4, Р-4а; x_5– соотношение водород/сырье; a_0, a_i, a_{ik} и $\widetilde{a}_0, \widetilde{a}_i, \widetilde{a}_{ik}$ – соответственно четкие и нечеткие коэффициенты регрессии.

Таким образом модели определения объема катализата с блока риформинга идентифицированы в виде уравнений множественной регрессии (1), а и его качества (октановое число) – в виде нечетких уравнений множественной регрессии (2). Октановое число в производственных условиях непосредственно не измеряется, поэтому определяется в заводской лаборатории специалистами на основе его опыта и знаний и специальных средств и методики.

На основе обработки экспериментально-статистических с использованием пакета программ REGRESS, реализующий метод наименьших квадратов проведена параметрическая идентификация неизвестных коэффициентов модели (1):

$$y_2^{R-4,R-4a} = 0.3989835x_1 + 11.1869231x_2 - 0.00315895x_3 - 1.0239130x_4 + 0.0253700x_5 +$$
$$+ 0.0050697x_1^2 + 9.2899408x_2^2 - 0.0000585x_3^2 - 0.0445179x_4^2 + 0.0000491x_5^2 +$$
$$+ 0.2301827x_1x_2 + 0.0001003x_1x_3 + 0.00216839x_1x_4 + 0.00049873x_2x_3$$
$$- 0.5250836x_1x_4 - 0.0006867x_3x_4, \qquad (3)$$

Для идентификации нечетких коэффициентов модели (2) с помощью множества уровня проведены α срезы на трех уровнях 0,5, 0,75 и 1. Функции принадлежности нечетких коэффициентов построены в виде гауссово типа с помощью приложения Fuzzy Logic Toolbox системы MATLAB. В результате получены четкие значения идентифицируемых коэффициентов на α=0,5; 0,75; 1; 0,75 и 0,5 (Оразбаев и др., 2021):

$$\tilde{y}_2 = \left(\frac{0.5}{0.430000} + \frac{0.75}{0.433000} + \frac{1}{0.435000} + \frac{0.75}{0.437000} + \frac{0.5}{0.440000}\right)x_{14} -$$
$$-\left(\frac{0.5}{20.076906} + \frac{0.75}{20.076916} + \frac{1}{20.076923} + \frac{0.75}{20.076930} + \frac{0.5}{20.076938}\right)x_{24} +$$
$$-\left(\frac{0.5}{0.052810} + \frac{0.75}{0.052824} + \frac{1}{0.052834} + \frac{0.75}{0.052844} + \frac{0.5}{0.052858}\right)x_{34} -$$
$$-\left(\frac{0.5}{0.724870} + \frac{0.75}{0.724950} + \frac{1}{0.720000} + \frac{0.75}{0.725050} + \frac{0.5}{0.725130}\right)x_{44} +$$
$$+\left(\frac{0.5}{0.042209} + \frac{0.75}{0.042339} + \frac{1}{0.042439} + \frac{0.75}{0.042539} + \frac{0.5}{0.042669}\right)x_{54} +$$
$$+\left(\frac{0.5}{0.005198} + \frac{0.75}{0.005328} + \frac{1}{0.005438} + \frac{0.75}{0.005548} + \frac{0.5}{0.005688}\right)x_{14}^2 -$$
$$-\left(\frac{0.5}{15.443467} + \frac{0.75}{15.443637} + \frac{1}{15.446787} + \frac{0.75}{15.443937} + \frac{0.5}{15.443112}\right)x_{24}^2 +$$
$$+\left(\frac{0.5}{0.000007} + \frac{0.75}{0.000057} + \frac{1}{0.000107} + \frac{0.75}{0.000157} + \frac{0.5}{0.000207}\right)x_{34}^2 -$$
$$-\left(\frac{0.5}{0.030058} + \frac{0.75}{0.030138} + \frac{1}{0.030138} + \frac{0.75}{0.030278} + \frac{0.5}{0.030358}\right)x_{44}^2 +$$
$$+\left(\frac{0.5}{0.000004} + \frac{0.75}{0.000054} + \frac{1}{0.000104} + \frac{0.75}{0.000154} + \frac{0.5}{0.000224}\right)x_{54}^2 +$$

$$+\left(\frac{0.5}{0.000100}+\frac{0.75}{0.000170}+\frac{1}{0.000220}+\frac{0.75}{0.000270}+\frac{0.5}{0.000340}\right)x_{14}x_{34}+$$
$$+\left(\frac{0.5}{0.000125}+\frac{0.75}{0.000205}+\frac{1}{0.000265}+\frac{0.75}{0.000325}+\frac{0.5}{0.000405}\right)x_{14}x_{54}-$$
$$-\left(\frac{0.5}{0.557242}+\frac{0.75}{0.557492}+\frac{1}{0.557692}+\frac{0.75}{0.557892}+\frac{0.5}{0.558142}\right)x_{24}x_{44}+$$
$$+\left(\frac{0.5}{0.00006}+\frac{0.75}{0.000046}+\frac{1}{0.000086}+\frac{0.75}{0.000126}+\frac{0.5}{0.000166}\right)x_{34}x_{54};$$

Затем с применением пакета программ REGRESS на основе метода наименьших квадратов определены четкие значения регрессионных коэффициентов на выбранных α уровнях. Для расчета значения октанового числа катализа и компьютерного моделирования определенные значения коэффициентов регрессии на α срезах объединены на основе соответствующей формулы (Рыжов, 2017). Таким образом, после идентификации параметров на α уровнях и их объединения получена следующая модель оценки качества, т.е. октанового числа катализата, удобная для компьютерного моделирования:

$$y_2 = 0.4356500x_1 - 20{,}0758500x_2 - 0.05598558x_3 - 0.7200577x_4 + 0.0423567x_5 +$$
$$+0{,}0055003x_1^2 - 15.4456725x_2^2 + 0.0001147x_3^2 - 0.0230225x_4^2 + 0.0001237x_5^2 + 0.0002533x_1x_3 + \quad (4)$$
$$+0.0002887x_1x_5 - 0.5577895x_2x_4 + 0.0000957x_3x_5.$$

Фракционный состав катализата и влияющий на качества катализа, описываются тоже нечетко и также оцениваются с участием экспертов. Нечетких моделей, описывающие фракционный состав, можно построить на основе описанного выше подхода к оценке октанового числа.

Структура интеллектуализированной системы компьютерного моделирования и оптимизации режимов работы реакторов риформинга установки ЛГ-35-11/300-95. В процессе управления режимами работы ХТС ЛПР, например, оператор-технолог, часто попадает в сложную, когда для принятия наилучшего решения требуется проанализировать большой объем информации, сравнивать множество альтернатив по вектору критериев. Причем такие задачи усложняются тем, что критерии являются противоречивыми и ЛПР необходимо оценить последствия принимаемого решения в условиях неопределенности в нечеткой среде.

Эффективным подходом к решению таких сложных, трудноформализуемых задач является использование интеллектуализированных систем компьютерного моделирования и оптимизации на основе математических моделей (ИСКМО) на базе современных компьютеров.

Такая система позволяют эффективно объединить достижения методов моделирования, оптимизации на основе использования способностей ЛПР решать нечеткие задачи и возможностями современных компьютеров. Такой симбиоз формальных и неформальных методов, компьютерных технологий позволяет улучшить и ускорить процедуру принятия эффективного решения ЛПР для управления режимами работы ХТС.

Можно предложить следующую структуру ИСКМО для оптимизации режимов работы ректоров риформинга, включающие в себя пакет их взаимосвязанных моделей, разрабатываемые с учетом нечеткости исходной информации, базы знаний, аккумулирующий знания ЛПР, специалистов-экспертов, и других необходимых блоков системы (рисунок 2).

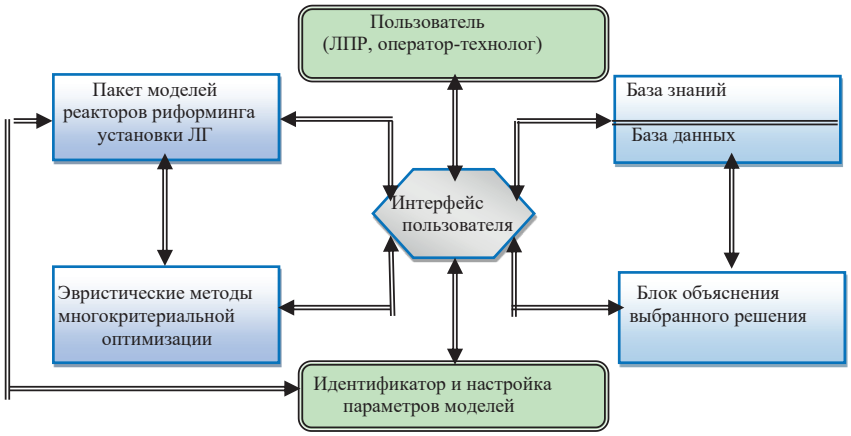


Рисунок 2. Структура ИСКМО для оптимизации режимов работы реакторов риформинга

Приведенные на рисунке 2 основные функциональные блоки ИСКМО связаны через информационные потоки и каждый из них в системе выполняет определенные функции, а все вместе за счет свойства синергизма и эмерджентности системы позволяют эффективно реализовать процесса компьютерного моделирования и оптимизации ХТС.

Пользователем ИСКМО является ЛПР, например, операторы-технологи управляющие реакторами риформинга установки каталитического риформинга. В процессе решения задачи ЛПР осуществляет с помощью системы выбор наилучшего режима работы реакторов, который обеспечивает оптимальные (компромиссные) значения критериев оценки качества работы реакторов. Выбор наилучшего решения

ЛПР реализует с учетом сложившейся производственной ситуаций, производственного плана, требований к объему и качеству производимой продукции, и т. д. Таким образом, решение принимается с учетом важности локальных критериев и ограничений и их изменений, т.е. ЛПР в диалоге с системой решает задачу многокритериальной оптимизации режимов работы реакторов риформинга на основе их моделей.

Пакет моделей реакторов риформинга установки ЛГ-35-11/300-95 объединяет математических моделей взаимосвязанных реакторов риформинга Р-2, Р-3, Р-4 и Р-4а блока риформинга в единый пакет. Эти модели, которые могут быть разработаны на основе различных методов в зависимости от характера исходной информации, объединяются в соответствии со схемой протекания технологического процесса риформинга. Соответственно пакет моделей позволяет системно моделировать работы реакторов риформинга блока риформинга. Кроме того, эти модели используются при расчете значений локальных критериев выбора режима работы объекта в зависимости от изменений входных, режимных параметров.

Эвристические методы многокритериальной оптимизации предназначены для формализации и решения задач многокритериальной оптимизации для управления режимами работы реакторов риформинга с учетом нечеткости исходной информации. Такие эвристические методы решения задач многокритериальной оптимизации разрабатываются на основе модификации различных принципов оптимальности для работы в нечеткой среде (Carreno, 2014), (Orazbayev et all, 2022). Данные методы на основе пакета моделей, базы знаний и данных, и при необходимости других блоков системы позволяют определить и выбрать ЛПР наилучшего режима работы реакторов риформинга установки ЛГ-35-11/300-95 по выбранным критериям с выдачей рекомендуемых значений входных и режимных параметров.

Базы знаний и база данных служить для создания базы знаний и данных и хранения в них формализованных знаний ЛПР, экспертов предметной области и статистических данных о показателях производства. На основе формализованных знаний и данных этого блока проводится анализ работы и основных показателей объекта, а также реализуются процессы подготовки и принятия решений для управления режимами работы реакторов риформинга. Кроме того, данные из этого блока используются при составлении отчетов и адаптации математических моделей.

Интерфейс пользователя предназначен организации удобного

дружественного, диалогового режима работы пользователя с компьютером при вводе и корректировке исходных данных ЛПР, а также при выполнении других функций ИСКМО.

Объяснения выбранного решения. Данный блок реализует функцию объяснения выбранных решений и подсказки по выполнению некоторых действий при эксплуатации ИСКМО.

Идентификатор и настройка параметров моделей представляет собой программу, которая при необходимости на основе алгоритмов параметрической идентификации идентифицирует параметров моделей.

Основной интерфейс созданной ИСКМО, предназначенной для оптимизации режимов работы реакторов риформинга на основе компьютерного моделирования в диалоговом режиме ЛПР-компьютер, представлен на рисунке 3.

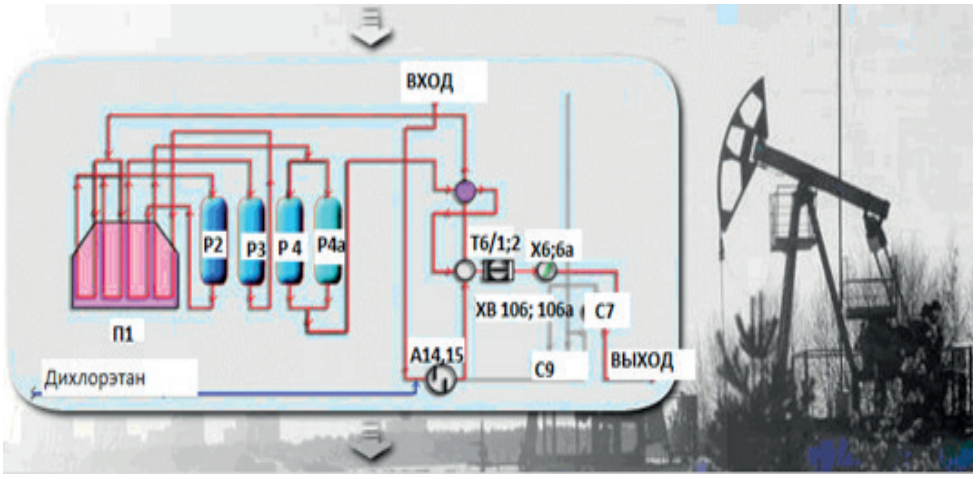


Рисунок 3. Интерфейс ИСКМО для оптимизации режимов работы реакторов риформинга на основе компьютерного моделирования

Как видно, из рисунка 3 в режиме моделирования для удобства в Интерфейсе на верхней части приведены наименование основных режимных-входных параметров (x_1, x_2, x_3, x_4, x_5), меняя которых осуществляется процесс компьютерного моделирования и поиск оптимального режима работы ректоров риформинга. Для удобства пользователю в правой части каждого входного, режимного параметра показаны его интервалы изменения.

Для выбора моделируемого реактора риформинга (Р-2, Р-3, Р-4,4а) в интерфейсе имеется соответствующее окно. Для изменения значения

каждого из параметров x_1, x_2, x_3, x_4, x_5 в правой стороне имеются соответствующие окна.

В нижней части окна приводятся результаты моделирования – значения выходных параметров процесса – y_j, $j=\overline{1,5}$, (количество вырабатываемой продукции и качественные показатели целевой продукции). Для вывода нового значения выходных параметров при изменении входных необходимо нажать на кнопку: стоящей перед соответствующими y_j, $i=\overline{1,5}$).

На рисунке 3 приведены результаты поиска оптимального режима работы реакторов риформинга на основе компьютерного моделирования.

Таким образом, с помощью данной системы, меняя значения входных, режимных параметров ($x_i, i=\overline{1,5}$) и моделируя различные режимы работы реакторов риформинга на компьютере, можно найти оптимальный режим их работы, т.е. найти такие значения $x_i, i=\overline{1,5}$, которые обеспечивают наилучшие значения выходных параметров y_j, $j=\overline{1,5}$.

Следует отметить, что описанный режим оптимизации требует опыт и знаний пользователя, а также времени, т.е. неудобно для производственников. Поэтому для улучшения данной системы в производственных условиях создается подсистема «Система принятия решений» на основе эвристических методов многокритериальной оптимизации для выбора оптимальных режимов работы реакторов риформинга с учетом нечеткости исходной информации, предложенные нами в (Carreno, 2014), (Orazbayev et all, 2022). Такие алгоритмы позволяют пользователю в удобном режиме решать задачи оптимизации, т.е. осуществляет автоматизированный поиск таких значений входных параметров, которые обеспечивают оптимальных значений выходных параметров – критериев.

Обсуждение результатов. Пакет математических моделей реакторов риформинга разработан на основе использование статистических данных и экспертной информации, выраженная нечетко на естественном языке в виде суждения и высказываний экспертов предметной области, лица, принимающего решения при управлении процессом риформинга. Для сбора, обработки и использования нечеткой информации, представляющей собой формализованного интеллекта человека-оператора, применены методы экспертной оценки и математический аппарат теорий нечетких множеств.

Математическая модель, определяющая объем целевой продукции с реакторов риформинга в зависимости от входных, режимных пара-

метров блока риформинга, построена в виде нелинейного уравнения множественной регрессии (1). Такая структура модели идентифицирована на основе метода последовательного включения регрессоров, а ее неизвестные параметры, т.е. регрессионные коэффициенты идентифицированы с помощью метода наименьших квадратов на основе экспериментально-статистических данных.

Так как качество катализата, т.е. его октановое число характеризуется нечеткостью, зависимость октанового числа от входных, режимных параметров x_i, $i = \overline{1,5}$, определена в виде нечеткой регрессионной модели с нечеткими коэффициентами (2). Параметрическая идентификация нечетких регрессионных коэффициентов модели (2) реализована с помощью модифицированного метода наименьших квадратов на трех α–срезах. Затем путем объединения значения нечетких коэффициентов на этих α–срезах получена удобная модель для оценки октанового числа катализата с помощью компьютерного моделирования вида (4).

Структура интеллектуализированной системы компьютерного моделирования для оптимизации режимов работы реакторов риформинга (рисунок 2) создана по принципу открытых систем, т. е. допускает добавлять дополнительные блоки и менять их при необходимости. Представленные основные функциональные блоки созданной системы компьютерного моделирования и оптимизации режимов работы реакторов связаны через информационные потоки и каждый из них в системе выполняет определенные функции. А все эти функциональные блоки вместе за счет свойства синергизма и эмерджентности системы позволяют эффективно реализовать процесса моделирования и оптимизации исследуемых объектов.

Заключение. Исследованы проблемы неопределенности, возникающие случайным и нечетким характерами исходной информации, которые усложняют процессы разработки математических моделей и оптимизации режимов работы сложным ХТС. В качестве таких ХТС рассматривается блок риформинга установки каталитического риформинга ЛГ-35-11/300-95, функционирующего на Атырауском нефтеперерабатывающем заводе, в условиях неопределенности из-за случайного характера и нечеткости исходной информации. На основе комплексного использования доступной информации различного характера предлагаются подходы к их решению. Для эффективной оптимизации создана интеллектуализированная система компьютерного моделирования на основе моделей, построенных с использование интеллекта человека-оператора. Такие модели, по сравнению с тради-

ционными моделями, при наличии компетентных специалистов-экспертов и правильном проведении экспертной оценки и обработки полученной нечеткой информации, являются более адекватными и содержательными, так как учитывают не формализуемые сложные связи между параметрами нечетко описываемых ХТС.

К основным результатам полученные в процессе исследования относятся:

1. Разработанная модель, для определения объема целевой продукции реакторов риформинга имеет структуру множественного регрессионного уравнения, параметры которого идентифицируются с помощью пакета программ, реализующих алгоритм метода наименьших квадратов.

2. Для определения нечетко описываемого качества продукции разработана нечеткая модель, имеющая структуру нечеткого уравнения множественной регрессии, имеющая нечетких коэффициентов регрессии. В этой модели нечеткие регрессионные коэффициенты идентифицированы с помощью модифицированного метода наименьших квадратов на основе множества уровня α.

3. Создана структура интеллектуализированной системы оптимизации режимов работы реакторов риформинга на основе компьютерного моделирования, описаны ее основные блоки. Описан основной интерфейс ИСКМО для оптимизации режимов работы реакторов риформинга на основе компьютерного моделирования.

Благодарности. Исследование финансируется Комитетом науки Министерства образования и науки Республики Казахстан (грант № AP08855680 - Интеллектуализированная система поддержки принятия решений для управления режимами работы установки каталитического риформинга).

Information about authors:

Zhumadillayeva A.K. – Candidate of Technical Sciences, Associate Professor, Deputy Dean of the Faculty of Information Technologies, L.N. Gumilyov Eurasian National University, e-mail: ay8222@mail.ru, https://orcid.org/0000-0002-1741-7553;

Kabibullin M.D. – doctoral student of the Department of System Analysis and Control, L.N.Gumilyov Eurasian National University, e-mail: madyar_kabibullin@mail.ru, https://orcid.org/0000-0002-3972-7894;

Orazbayev B.B. – doctor of technical sciences, academician of the

Engineering academy of the republic of Kazakhstan, professor of the department of System analysis and Control, L.N. Gumilyov Eurasian National University, e-mail: batyr_o@mail.ru https://orcid.org/0000-0003-2109-6999;

Orazbayeva K.N. – doctor of technical sciences, professor of the department of management, Esil University, e-mail: kulman_o@mail.ru, https://orcid.org/0000–0002–1741–7553;

Tuleuov Zh.N. – Managing Director for Processes of Oil Refining and Desulfurization of Atyrau Oil Refinery LLP, e-mail: zh.tuleuov@anpz.kz, https://orcid.org/0000-0003-2353-2416.

ЛИТЕРАТУРА

Biegler L.T., Lang Y.D, Lin W.J. (2016) Multi-scale Optimization for Process Systems Engineering, Computers and Chemical Engineering, №10, pp. 17–35.

Carreno J.E. (2014) Multi-objective optimization by using evolutionary algorithms: The p-Optimality Criteria, IEEE Transactions on Evolutionary Computation., vol. 18. N 2, pp. 167–179. (in Eng.).

Chen Y., He L., Li J., Zhang S. (2018) Multi-criteria design of shale-gas-water supply chains and production systems towards optimal life cycle economics and greenhouse gas emissions under uncertainty, Computers & chemical engineering, vol. 109, pp. 216–235 (in Eng.).

Isakov Yu.A. Artificial intelligence (2018) ModernScience.No 6-1. pp. 25-27.

L. Sansyzbay, B. Orazbayev, W. Wojcik (2020) Development and Analysis of Models for Assessing Predicted Mean Vote Using Intelligent Technologies, International Journal of Fuzzy Logic and Intelligent Systems. DOI: 10.5391/IJFIS.2020.20.4.324 (in Eng.).

M. Fayaz, S. Ahmad, I. Ullah, D. Kim. (2018) A Blended Risk Index Modeling and Visualization Based on Hierarchical Fuzzy Logic for Water Supply Pipelines Assessment and Management, Processes, V. 6. № 5. pp. 102–112 (in Eng.).

Motlatsi C. Lehloka, James A. Swart, and Pierre E. Hertzog (2021) Linear Regression Algorithm Results for a PV Dual-Axis Tracking-Type System, International Journal of Electrical and Electronic Engineering & Telecommunications, Vol. 10, No. 2, pp. 139-144. Doi: 10.18178/ijeetc.10.2.139-144.

Orazbayev B.B, Ospanov E.A., Orazbayeva K.N., Kurmangazieva L.T. (2018) Hybrid Method for the Development of Mathematical Models of a Chemical Engineering System in Ambiguous Condition, Mathematical Models and Computer Simulations, V.10, -pp. 748-758 (in Eng.).

Orazbayev B., Kozhakhmetova D., Wójtowicz R., Krawczyk J. (2020) Modeling of a Catalytic Cracking in the Gasolin Production Installation with a Fuzzy Environment, Energies, 4736. DOI: 10.3390/en13184736 (in Eng.).

Orazbayev B., Zhumadillayeva A., Orazbayeva K., Iskakova S., Utenova, B., Gazizov F., Ilyashenko S., Afanaseva O. (2022) The System of Models and Optimization of Operating Modes of a Catalytic Reforming Unit Using Initial Fuzzy Information, Energies, V. 15, Issue 4, -P. 1-26. 1573. Doi: 10.3390/en15041573 (in Eng.).

Reverberi A.P., Kuznetsov N.T., Meshalkin V.P., Salerno M., Fabiano B. (2016) Systematical Analysis of Chemical Methods in Metal Nanoparticles Synthesis, Theor. Found. Chem. Eng, V. 50. № 1, – P. 63–75 (in Eng.).

Sharikov Yu.V., Petrov P.A. (2013) Universal model for catalytic reforming, Chemical and Petroleum Engineering. DOI: 10.1007/s10556-007-0103-z (in Eng.).

Абдулов П.В. (2018) Введение в теорию принятия решений, Москва, Медиа.146 с.

В.Г. Засканов, Д.Ю. Иванов, Г.М. Гришанов (2013) Системы поддержки принятия решений. Самара: СГАКУ им. С.П. Королева (нац. исслед.ун-т).175 с.

Гуцыкова С. (2017) Метод экспертных оценок. Теория и практика, Москва, Когито-Центр. 509 с.

Карманов Ф.И. (2017) Статистические методы обработки экспериментальных данных с использованием пакета MathCad, Москва, Инфра-М. 287 с.

Кашин О.Н. (2011) Оптимизация химико-технологической системы нефтеперерабатывающего завода с использованием энергосберегающих методов. Дисс.кандидата технических наук. Санк-Петербург, 135 с.

Липин А.Г. (2018) Математическое моделирование химико-технологических систем. Иваново. 135 с.

Макаров И.М., Лохин В.М., Манько С.В., Романов М.П. (2012) Искусственный интеллект и интеллектуальные системы управления, Москва, Наука. 336 с.

Муленко В.В (2015) Компьютерные технологии и автоматизированные системы. Москва, РГУ нефти и газа им. И.М. Губкина. 73 с.

Оразбаев Б.Б., Жумадиллаева А.К., Дюсекеев К.А., Сантеева С.А., Xiao-Guang Yue (2021) Разработка математических моделей реакторов риформинга бензина установки ЛГ-35-11/300-95 на основе системного подхода, Известия национальной академии наук РК, Series physic-mathematical, 5 (339). с.145-152.

Попов А.Л (2019) Проектирование и создание систем поддержки принятия решений, Екатеринбург, Урал. гос. ун-т. 137 с.

Прокопюк С.Г., Масгутов Р.Н. (2018) Промышленные установки каталитического крекинга, Москва. 278 с.

Рыжов А.П. (2017) Теория нечетких множеств и ее приложений: монография, Москва, МГУ. 115 с.

Чернов В.Г. (2019) Модель поддержки принятия решений при планировании проекта внедрения КИС на основе нечетких множеств, Москва, Наука. 347с.

REFERENCES

Biegler L.T., Lang Y.D., Lin W.J. (2016) Multi-scale Optimization for Process Systems Engineering, Computers and Chemical Engineering, №10, pp. 17–35.

Carreno J.E. (2014) Multi-objective optimization by using evolutionary algorithms: The p-Optimality Criteria, IEEE Transactions on Evolutionary Computation., vol. 18. N 2, pp. 167–179. (in Eng.).

Chen Y., He L., Li J., Zhang S. (2018) Multi-criteria design of shale-gas-water supply chains and production systems towards optimal life cycle economics and greenhouse gas emissions under uncertainty, Computers & chemical engineering, vol. 109, pp. 216–235 (in Eng.).

Isakov Yu.A. Artificial intelligence (2018) ModernScience.No 6-1. pp. 25-27.

L. Sansyzbay, B. Orazbayev, W. Wojcik (2020) Development and Analysis of Models for Assessing Predicted Mean Vote Using Intelligent Technologies, International Journal of Fuzzy Logic and Intelligent Systems. DOI: 10.5391/IJFIS.2020.20.4.324 (in Eng.).

M. Fayaz, S. Ahmad, I. Ullah, D. Kim. (2018) A Blended Risk Index Modeling and

Visualization Based on Hierarchical Fuzzy Logic for Water Supply Pipelines Assessment and Management, Processes, V. 6. № 5. pp. 102–112 (in Eng.).

Motlatsi C. Lehloka, James A. Swart, and Pierre E. Hertzog (2021) Linear Regression Algorithm Results for a PV Dual-Axis Tracking-Type System, International Journal of Electrical and Electronic Engineering & Telecommunications, Vol. 10, No. 2, pp. 139-144. Doi: 10.18178/ijeetc.10.2.139-144.

Orazbayev B.B, Ospanov E.A., Orazbayeva K.N., Kurmangazieva L.T. (2018) Hybrid Method for the Development of Mathematical Models of a Chemical Engineering System in Ambiguous Condition, Mathematical Models and Computer Simulations, V.10, -pp. 748-758 (in Eng.).

Orazbayev B., Kozhakhmetova D., Wójtowicz R., Krawczyk J. (2020) Modeling of a Catalytic Cracking in the Gasolin Production Installation with a Fuzzy Environment, Energies, 4736. DOI: 10.3390/en13184736 (in Eng.).

Orazbayev B., Zhumadillayeva A., Orazbayeva K., Iskakova S., Utenova, B., Gazizov F., Ilyashenko S., Afanaseva O. (2022) The System of Models and Optimization of Operating Modes of a Catalytic Reforming Unit Using Initial Fuzzy Information, Energies, V. 15, Issue 4, -P. 1-26. 1573. Doi: 10.3390/en15041573 (in Eng.).

Reverberi A.P., Kuznetsov N.T., Meshalkin V.P., Salerno M., Fabiano B. (2016) Systematical Analysis of Chemical Methods in Metal Nanoparticles Synthesis, Theor. Found. Chem. Eng, V. 50. № 1, – P. 63–75 (in Eng.).

Sharikov Yu.V., Petrov P.A. (2013) Universal model for catalytic reforming, Chemical and Petroleum Engineering. DOI: 10.1007/s10556-007-0103-z (in Eng.).

Abdulov P.V. (2018) Introduction to the theory of decision-making, Moscow, Media.146 p.

V.G. Zaskanov, D.Y. Ivanov, G.M. Grishanov (2013) Decision support systems. Samara: SSAKU named after S.P. Korolev (National Researchun-t).175 p.

Gutsykova S. (2017) The method of expert assessments. Theory and Practice, Moscow, Kogito-Center. 509 p.

Karmanov F.I. (2017) Statistical methods of experimental data processing using the MathCad package, Moscow, Infra-M. 287 p.

Kashin O.N. (2011) Optimization of the chemical-technological system of an oil refinery using energy-saving methods. Diss.candidate of technical Sciences. St. Petersburg, 135 S.

Lipin A.G. (2018) Mathematical modeling of chemical and technological systems. Ivanovo. 135 p.

Makarov I.M., Lokhin V.M., Manko S.V., Romanov M.P. (2012) Artificial Intelligence and Intelligent control systems, Moscow, Nauka. 336 c.

Mulenko V. In (2015) Computer technologies and automated systems. Moscow, Gubkin Russian State University of Oil and Gas. 73 p.

Orazbayev B.B., Zhumadillayeva A.K., Dyusekeev K.A., Santeeva S.A., Xiao-Guang Yue (2021) Development of mathematical models of gasoline reforming reactors of the LG plant-35-11/300-95 based on a systematic approach, Proceedings of the National Academy of Sciences of the Republic of Kazakhstan, Series physical-mathematical, 5 (339). p. 145-152.

Popov A.L. (2019) Design and creation of decision support systems, Yekaterinburg, Ural State University. 137 p.

Prokopyuk S.G., Masgutov R.N. (2018) Industrial catalytic cracking plants, Moscow. 278 p.

Ryzhov A.P. (2017) Theory of fuzzy sets and its applications: monograph, Moscow, MSU. 115 p.

Chernov V.G. (2019) Decision support model for the planning of a CIS implementation project based on fuzzy sets, Moscow, Nauka. 347c.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 91-116
https://doi.org/10.32014/2022.2518-1726.141
УДК 004.056
МРНТИ 81.93.29

Ж.Д. Изтаев[1], Г.Т. Джусупбекова[1], Г.К. Ордабаева[2*]

[1]М. Әуезов атындағы Оңтүстік Қазақстан университеті,
Қазақстан, Шымкент;

[2]Әл-Фараби атындағы Қазақ ұлттық университеті, Қазақстан, Алматы.
E-mail: gulzi200988@mail.ru

УНИВЕРСИТЕТ ҮШІН АҚПАРАТТЫҚ ҚАУІПСІЗДІК ҚАТЕРЛЕРІНІҢ ЖЕКЕ МОДЕЛІН ӘЗІРЛЕУ

Аннотация. Бүгінгі таңда ақпараттық технологиялардың дамуымен есептеу және желілік қосымшалар университет ортасының ажырамас бөлігіне айналды. Қазіргі университеттер технологиялық прогрестің алдыңғы қатарында. Технологияға кең қолжетімділік құнды оқу орта-сына әкеледі, екінші жағынан, қауіпсіздікке қауіп төндіретін есептеу ортасының осалдығына әкелуі мүмкін. Университет кампустары Wi-Fi-ға кең қолдау көрсету, дәріс жазу бағдарламалық жасақтамасын қолдана отырып онлайн оқыту, сандық кітапхана, сыныптағы виртуализация, веб-конференциялар және т.б. сияқты мүмкіндіктерді ұсына отырып, әлемдегі технологиялық дамыған орындардың бірі ретінде өзін дәлел-деді.

Ашық үлкен кампусты үнемі өзгеріп отыратын қауіптер мен осал-дықтардан қорғау өзекті мәселенің бірі. Университеттің ашық есептеу ортасының қолданушылары - студенттер, оқытушылар және әкімшілік. Университет кампусының желісі пайдаланушыларға қауіпсіз қол жетімділікті қамтамасыз етіп қана қоймай, оларды осалдықтардан да қорғауы керек. Әр қолданушы университеттік ресурстардың әр түрлі деңгейімен университеттік есептеу ортасына қол жеткізе алады. Кампустар желісінде тәуекел дәрежесін және қауіпсіздік тиімділігін арттыру қажет. Бұл өте маңызды қауіптерді анықтауды, кампус

желісін үздіксіз желілік бақылау арқылы тәуекел деңгейін өлшеу үшін осалдықтарды бағалауды талап етеді.

Мақалада әл-Фараби атындағы Қазақ ұлттық университетінің (ҚазҰУ) кампус желісінде болатын қауіпсіздік қатерлерін ескере отырып, университеттің есептеу ортасы үшін арнайы жасалған ақпараттық қауіпсіздік тәуекелдерін сандық бағалау моделі ұсынылған. Ұсынылған модель университет желісінің конфигурациясындағы ықтимал қауіптер мен ақпараттық процестерді анықтау арқылы қауіпсіздік тәуекелдерін сандық түрде өлшейді. Бұл модельді тәуекелдерді талдаушы және университеттің қауіпсіздік менеджері нақты және қол жетімді түрде сенімді және қайталанатын тәуекелдерді талдауды жүзеге асыру үшін қолдана алады.

Түйін сөздер: ақпараттық жүйелер, ақпараттық қауіпсіздік, модель, Nmap, Metasploit, Acunetix, мобильді қосымша.

Ж.Д. Изтаев[1], Г.Т. Джусупбекова[1], Г.К. Ордабаева[2*]

[1]Южно-Казахстанский университет им. М. Ауезова,
Казахстан, Шымкент;
[2]Казахский национальный университет имени аль-Фараби,
Казахстан, Алматы.
E-mail: gulzi200988@mail.ru

РАЗРАБОТКА ЧАСТНОЙ МОДЕЛИ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ДЛЯ УНИВЕРСИТЕТА

Аннотация. Сегодня с развитием информационных технологий вычислительные и сетевые приложения стали неотъемлемой частью университетской среды. Современные университеты лидируют в технологическом прогрессе. Широкий доступ к технологии приводит к ценным средам обучения, с другой стороны, может привести к уязвимости вычислительной среды, угрожающей безопасности. Кампусы университета зарекомендовали себя как одно из технологически развитых мест в мире, предлагая такие возможности, как широкая поддержка Wi-Fi, онлайн-обучение с использованием лекционного программного обеспечения, цифровая библиотека, виртуализация в классе, веб-конференции и т.д.

Защита открытого большого кампуса от постоянно меняющихся угроз и уязвимостей является одной из актуальных проблем. Пользователи открытой вычислительной среды университета – студенты, преподаватели и администрация. Сеть кампуса университета должна обеспечивать не только безопасный доступ к пользователям, но и защищать их от уязвимостей. Каждый пользователь имеет доступ к университетской вычислительной среде с разным уровнем университетских ресурсов. Необходимо повысить степень риска и эффективность безопасности в сети кампусов. Это требует выявления очень важных рисков, оценки уязвимостей для измерения уровня риска путем непрерывного сетевого контроля сети кампусов.

В статье представлена модель количественной оценки рисков информационной безопасности, разработанная специально для вычислительной среды университета с учетом рисков безопасности в сети кампусов Казахского национального университета имени аль-Фараби (КазНУ). Предложенная модель количественно измеряет риски безопасности путем выявления возможных рисков и информационных процессов в конфигурации сети университета. Данная модель может применяться аналитиком рисков и менеджером по безопасности университета для осуществления анализа надежных и повторяющихся рисков в реальном и доступном виде.

Ключевые слова: информационные системы, информационная безопасность, модель, Nmap, Metasploit, Acunetix, мобильное приложение.

Zh.D. Iztayev[1], G.T. Dzhusupbekova[1], G.K. Ordabaeva[2*]

[1]South Kazakhstan University named after M. Auezov,
Kazakhstan, Shymkent;
[2]Al-Farabi Kazakh National University, Kazakhstan, Almaty.
E-mail: gulzi200988@mail.ru

DEVELOPMENT OF A PRIVATE MODEL OF INFORMATION SECURITY THREATS FOR THE UNIVERSITY

Abstract. Today, with the development of information technology, computing and networking applications have become an integral part of the university environment. Modern universities are leading the way in technological progress. Broad access to technology leads to valuable

learning environments, on the other hand, can lead to a security-threatening computing environment vulnerability. The university's campuses have established themselves as one of the technologically advanced places in the world, offering opportunities such as extensive Wi-Fi support, online learning using lecture software, digital library, classroom virtualization, web conferences, etc.

Protecting an open, large campus from ever-changing threats and vulnerabilities is one of the pressing challenges. Users of the open computing environment of the university are students, teachers and administration. The university's campus network should provide not only secure access to users, but also protect them from vulnerabilities. Each user has access to a university computing environment with a different level of university resources. It is necessary to increase the degree of risk and security efficiency in the network of campuses. This requires identifying very important risks, assessing vulnerabilities to measure the level of risk by continuously monitoring the network of campuses.

The article presents a model for quantitative assessment of information security risks, developed specifically for the computing environment of the university, taking into account security risks in the network of campuses of the Al-Farabi Kazakh National University (KazNU). The proposed model quantifies security risks by identifying possible risks and information processes in the configuration of the university network. This model can be used by a Risk Analyst and a University Security Manager to perform a real and available analysis of reliable and recurring risks.

Key words: information systems, information security, model, Nmap, Metasploit, Acunetix, mobile application.

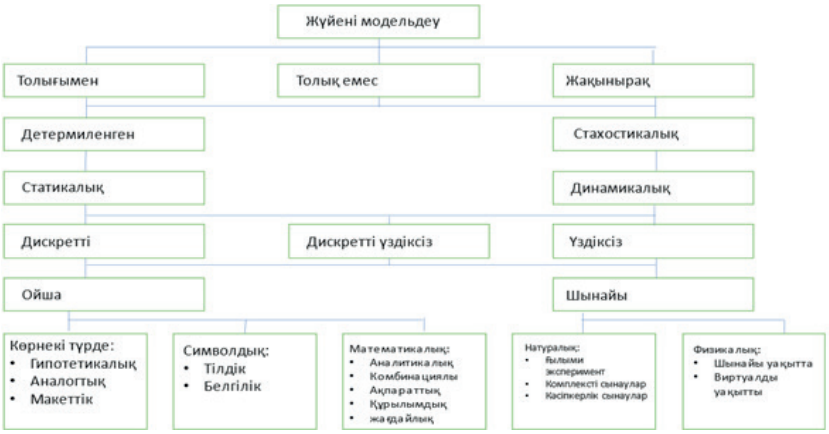
Кіріспе. Қазіргі заманғы типтік ақпараттық жүйелер (АЖ) – бұл клиент-серверлік аумақтық бөлінген көп қолданушы архитектурасы. Бағдарламалық жасақтама (әдетте қолданбалы бағдарламалық жасақтама) жалпы бағдарламалау тілдерін (мысалы, C#, TypeScript (бұрыштық даму платформасы, dotnet әзірлеу ортасы) қолдана отырып, ашық бағдарламалық интерфейске (API) негізделген функционалды модификацияларды ұсынады.

Жүйелерді модельдеу барлық ғылым салаларында басты зерттеу әдістерінің бірі және күрделі жүйелерді бағалаудың бірден-бір негізгі ғылыми әдісі. АЖ моделдеу - АЖ моделінің көмегімен түпнұсқа АЖ-нің маңызды қасиеттері туралы ақпарат алу мақсатында, нақты АЖ-ні басқасымен алмастыру түсініледі.

АЖ деректерді орталықтандырылған сақтау, жинақтау және бірнеше рет пайдалану қағидатын қамтамасыз етеді. Пайдаланушылардың автоматтандырылған жұмыс орындарында (АЖО) ресурстарды үнемдеу және ақпараттық қауіпсіздікті қамтамасыз ету үшін деректерді сақтау, өңдеу сервер көмегімен жүзеге асырылады.

АЖ жергілікті есептеу желісінен тыс орналасқан және АЖ серверлерімен ақпарат алмасуды жүзеге асыратын мобильді АЖО-да ақпаратты криптографиялық қорғау құралдары қолданылады.

Зерттеу әдістемесі мен материалдары. Жүйелерді модельдеу ұқсастық теориясына негізделген, оның негізгі мәні абсолютті ұқсастық тек бір объектіні басқасымен алмастыра алатын кезде ғана болуы мүмкін (1-сурет).



1-сурет. Жүйелерді модельдеу түрлерінің жіктелуі

Детерминленген модельдеу - кездейсоқ әсерлері жоқ процестерді көрсетеді. Стахостикалық модельдеу - ықтималды процестер және оқиғаларды көрсетеді. Статикалық модельдеу - АЖ-ның күйін кезкелген уақыт аралығында не болғанын сипаттайды. Динамикалық модельдеу - қазіргі уақыттағы АЖ-ның күйін көрсетеді. Дискретті модельдеу - АЖ-дағы дискретті, ал үздіксіз модельдеу - үздіксіз процестерді сипаттайды. Дискретті үздіксіз модельдеу - АЖ-дағы дискретті және үздіксіз процестерді сипаттауда қолданылады. Ойша модельдеу - шынайы объекттерді модельдегенде қолданылады, егер олар белгіленген уақыт аралығында қолжетімсіз болса немесе қолайсыз жағдайлар туындаса. Көрнекі модельдеу кезінде - АЖ-да болып жатқан құбылыстар мен процестерді бейнелейтін АЖ-ның көрнекі модельдері қалыптасады. Гипоткалық модельдеуде - нақты

АЖ-дағы процестердің заңдылықтары туралы гипотеза жасалады, ол сарапшының АЖ туралы білім деңгейін көрсетеді және зерттелетін АЖ-ның кірісі мен шығысының арасындағы себептік қатынастарға негізделген. Мұндай модельдеу АЖ құру үшін ақпараттар жеткіліксіз болған кезде қолданылады. Аналогтық модельдеу - әр деңгейдегі анаогтарды қолданады. Ең үздік деңгей зерттеліп отырған АЖ-ның толық аналогы болып табылады.

Математикалық модельдеу - нағыз АЖ модельіне қарап жасалынған математикалық модельдеу процесі. Аналитикалық модельдеу кезінде АЖ элементтерінің жұмыс істеу процестері кейбір функционалдық қатынастар (алгебралық, интегродифференциалды, айырмашылық және т.б.) немесе логикалық шарттар түрінде жазылады. Біріктірілген (аналитика - имитациялық) АЖ модельдеу екі модельдеудің жақсы жақтарын алады. Ішкі процестерге аналитикалық модель сәйкес келсе сол қолданылады, ал басқа кезде имитациялық модель қолданылады.

Физикалық модельдеу кезінде АЖ құбылыстардың табиғатын сақтайтын және физикалық ұқсастыққа ие жұмыс ортасында жүзеге асырылады. Физикалық модельдеу нақты және нақты емес уақыт шкаласында, сондай–ақ уақытты ескерусіз жүруі мүмкін.

Модельдерді құру әдістерінің жіктелуі 2 - суретте көрсетілген.

2 - сурет. Модельдерді құру әдістерінің жіктелуі

Бұл модельдердің негізгі кемшіліктері:

- модельдеу кезінде жаңа сапалық сипаттамаларды анықтау әрдайым мүмкін емес;

- кез-келген модель мүмкін құбылыстардың түсіндірілуін азайтады;
- статистикалық модельдер модель құрудың эмпирикалық жиынтығы аясында ғана объективті бола алады.

Мемлекеттік жалпыодақтық стандартта (государственный общесоюзный стандарт, ГОСТ, СТ РК ГОСТ Р ИСО/МЭК 15408-1-2006:52) ұғымдарды пайдалану мәнмәтінін және оларды қолдану тәсілін қоса алғанда, жалпы критерийлердің (ЖК) барлық бөліктерінде қолданылатын жалпы ұғымдар берілген. Қауіпсіздік активтерді қауіп-қатерден қорғаумен байланысты, онда қауіптер қорғалатын активтерді теріс пайдалану әлеуеті негізінде жіктеледі. Қауіптердің барлық түрлерін ескере отырып қауіпсіздік саласында адамның іс-әрекетіне, зиянды немесе басқа да байланысты мәселелерге көңіл бөлінеді.

Сонымен қатар, (СТ РК 1698-2007:33) «компьютерлік шабуыл - ақпаратқа, автоматтандырылған ақпараттық жүйенің ресурсына немесе бағдарламалық немесе бағдарламалық-аппараттық құралдарды қолдана отырып, оларға рұқсатсыз қол жеткізуге бағытталған рұқсатсыз әсер ету» деген анықтама берілген.

Қазіргі уақытта шабуылдардың көптеген модельдері, шабуылдарды модельдеу әдістері мен құралдары бар. Ақпараттық жүйелерге шабуылдардың негізгі модельдері 3 - суретте көрсетілген.

3 - сурет. Ақпараттық жүйелерге шабуыл модельдері

Шабуылдардың негізгі модельдерінің артықшылықтары мен кемшіліктері 1-кестеде келтірілген.

1-кесте. Шабуыл модельдерінің артықшылықтары мен кемшіліктері

№ р/н	Модель	Артықшылықтары	Кемшіліктері
1.	Кестелік (матрицалық)	Ең қарапайым	Циклдік шабуылдарды, оқыс оқиғалар немесе құқық бұзушының әрекеттері арасындағы көптеген байланыстарды модельдеу қиын.
2.	Логикалық	Инциденттерді өңдеу және пәндік сала туралы білімді ұсыну тілдерін қолдану. Модельдеу шабуылдары туралы белгісіздік жағдайларын ескереді.	Логикалық шығару механизмдерін қамтамасыз ететін қашықтықтан банктік қызмет көрсету (ҚБҚ) пайдалану;
3.	Графикалық	"Инциденттерді талдау, шабуылдарды анықтау, АҚ қамтамасыз ету үшін тәуекелдер мен ресурстарды барынша азайту" сияқты көптеген міндеттерді шешуге арналған	Элементтердің көптігі бар ақпаратты бағалау үшін графиктің пайда болуымен байланысты масштабталу.
4.	Шабуыл ағаштарындағы графтар	Көрнекілік, масштабталу, бейімделу, әмбебаптылық	Циклдік шабуылдарды модельдеу қиын; динамикалық модельдеудің болмауы
5.	Байес графтары	Көрнекілік, масштабталу, бейімделу, әмбебаптылық. шабуылдар туралы белгісіздік жағдайларын ескереді	Циклдік шабуылдарды модельдеу қиын; динамикалық модельдеудің болмауы
6.	Петри Желісі	Динамикалық және параллель процестерді модельдеудің ыңғайлылығы ықтималдық процестерді, уақыт параметрлерін қолдануды, зерттеу мен қолданудың қарапайымдылығын, көптеген көрсете алады.	Құқық бұзушының мінез-құлқы мен шабуылдың мақсаттарын сипаттай алмау
7.	Имитациялық	Құқық бұзушының мінез-құлық сипаттамаларын және шабуыл мақсаттарын модельдеуге мүмкіндік береді. Таратылған шабуылдарды модельдеу үшін ыңғайлы, көптеген құралдар бар және олар кең таралған	Үлкен есептеу ресурстарын қажет етеді

Шабуыл модельдерінің бірқатар кемшіліктері бар, атап айтқанда:

- модельдеудің күрделілігі;
- есептеу ресурстарын қажет етеді;
- ақпараттық қауіпсіздік (АҚ) саласындағы жоғары білікті мамандарды тартуды талап етеді;
- сараптамалық әдістердің қателіктері (сараптамалық бағалау).

Жүргізілген талдау негізінде қолданыстағы кемшіліктерді болдырмайтын өзекті ақпараттық қауіпсіздікті қорғауды (АҚҚ) анықтаудың жаңа әдістемелерін жетілдіру және әзірлеу қажеттілігі туралы қорытынды жасауға болады.

Khando (Khando et.al., 2021:22) ақпараттық қауіпсіздік саласындағы қызметкерлердің хабардар болуы (information security awareness, ISA) бойынша әдебиеттерге жүйелі шолу жасаған және жеке меншік ұйымдар мен мемлекеттік секторда қызметкерлерде ISA-ны арттыру үшін ISA әдістері мен факторларының заманауи жинағын ұсынады. Бұл зерттеу ISA контентін әзірлеу әдістері мен факторларындағы соңғы үрдістер туралы біраз түсінік береді, сондай-ақ ISA-ны өз ұйымдарында дамытудың жан-жақты бағдарламасын әзірлеуге көмектесу үшін ақпараттық қауіпсіздік жөніндегі мамандар арасында ақпарат пен білімді тарату жолымен ISA озық тәжірибесін енгізуге ықпал етеді.

У.А. Төкеев (Төкеев т.б., 2011:161) ақпараттық қауіпсіздік қатерін бағалау мен анализдеу әдістерін дамытуда айтарлықтай үлес қосқан отандық ғалымдар. Берілген оқу құралы ақпараттар қауіпсіздігін басқару жағынан мемлекеттік тілде жазылған алғашқы оқу-әдістемелік нұсқаулық. Оқу құралы 7 тараудан және біренше қосымшалардан тұрады. Әр тараудан соң шағын жаттығулар мен есептер берілген.

Ахметов Б.С. (Ахметов т.б., 2015:8) зерттеулерінде ұлттық және халықаралық құжаттарда ашылып отырған қауіп түсінігінің анализіне талдау жасалған. Ақпараттық қауіпсіздік саласында көптеген түсініктер ішінен қауіптің базалық сипаттамасы анықталған. Сонымен бірге, ақпараттық қауіпсіздік саласындағы бейнеде қауіптің базалық сипаттамасын п-компоненттік кортежді моделі түрінде таныстыру ұсынылған. Мақалада ақпараттық қауіпсіздік саласындағы кезекті талдау жасау үшін қауіп түсінігі ашылған және оның базалық сипаттамасы анықталған. Бұл ақпаратты қорғау тапсырмаларын тиімді шешуді жоғарылату мүмкіндігін кеңейтеді.

М.Н. Калимолдаев (Калимолдаев т.б., 2017:7) математикалық логиканы пайдалана отырып ақпараттық ресурстарға қолданушылардың қол жетімділік құқықтары мен мүмкіндіктерінің моделін қарастырған.

Логикалық жүйе түрінде қолжетімділікті типтелген атрибуттық шектеу моделін ұсыну субъектілердің объектілерге қауіпсіз қол жеткізуін қамтамасыз ете отырып, мәндерге артықшылықтар берудің дұрыстығын және жүйе жұмысының дұрыстығын формальды түрде дәлелдеуге мүмкіндік береді. Сонымен қатар, логика түріндегі қолжетімділікті типтелген атрибуттық ажырату моделі логикалық және функционалдық бағдарламалау тілдерінде тікелей іске асырылады.

Akhmetov B. зерттеулерінде (Akhmetov et.all., 2015:9) ірі оқу орындарының, атап айтқанда, Иордания, Қазақстан және Украина университеттерінің киберқауіпсіздік жүйелерінде өзара инвестициялық бақылау стратегияларының ұтымды нұсқаларын іздеу моделі қарастырылған. Мақала киберқауіпсіздік жүйелерінің инвестициялық параметрлері мен біздің мемлекеттің ірі білім беру мекемелерінің ақпараттық-білім беру ортасын қорғауға байланысты басқа да міндеттерді қамтамасыз ету арасындағы әртүрлі өзара байланыстарға арналған. Ғылыми зерттеулердің осы сегментегі басқа авторлардың жұмыстарына қарағанда өзгешілігі, өзара инвестициялар процесіндегі нақты параметрлер мен ұсынымдарды айқындау қабілеттігі. Сондықтан Иордания, Қазақстан және Украина жоғары оқу орындарының үлгісінде ақпараттық-білім беру платформаларында және оқу орындарының киберқауіпсіздігіндегі басқарушылық шешімдерді оңтайландыру үшін алғышарттар жасалған.

Khairur Razikin (Khairur Razikin et.al., 2022:22) ISO/IEC 27001 киберқауіпсіздік құрылымы негізінде ақпараттық технологиялар қауіпсіздігі жүйесін құру кезінде киберқауіпсіздік саласында шешімдер қабылдауды әзірлеуге арналған модель зерттелген. Ұсынылып отырған модель қауіп-қатерлерді жеңілдету үшін ең үздік қауіпсіздік жүйесін алуға бағытталған. Бұл құжат ақпараттық технологиялар қауіпсіздігі жүйелерін әзірлеу кезінде ең жақсы қадамдарды айқындау үшін киберқауіпсіздік саласында шешімдер қабылдауды қолдау жөніндегі ұсынымдарды әзірлеуде стратегиялық директивті органдарға ықпал етті.

Лахно В. (Лахно и др., 2021:11) өз зерттеулерінде кибершабуылды анықтау кезінде ақпараттық-коммуникациялық желілеріне кіру қаупі мен кезеңдерін болжау барысында шешімдер қабылдауды қолдау жүйесінің есептеу ядросы үшін байесовтік желілердің (БС) үлгілерін әзірлеген. Ұсынылған БС үлгілері көптеген кездейсоқ айнымалыларды операцияға және берілген шарттар кезінде кибернетикалық қатерді немесе кіру нақты кезеңін іске асыру ықтималдығын анықтауға

мүмкіндік береді. Байестің динамикалық желілерін (DSB) қолдану негізінде желілік басып кірулерді анықтаудың ықтимал модельдерімен толықтырылған. Әзірленген модельдердің тиімділігі бұрын оқытуда қолданылмаған тестілік негізде тексерілген.

Барабанов А.В. (Барабанов, и др., 2017:224) оқу құралында ұйымдарда ақпаратты көп деңгейлі қорғау кезінде қолданылатын базалық қағидаттар, тұжырымдамалық тәсілдер мен ақпараттық технологиялар баяндалған. Компьютерлік жүйелердің ақпараттық ресурстарының қауіпсіздігін қамтамасыз ету әдістері, желілік қауіпсіздік, техникалар мен құралдарды құрылымдау және жіктеу жолдары терең баяндалған.

Лившиц И.И. (Лившиц и др., 2018:407) ғылыми жұмысында қазіргі заманғы кешенді тәуекел-бағдарланған тәсілге, ақпараттық қауіпсіздік менеджменті жүйесінде ақпараттық қауіпсіздік аудитін (АҚА) орындаудың арнайы модельдері мен әдістеріне негізделген өнеркәсіптік объектілер үшін ақпараттық қауіпсіздік аудитін қамтамасыз етуге арналған ғылыми-әдістемелік аппарат қалыптастырылған. Модельдер мен әдістер кешенінің жаңалығы АҚА орындау үшін функционалдық аяқталған құрылымды қалыптастыру болып табылады.

Jaafer Al-Saraireh (Jaafer Al-Saraireh, et al., 2022:11) зерттеу жұмысында кеңейтілген тұрақты қатер (Advanced Persistent Threat, APT) шабуылдарының деректер жиынтығын пайдалана отырып, APT шабуылдарын анықтау тәсілі ұсынылған. Деректер жиыны түрлі шабуылдар түрлерінің негізінде APT шабуылдарын анықтау үшін ұсынылған машиналық оқыту (ML) моделіне орналастырылды. Деректердің бес түрі жиналды, атап айтқанда, қалыпты, рекогносцирлік, бастапқы ымыраға келу, бүйірлік қозғалыс және деректерді сүзгілеу. Деректердің әрбір түрі қаскүнем өтуі мүмкін кезеңді көрсетеді. Көрнекті өнімділікке деректер жиынтығындағы 65 функцияның тек 12 функциясын пайдалана отырып, 99,89% дәлдікпен қол жеткізілді.

Зерттеу нәтижелері. Бағдарламалық жасақтаманы автоматтандырылған өңдеу және әзірлеу үшін мәліметтер жиынтығын ұсыну мақсатында Python 3 бағдарламалау тілі және Data Science технологиясы қолданылды.

Деректерді автоматтандырылған өңдеуді дайындау үшін деректерді түрлендіру қажет. Деректерді түрлендіруге арналған Python 3 бағдарламалау тіліндегі код 4-суретте келтірілген:

```
train = pd.read_csv('threats.csv', encoding='utf-8')
df = pd.get_dummies(train)
# Жол деректерін түрлендіру
for col in list(df.columns):
if ('ft²' in col or 'kBtu' in col or 'Metric Tons CO2e' in col or 'kWh' in
col or 'therms' in col or 'gal' in col or 'Score' in col):
# Конвертация
df[col] = df[col].astype(float)
```

4-сурет. Python 3 бағдарламалау тілінде деректерді түрлендіру

Тәуекелдерді бағалаудың көптеген модельдері бар, алайда, ұйымдарға ұйым ішінде қай модельді қолдануға болатындығын анықтауға көмектесетін механизм жоқ. Сонымен қатар, бұл модельдер банктер сияқты мақсатты ұйымдарды бұзу кезінде анықталған қауіпсіздік мәселелерін ескереді. Қауіпсіздік тәуекелдерін бағалау бұл ұйымдар үшін өте маңызды болғанымен, біріншіден, ұйымдар қауіпсіз және жабық желілік ортаға ие. Екіншіден, ақпараттық қауіпсіздік тәуекелдерін бағалау негізгі және басым міндет болып табылатын университеттер сияқты жоғары оқу орындарында үлкен және ашық есептеу ортасы бар. Университеттің үлкен есептеу ортасына әртүрлі желілік құрылғылар, бағдарламалық қосымшалар және көптеген серверлер кіреді. Ұсынылған модельдің маңыздылығы мен тиімділігін бағалау үшін ҚазҰУ -нің есептеу ортасы қолданылды (5-сурет).

5-сурет. ҚазҰУ университетінің есептеу ортасы үшін желіні орнату

Ақпараттық қауіпсіздік тәуекелдерін басқарудың ұсынылып отырған құрылымы тәуекелдерді басқарудың үздіксіз процесін айқындайды, әр түрлі әрекеттер тізбегінен тұрады. Ұсынылған модельдің бірін-

ші кезеңінде ұйымның ақпараттық активтерінің тәуекелдеріне толық талдау жасалады және нәтижелер негізінде келесі әрекеттер анықталады: бірінші кезеңдегі іс-әрекеттің мақсаты - осалдық сканері немесе ену тесті сияқты әртүрлі көздерден университеттің есептеу ортасында көрінетін және қолдануға болатын әлсіздіктер мен осалдықтарды анықтау. Бірінші кезеңнің нәтижелері екінші кезеңдегі тәуекелдерді талдауда қолданылады, келесі қадамда барлық активтер үшін тәуекелдерді бағалау жүргізіледі.

Бірінші кезеңде кампус желісінің қауіпсіздігін қамтамасыз етуге ұсынылған тәсіл ақпаратты маңызды активтердің бірі ретінде анықталады. Университеттің есептеу ортасы үшін қауіпсіз инфрақұрылымды дамыту қауіпсіздік мамандары байланыс инфрақұрылымының бөлігі бола алатын техникалық активтерге (желіге қосылу кабельдері, маршрутизаторлар мен коммутаторлар) немесе құрылғы инфрақұрылымына (физикалық немесе виртуалды хосттар) назар аударады және бағдарламалық жасақтама болуы мүмкін.

Университеттің есептеу жүйесінің сипаттамасы есептеу жүйесінің шектеулерін, сондай-ақ желілік ортаны құрайтын ресурстар мен ақпаратты анықтау арқылы тәуекелді бағалауға күш салады. ҚазҰУ университеті кампусының үлкен және ашық желілік ортасы негізінен келесі қауіпсіздік қатерлеріне ұшырайды - Фишинг, Ransomware және зиянды бағдарламалар. Киберқылмыскерлер қаржылық пайда алу үшін ресми хабарламаларды бұрмалайтын электрондық пошталарды немесе веб-шоттарды пайдаланады. Университеттің жас студенттері көбінесе фишингтік шабуылдардың құрбаны болады, нәтижесінде зиянды бағдарламалар немесе бопсалау бағдарламалары жүктеледі.

ҚазҰУ қалашығының аумағында Wi-Fi-ға қол жетімділік қамтамасыз етіледі, бұл техникалық прогресс тұрғысынан өте жақсы, бірақ күтпеген жерден қауіпсіздік мәселелерін тудыруы мүмкін.

Университет жастары Facebook, Telegram және YouTube сияқты әлеуметтік желілердің ең белсенді қолданушылары болып табылады. Бұл университет желісінде зиянды бағдарлама әлеуметтік медиа көмегімен таралуы мүмкін дегенді білдіреді.

Көптеген мобильді құрылғылар негізінде қауіп-қатер де көп. Студенттер технологияны бірінші болып игереді және кампуста жаңа құрылғылар жиі пайда болады - iPad–тан бастап жаңа Android телефондары.

Осалдықтарды анықтау және бағалау мақсатында келесі желіні сканерлеу қосымшалары қолданылды: Nmap, Nexpose және Acunetix.

Сагар Рахалкар Пуна (Сагар Рахалкар Пуна, 2019:144) ақпараттық қауіпсіздіктің саласындағы 11 жылдық тәжірибесі бар маман. Ол киберқылмыстарды тергеуге, цифрлық криминалистикаға, қосымшалардың қауіпсіздігіне, осалдықты бағалауға және енуге тестілеуге, мандаттар мен нормативтік актілерді сақтауға маманданған. Бұл оқулықта NMAP, OpenVAS және Metasploit сияқты үш құралдың негіздерімен танысасыз және осы үш құралды пайдалана отырып, ену үшін тестілеудің кең мүмкіндіктеріне ие боласыз.

Ордабаева Г.К. (Ордабаева, 2020:6) зерттеуінде OSI моделінің физикалық, арналық және желілік деңгейіндегі таратылған есептеу жүйесі объектілерінің өзара әрекеттесуін сипаттайтын бағытталған графты қолдана отырып жасалған ақпараттық жүйенің моделі берілген. Қолданылуы қарапайым Nessus Vulnerability Scanner негізінде желінің осалдығын анықтау әдісі келтірілді. Nessus Vulnerability Scanner осалдықты анықтаудағы үздік сканер болып табылады. Басты артықшылығы – деректер базасындағы қауіпті моделдер негізінде желіні тез әрі сапалы тексереді.

Eric Filiol (Eric Filiol, et al., 2021:8) зерттеуінде Red Hat хакерлері орындайтын осалдықты анықтау және жою процесін автоматтандыруға бағытталған әдіс ұсынылған. Ұсынылған әдіс Metaesploitable Linux дистрибутивін пайдалана отырып бағаланған. Нәтижеде кең таралған алты сервиске назар аударылған: ftp, ssh, telnet, rdp, StartViewer және Printer. Ұсынылған әдіс кең таралған осалдықтарды автоматты түрде жоюға қабілетті екені көрсетілген.

Университет ортасында қауіпсіздіктің ықтимал қауіптерін анықтағаннан кейін, келесі қадам–сәтті шабуыл нәтижесінде шығындардың ықтимал әсерін анықтайтын осалдықты бағалау. Осалдықтарды сканерлеу әкімшіге желідегі қауіпсіздіктің нақты жағдайы туралы хабарлайды және шабуылдаушы кез–келген осалдықты бірінші болып тапқанға дейін түзетуді анықтауға көмектеседі. Университет желісі үлкен және ашық, сондықтан бүкіл желіні сканерлеудің орнына біз хосттарды топтарға жіктеп, әр топты сканерлейміз.

Сканерлеу мақсаты – жүйенің конфигурациялары туралы егжей-тегжейлі түсінік алу үшін сыртқы қауіпсіздікке қарсы шараларды болдырмау. Сканерлеу қауіпсіздік жағдайын анықтауға арналған, яғни, егер ҚазҰУ университетінің желісін зерттеуге тырысса, хакердің не көретінін анықтау. ҚазҰУ университетін сканерлеу үшін Nmap, Nexpose, Metasploit және Acunetix сияқты құралдар пайдаланылды. Nmap және Nexpose құралы осалдықтар үшін сканерленуі керек

желідегі хосттарды іздеу үшін, Acunetix веб-осалдықтарды сканерлеу үшін, ал Metasploit Nexpose-пен бірге енуді тестілеу үшін қолданылады.

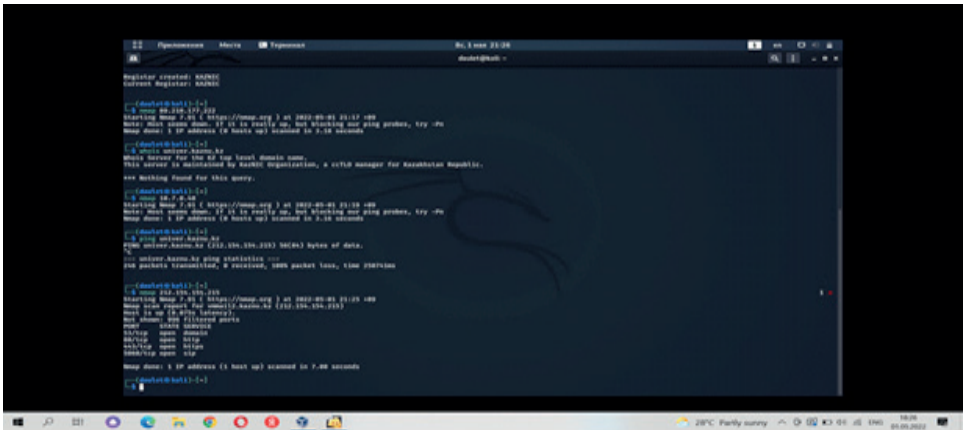
Nmap (Network Mapper) – желіні және қауіпсіздік аудитін анықтауға арналған ақысыз, ашық көзі бар желілік сканерлеу құралдарының бірі. Nmap міндеттері: хосттарды анықтау және портты сканерлеу.

Nmap және Nexpose желідегі хосттар мен қызметтерді іздеу үшін университеттің маршрутизатор аймағында орналасқан. Осы уақытта Nexpose 996 қызметі бар 5 хостты тапты, олардың ішінде университеттің негізгі сервері 27 жоғары, 15 орташа және 9 төмен осалдықпен жұмыс істейді. 2-кестеде Nexpose жасаған нәтижелер, хосттар туралы кейбір мәліметтер берілген.

2-кесте. Nexpose қосымшасымен анықталған хосттар

Табылған	IP	Хост атауы	Операциялық жүйелер	Қызметтер
19.04.22 17:31	212.154.154.215	univer.kaznu.kz	Белгісіз	465
19.04.22 17:31	212.154.154.215	ИКС	Windows 10	7
19.04.22 00:44	212.154.154.215	212.154.154.215	Linux	4

TCP қосылымын (–sT) сканерлеу белсенді хостты анықтау үшін Nmap қосымшасы арқылы жүзеге асырылады. Берілген уақыт белгісінде Nmap 996 портты сканерледі, соның ішінде ашық түрде анықталған порттар 6- суретте берілген.



6-сурет. Nmap қосымшасымен ашық порттарды анықтау

Ашық порттың күйі шабуылдаушының осы порттарды қолдана отырып желіге кіре алатындығын көрсетеді. Бағдарлама TCP қосылымдарын, UDP датаграммаларын белсенді қабылдайды. Қауіпсіздік

қызметкерлері әрбір ашық порт шабуылдың жолы екенін біледі. Зиянкестік әрекеттер ашық порттар көмегімен орындалады, ал желілік әкімші міндеті осы порттарды брандмауэрмен жабу және қорғау.

Порттарды қарап шығу кіру нүктелеріне әкеледі, олар арқылы зиянкестер желіге кіре алады. Біздің желідегі әлсіздіктерді жою үшін қауіпсіздік мамандары біздің желідегі осы осалдықтар туралы білуі керек. Nmap және Nessus осалдық сканерлері ҚазҰУ университетінің есептеу ортасында болатын желілік осалдықтар туралы ақпаратты анықтау үшін қолданылады. 7-суретте Nmap көмегімен осалдықты анықтау нәтижесі берілген. Портты сканерлеу нәтижесі 445 портының 208.91.199.121 хостында ашылғанын көрсетеді. 445 порт - бұл TCP порты, ол сервердің хабарлама блогының (SMB) осалдығына ұшырауы мүмкін. Осалдықты тексеру үшін Nmap SMB-vuln-ms08067 сценарийі іске қосылады.

```
root:~# nmap -p 445  212.154.154.213-script=smb-vuln-ms08-067.nse

Starting Nmap 7.40 ( https://nmap.org ) at 2022-04-25 10:56 IST
Nmap scan report for 192.168.1.212
Host is up (0.00050s latency).
PORT    STATE SERVICE
445/tcp open  microsoft-ds
MAC Address: 08:00:27:25:C3:51

Host script results:
| smb-vuln-ms08-067:
|   VULNERABLE:
|   Microsoft Windows system vulnerable to remote code execution (MS08-067)
|     State: VULNERABLE
|     IDs:  CVE:CVE-2008-4250
|           The Server service in Microsoft Windows 2000 SP4, XP SP2 and SP3, Server 2003 SP1 and SP2,
|           Vista Gold and SP1, Server 2008, and 7 Pre-Beta allows remote attackers to execute arbitrary
|           code via a crafted RPC request that triggers the overflow during path canonicalization.
|
|     Disclosure date: 2008-10-23
|     References:
|       https://technet.microsoft.com/en-us/library/security/ms08-067.aspx
|_      https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2008-4250

Nmap done: 1 IP_address (1 host up) scanned in 7.08 seconds
```

7 - сурет. Metasploit енуге тестілеу нәтижесі

Сканерлеу нәтижесі хосттың MS08-067 деп аталатын кодты қашықтан орындау осалдығына ұшырайтындығын көрсетеді. Бұл тексеру қауіпті және жүйенің бұзылуына әкелуі мүмкін. Ұйымның қауіпсіздік қызметкерлері үшін мұндай тексерулерді жүргізу өте маңызды, өйткені бұл осалдықты пайдалана отырып жүйеге көптеген зияндық келтіруі мүмкін.

Біз өз зерттеуімізде Metasploit-ты осалдықтарды бағалау үшін енуге тестілеу құралы ретінде пайдаланамыз. 8 - суретте Metasploit ену тестілеу нәтижесі көрсетілген.

```
msf exploit(ms08_067_netapi) > show options

Module options (exploit/windows/smb/ms08_067_netapi):

   Name      Current Setting  Required  Description
   ----      ---------------  --------  -----------
   RHOST     212.154.154.213  yes       The target address
   RPORT     445              yes       The SMB service port (TCP)
   SMBPIPE   BROWSER          yes       The pipe name to use (BROWSER, SRVSVC)

Exploit target:

   Id  Name
   --  ----
   0   Automatic Targeting

msf exploit(ms08_067_netapi) > check
[+] 192.168.1.212:445 The target is vulnerable.
msf exploit(ms08_067_netapi) >
```

8-сурет. Metasploit қосымшасымен енуді тестілеу нәтижесі

Nmap және Nessus осалдықтарын сканерлеп, Metasploit ену тестінен кейін біз ҚазҰУ университетінің желілік ортасында болатын кейбір маңызды осалдықтарды таптық. Маңызды осалдықтар жылдам әрекет етуді талап етеді, өйткені шабуылдаушыға оларды пайдалану оңай және олар жүйені толық бақылауға мүмкіндік береді.

Веб-қосымшалар мен веб-қызметтер университеттің есептеу ортасының негізгі құрамдас бөлігі болып табылады. Алайда, веб-сайттар мен веб-қосымшаларда құпия корпоративтік деректердің, несие карталарының, тұтынушы туралы ақпараттың және жеке басын анықтайтын ақпараттың ұрлануына әкелуі мүмкін осалдықтар бар.

Ұйымның қауіпсіздік деңгейін жоғарылату үшін веб-қосымшалардың қауіпсіздігін тек басымдыққа ғана емес, сонымен қатар негізгі талап ретінде қарастыру керек. Осы бағытта веб-сканерлеу ҚазҰУ университетінің веб-серверіне арналған Acunetix веб–сканерінің көмегімен орындалды (9- сурет).

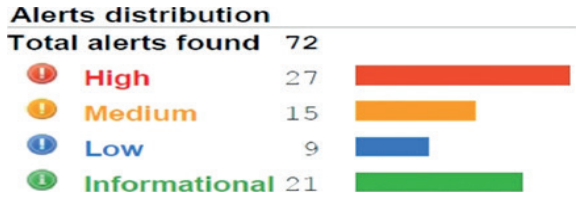
Alerts (72) Knowledge Base (4) 27 15 9 21 Generate Report

Start Date	19 Apr 2022 21:40	Files	15	Requests	17287	Host Name	http://univer.kaznu.kz
End Date	19 Apr 2022 21:40	Directories	6	Avg. Response Time	142.78 ms	Scan Target Name	Univer Kaznu Web Scan
Duration	7h 56m 5s	Variations	14	Responsive	Yes	Scan Type	Web

	Name	Module
+	Blind SQL Injection (6)	Scripting (Blind_Sql_Injection.script)
+	Cross site scripting (verified) (1)	Scripting (XSS.script)
+	Directory traversal (1)	Scripting (Directory_Traversal.script)
+	Microsoft IIS tilde directory enumeration (1)	Scripting (IIS_Tilde_Dir_Enumeration.script)
+	Script source code disclosure (1)	Scripting (Script_Source_Code_Disclosure.script)
+	SQL injection (verified) (15)	Scripting (Sql_Injection.script)
+	Weak password (2)	Scripting (Html_Authentication_Audit.script)
+	Application error message (10)	Scripting (Error_Message.script)
+	HTML form without CSRF protection (3)	Crawler
+	User credentials are sent in clear text (2)	Crawler
+	ASP.NET version disclosure (1)	Scripting (ASP_NET_Error_Message.script)
+	Clickjacking: X-Frame-Options header missing (1)	Scripting (Clickjacking_X_Frame_Options.script)
+	Cookie without HttpOnly flag set (1)	Crawler
+	Cookie without Secure flag set (1)	Crawler
+	Login page password-guessing attack (4)	Scripting (Html_Authentication_Audit.script)

9-сурет. Acunetix веб-сканерлеу нәтижесі

10 - суретте анықталған хост ескертулері жинақталған сыртқы сканерлеу нәтижесі берілген.



10-сурет. Acunetix қосымшасымен 212.154.154.213 хостты веб-сканерлеу нәтижесі

Тәуекелді бағалау үшін ықтималдық, статистика және ақпараттық технологияларды білетін білікті мамандар қажет. Тәуекелдерді өлшеудің алғашқы қадамы әртүрлі сканерлерден, атап айтқанда, Nexpose, Acunetix және Metasploit–тен алынған барлық сканерлеу нәтижелерін орталықтандыруды қажет етеді (3-кесте).

3-кесте. Интеграцияланған сканерлеу нәтижелері

Осалдық	Қатаңдық	Барлық ескертулер	Категория
Әлсіз пароль	7,5	2	Қатыгез шабуыл
Әлсіз пароль	7,5	2	Аутентификацияның жеткіліксіздігі
Сайт аралық скриптинг (тексерілген)	4,4	1	Сайт аралық скриптинг
Соқыр SQL инъекциясы	7,8	6	SQL–инъекция
SQL инъекциясы (тексерілген)	7,8	15	
Microsoft IIS Tilda каталогтарын тізімдеу	2,6	1	Ақпараттың ағуы
Сценарийдің бастапқы кодын ашу	2,6	1	
Әлсіз пароль	7,5	2	
Қолданба қатесі туралы хабар	5,0	10	
Нұсқаны ашу ASP.NET	0,0	1	
Microsoft IIS нұсқасын ашу	0,0	1	
Автотолтыру қосылған пароль түрін енгізу	0,0	4	
Каталогты айналып өту	6,8	1	Айналма жолдары
CSRF қорғаусыз HTML пішіні	8,6	6	Функционалдылықты теріс пайдалану
Clickjacking: x–Frame-Options тақырыбы жоқ	6,8	1	
Кіру бетіндегі парольді шабуылдау	6,8	4	

Сканерлеу нәтижелері бойынша жиналған осалдық туралы барлық мәліметтермен қауіпсіздік мамандары жергілікті желілік белсенділік пен құрылғы конфигурацияларымен қатар осалдықтарды бағалаудың жалпы жүйесі (Common Vulnerability Scoring System, CVSS) ретінде тәуекелдерді басымдыққа ие болуы керек. Тәуекел деңгейі анықталған осалдықтардың қайсысы жүйеге шынымен қауіп төндіретінін анықтайды, нәтижеде осалдықтар тәуекел мөлшеріне сәйкес жойылады. Тәуекелдің мәні эксплуатацияны қолдану ықтималдығына байланысты, сондай-ақ, осалдықтың пайда болу жиілігі жүйеде осалдықтың пайда болу күніне байланысты. Осалдық тәуекелінің жиілігі мен сандық деңгейі қауіпсіздік тәуекелінің сандық моделінің математикалық теңдеулерімен анықталады, ол CVSS базалық бағаларын жоғарылату үшін уақыт пен экологиялық өлшемдерді есептейді, содан кейін тәуекелдің соңғы мәнін шығарады. Тәуекел деңгейін сандық бағалау 0-ден 10-ға дейінгі диапазонда болады, бұл сандық бағалауды ұйымдарға осалдықты басқару процестерін дұрыс бағалауға және басымдық беруге көмектесу үшін сапалы көрініске аударуға болады (4-кесте).

Тәуекел деңгейін сандық өлшеу кезінде, эксплуатацияның деңгейімен қатар, біз ескертулердің жалпы саны, эксплуатациялық элемент, әсер етілген параметр және осалдықтарды сканерлеу кезінде анықталған опциялар сияқты көптеген факторларды ескереміз. Тәуекелдерді бағалау нәтижелеріне сүйене отырып, ұсынылған схеманың келесі кезеңі тәуекел деңгейін төмендететін қарсы шаралардың жаңартуларын анықтау болып табылады.

4-кесте. Сапа тәуекелдерін бағалау шкаласы

Тәуекелдің сандық шамасы	Тәуекел санаты	Сипаттамасы
9,0-ден 10,0-ге дейін	Критикалық	Тәуекел мүлдем қолайсыз; пайда болу ықтималдығын азайту үшін дереу әрекет етуді талап етуі керек.
7,0-ден 8,9-ға дейін	Жоғары	Тәуекел қабылданбайды; қалпына келтіру жоспарын мүмкіндігінше тезірек орындау қажет.
4,0-ден 6,9-ға дейін	Орташа	Тәуекел қысқа мерзім ішінде қолайлы болуы мүмкін; болашақ іс-әрекеттерге және бюджеттік жоспарларға тәуекелді азайту жөніндегі шараларды енгізуді талап етеді.
0,1-ден 3,9-ға дейін	Төмен	Тәуекелдер қолайлы; тәуекелді одан әрі төмендету жөніндегі жоспарлар қауіпсіздіктің басқа жаңартуларымен іске асырылуы тиіс.

ҚазҰУ желісінің қауіпсіздігін арттыру үшін анықталған тәуекелдер бойынша келесі ұсыныстар берілген:

1) SQL енгізу: ҚазҰУ университетінің есептеу ортасы SQL енгізу туралы қауіпсіздік жүйесінің тек 21 ескертуін анықтады және келесі элементтер қозғалды: / Login.asp, /Register.asp, /Search.asp, /showforum.asp және / showthread.asp. SQL енгізу шабуылдары хакердің мүддесі үшін бағдарламаның сипатын өзгертетін SQL сұрауларының нысанын өзгертеді. Prepared Statement көмегімен серверлік қорғаныс SQL инъекцияларынан қорғаудың ең тиімді әдісі болып табылады, өйткені сұрау ниеті өзгермейді.

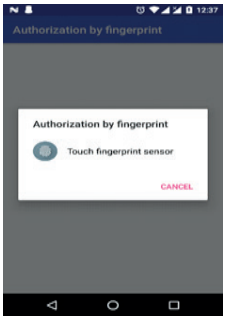
2) Әлсіз пароль: университет желісінде /Login–де әлсіз пароль туралы 6 ескерту табылды. 30 күннен асқан бірнеше пароль шоттары бар, ал кейбіреулері тіпті бір жылға жуық. Парольді бұзу -қазіргі қауіпсіздік деңгейін бағалау үшін қолданылатын ең көп таралған элементтердің бірі. Әлсіз парольдің осалдығын жеңудің қарапайым әдісі - пароль саясатын қолдану, яғни, парольдің ұзындығы 8 таңбадан артық болуы керек, кем дегенде бір бас әріп болуы керек, парольді таңдағанда кем дегенде бір сан және бір арнайы таңба болуы керек.

3) CSRF шабуылдары: осал элементтері /логині бар барлығы 6 нұсқа табылды: asp, /Register.asp және /Іздеу.asp. Негізгі қауіп браузердің сұраныстарды қалай өңдейтініне байланысты. Қарапайым мысал: веб–бағдарлама пароль туралы ақпаратты беру үшін HTTP сұрауында GET әдісін қолданады; get пайдалану кезінде шолғыш форма деректерін URL мекен–жайына кодтайды. Пішін деректері URL мекенжайында болғандықтан, олар шолғыштың мекен-жай жолында көрсетіледі, нәтижеде ақпарат қол жетімді. Ең оңай шешім-POST әдісін қолдану, POST әдісін қолданған кезде форма деректері URL мекенжайында емес, HTTP сұрау хабарламасында көрсетіледі.

Талқылау. Зерттеу көрсеткендей, тәуекелдерді бағалау нәтижелері жоғары басшылыққа, процедуралық, бюджеттік және жүйелік операциялық және басқарушылық өзгерістер туралы шешім қабылдауға көмектесетін ресми есеп форматында құжатталады. Тәуекелдерді бағалау рекурсивті процедура болғандықтан, бұл түпкілікті жасалған есеп тәуекелдерді бағалау процедурасының келесі циклінде ұсынылған құрылымның 1-кезеңі үшін кіріс ретінде пайдаланылады.

ҚазҰУ жүйесінің құпиясөзге байланысты осалдықтарының алдын алу мақсатында мобильді қосымшаға жаңа тексерулер енгізуді ұсынамыз. Ең алдымен, авторизацияланған қолданушы мобильді қосымшаға кірген сәтте саусақ ізі арқылы аутенификация жасалынады, екіншіден авторизацияланбаған қолданушы өз аккаунтына кірген сәтте, мобильді хабарлама аутентификация сұрайды.

11-суретте қосымшаға кіру үшін жасалған жаңа функционал, мұнда мобильді қосымшаға кірген сәтте шығатын аутентификация көрсетілген.

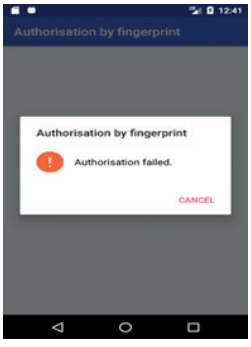


11-сурет. Саусақ ізі бойынша аутентификация

12-суретте қолданушының саусақ іздері, телефонға тіркелген қолданушыны саусақ ізімен сәйкестігі тексеріледі, дұрыс жағдайда біз өзімізге қажетті әрекеттерді жасай аламыз.

12-сурет. Саусақ іздерінің сәйкестігі анықталғаннан кейінгі нәтиже

13-суретте егер қолданушының саусақ іздері телефонда тіркелген саусақ ізіне сәйкес келмеген жағдайда шығатын хабарлама көрсетілген.



13 -сурет. Саусақ іздері сәйкестігі анықталмаған жағдай

14-суретте ҚазҰУ жүйесіне кірген сәтте сұрайтын аутентификация түрі, яғни мобильді қосымшада авторизация жасалған кезде сұралады, ал веб қосымшада міндетті түрде хабарлама сұралады.

14-сурет. Мобильді хабарлама авторизациясы

Қорытынды. Бұл құжат университеттің есептеу ортасы үшін ақпараттық қауіпсіздік тәуекелдерін сандық бағалау құрылымын ұсынады. Ұсынылған модельдің мақсаты қауіпсіздік ережелерін бұзу қаупін азайту болып табылады, бұл кампус желісін осал ететін себептерді түсінуді білдіреді. ҚазҰУ кампус желісіне ұсынылған құрылымды қолдана отырып, желінің қауіпсіздігін қамтамасыз етудің қазіргі тәсілдері университет ортасы тұрғысынан тиімсіз екендігі белгілі болды; өйткені университеттің есептеу ортасы банктер сияқты бұзу мақсаттарынан өзгеше. Бағалау зерттеулері ҚазҰУ желісінде пароль саясатын мәжбүрлеп қолдану сияқты мәселелерді қарастырады. Қашықтан кіруді басқару және рұқсаттарды міндетті есептік жазбалармен шектеу.

Ұсынылған модель университеттің желілік конфигурациясы үшін тәуекел мөлшерін сандық түрде өлшейді және оны нақты және қол жетімді түрде сенімді және қайталанатын тәуекелдерді талдауды жүзеге асыру үшін тәуекел талдаушысы және университеттің қауіпсіздік менеджері қолдана алады. ҚазҰУ зерттелу бойынша авторизация осалдықтары анықталды, сол себепті солардың алын алу шаралары жасалынды, оның ішінде саусақ ізі аутентификациясы және хабарлама авторизациясы. Осы ұсынылған алдын алу шаралары Андроид операциялық жүйесінде қолданба жасалынды, ол REST сұраныс арқылы смартфонмен байланыс жасайды.

Ұсынылған құрылымды кез-келген жоғары білім беру ұйымына

немесе университеттің IT-ортасына қолдануға болады. Бұл университеттерге қауіпсіздікке төнетін қауіп-қатерлерден бір қадам алда тұруға, сонымен қатар, қауіп-қатерге ұшыраған маңызды активтерге назар аудара отырып, қауіпсіздік бюджетінен көбірек алуға мүмкіндік береді.

Information about the authors:

Iztayev Zhalgasbek – Cand. Sci. (Pedagogical), M. Auezov South Kazakhstan University, Shymkent, Kazakhstan, jalgasbek_uko@mail.ru, https://orcid.org/0000-0002-3210-2963;

Dzhusupbekova Gulzat – Cand. Sci. (Pedagogical), M. Auezov South Kazakhstan University, Shymkent, Kazakhstan, gulzat20.10@mail.ru, https://orcid.org/0000-0003-1727-0966;

Ordabayeva Gulzinat – Senior Lecturer, Al-Farabi Kazakh National University; Almaty, Kazakhstan, gulzi200988@mail.ru; https://orcid.org/0000-0001-9952-1620.

ӘДЕБИЕТТЕР:

Ахметов Б.С., Корченко А.Г., Жекамбаева М.Н., Казмирчук С.В. (2015). Қауіптің базалық сипаттамасының кортежді моделі // Қазақстан Республикасының ұлттық ғылым академиясының баяндамалары. – 2015. - №6. – 12-19б. (қазақ тілінде).

Ахметов Б., Қыдыралина Л., Лахно В., Могилный Г., Ахметова Ж., Ташимова А. (2018). Оқу орындарының кибер қауіпсіздігіне өзара инвестициялар бойынша компьютерлік шешімдерді қолдау жүйесінің моделі // Халықаралық машина жасау және технологиялар журналы. – 2018. - Vol.9.-Iss.10.-P.1114-1122 (ағылшын тілінде).

Барабанов А.В., Дорофеев А.В., Марков А.С., Цирлов В.Л. (2017) Жеті қауіпсіз ақпараттық технология/ А.С. Марков ред. басқаруымен.: ДМК Пресс. - 2017. - 224 б.: сурет. (орыс тілінде).

Джаафер Әл-Сарайре, Ала ‘Масарве (2022). Алдыңғы қатарлы қауіп-қатерлерді анықтауға арналған жаңа тәсіл, Мысыр информатика журналы, 2022, ISSN 1110-8665, https://doi.org/10.1016/j.eij.2022.06.005. (https://www.sciencedirect.com/science/article/pii/S1110866522000470, 05.07.2022) (ағылшын тілінде).

И.И. Лившиц (2018). Күрделі өнеркәсіптік объектілерді басқарудың интеграцияланған жүйелерінің ақпараттық қауіпсіздік аудитінің модельдері мен әдістері: техн. ғылымдары докторы дисс.: 05.13.19/ СПИИРАН. - Санкт-Петербург. 2018. -407 б. (орыс тілінде).

Кхайрур Разикин, Бенф ано Соевито (2022). Тәуекелдерді талдау және киберқауіпсіздік құрылымы негізінде ақпараттық қауіпсіздік жүйесін әзірлеу үшін киберқауіпсіздік бойынша шешімдер қабылдауды қолдау моделі, Egyptian Informatics Journal, 2022, ISSN 1110-8665, https://doi.org/10.1016/j.eij.2022.03.001. (https://www.sciencedirect.com/science/article/pii/S1110866522000226, 01.07.2022) (ағылшын тілін-

де).

Кхандо Кхандо, Шанг Гао, Сираджул М. Ислам, Али Салман (2021). Жеке және мемлекеттік ұйымдардағы ақпараттық қауіпсіздік туралы қызметкерлердің хабардарлығын арттыру: Әдебиетке жүйелі шолу, Computers & Security, Volume 106, 2021, 102267, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2021.102267. (https://www.sciencedirect.com/science/article/pii/S0167404821000912, 30.06.2022) (ағылшын тілінде).

Қалимолдаев М.Н., Бияшев Р.Г., О.А. Рог (2017). Ақпаратқа қолжетімділікті шектеу үлгілерін құру үшін логиканы қолдану // Қазақстан Республикасы Ұлттық ғылым академиясының баяндамалары. – Almaty. – 2017.- Vol.3 – Num. 313 (2017). -P.48-54 (орыс тілінде).

ҚР СТ 1698-2007 (2007). Ақпаратты қорғау. Ақпаратты техникалық барлаудан және техникалық есептеу техникасы бойынша жылыстаудан қорғау. Қорғау әдістері. [Электрондық ресурс]// https://online.zakon.kz/Document/?doc_id=30374214&pos=6; -108 #pos = 6; -108. 23.06.2022 (орыс тілінде).

ҚР СТ ГОСТ Р ИСО/МЭК 15408-1-2006 (2006). Ақпараттық технология. Қауіпсіздікті қамтамасыз ету әдістері мен құралдары. Ақпараттық технологиялардың қауіпсіздігін бағалау критерийлері. 1-бөлім. Қазақстан Республикасы Индустрия және сауда министрлігінің Техникалық реттеу және метрология комитеті. - 2006. - 100 с. [Электрондық ресурс]// http://db4.sbras.ru/elbib/data/show_page.phtml?22+82. 23.06.2022 (орыс тілінде).

Лахно В., Ахметов Б., Ыдырышбаева М. және Ербол А. (2021). Кибершабуылдарды тану үшін шешімдер қабылдауды қолдау жүйелерінің білім базаларын қалыптастыру модельдері. «Физика-математика ғылымдары». 76. 4 (2021 желтоқсан), 88-98. DOI: https://doi.org/10.51889/2021-4.1728-7901.12. (қазақ тілінде).

Ордабаева Г.К. (2020). Желілік топологияның қауіпсіздік моделі мен әдістерін әзірлеу. Информатика және қолданбалы математика: V Халықаралық ғылыми конференция материалдары (29 қыркүйек -1 қазан 2020 ж.). Алматы, 2020. – б. 367-373 (қазақ тілінде).

Сагар Рахалкар Пуна (2019). Енуді тестілеу бойынша қысқаша нұсқау (NMAP, OpenVAS және Metasploit қолдану арқылы), Махараштра, Үндістан. ISBN-13 (pbk): 978-1-4842-4269-8\ ISBN-13 (electronic): 978-1-4842-4270-4 https: //doi.org/10.1007/978-1-4842-4270-4. https: //www.baikalctf.ru/data/documents/Kratkoe_rukovodstvo_po_testirovaniyu_na_proniknovenie.pdf, 03.07.2022 (орыс тілінде).

Төкеев У.А., Ахметов Б.Б. (2011). Ақпараттық қауіпсіздікті басқару: оқу құралы. – Алматы: әл-Фараби атындағы Қазақ ұлттық университеті, 2011. – 161 б. (қазақ тілінде).

Цифрлық экожүйенің дамытудың 2022-2027 жылдарға арналған («Киберқалқан-2») тұжырымдамасы (2022). [Электронды ресурс] https://www.gov.kz/memleket/entities/mdai/documents/details/320322?lang=kk. 23.06.2022 (қазақ тілінде).

Эрик Филиоль, Франческо Меркальдо, Антонелла Сантоне (2021). Енуге және салдарын жеңілдетуге автоматты тестілеу әдісі: қызыл қалпақ, Procedia Computer Science, том 192, 2021, 2039-2046, ISSN 1877-0509, *https://doi.org/10.1016/j.procs.2021.08.210*. (https://www.sciencedirect.com/science/article/pii/S1877050921017063, 05.07.2022) (ағылшын тілінде).

\

REFERENCES:

Akhmetov B.S., Korchenko A.G., Zhekambaeva M.N., Kazmirchuk S.V. (2015). Motorcade model of basic hazard characteristics//Reports of the National Academy of Sciences of the Republic of Kazakhstan. – 2015. - №6. - gr. 12-19.

Akhmetov B., Kydyralina L., Lakhno V., Mohylnyi G., Akhmetova J., Tashimova A. (2018). Model for a computer decision support system on mutual investment in the cybersecurity of educational institutions // International Journal of Mechanical Engineering and Technology. – 2018. – Vol.9.-Iss.10.-P.1114-1122.

Barabanov A.V., Dorofeev A.V., Markov A.S., Tsirlov V.L. (2017). Seven safe information technologies/edited by A.S. Markova. - M.: DMK Press. - 2017. - 224 S.: silt.

Eric Filiol, Francesco Mercaldo, Antonella Santone (2021). A Method for Automatic Penetration Testing and Mitigation: A Red Hat Approach, Procedia Computer Science, Volume 192, 2021, Pages 2039-2046, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.08.210. (https://www.sciencedirect.com/science/article/pii/S1877050921017063, 05.07.2022).

Jaafer Al-Saraireh, Ala' Masarweh (2022). A novel approach for detecting advanced persistent threats, Egyptian Informatics Journal, 2022, ISSN 1110-8665, https://doi.org/10.1016/j.eij.2022.06.005. (https://www.sciencedirect.com/science/article/pii/S1110866522000470, 05.07.2022).

Kalimoldaev M.N., R.G. Biyashev, O.A. Horn (2017). Application of logic for building models of delimitation of access to information//Reports of the National Academy of Sciences of the Republic of Kazakhstan. – Almaty. – 2017.- Vol.3 – Num. 313 (2017).- P.48-54.

Khairur Razikin, Benf ano Soewito (2022). Cybersecurity decision support model to designing information technology security system based on risk analysis and cybersecurity framework, Egyptian Informatics Journal, 2022, ISSN 1110-8665, https://doi.org/10.1016/j.eij.2022.03.001. (https://www.sciencedirect.com/science/article/pii/S1110866522000226, 01.07.2022).

Khando Khando, Shang Gao, Sirajul M. Islam, Ali Salman (2021). Enhancing employees information security awareness in private and public organisations: A systematic literature review, Computers & Security, Volume 106, 2021, 102267, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2021.102267. (https://www.sciencedirect.com/science/article/pii/S0167404821000912, 30.06.2022).

Lakhno V., Akhmetov B., Ydyryshbaeva M. and Erbol A. (2021). Models for building knowledge bases of decision support systems for recognizing cyber attacks. «Physical and Mathematical Sciences.» 76, 4 (Dec. 2021), 88-98. DOI:https://doi.org/10.51889/2021-4.1728-7901.12.

Livshits I.I. (2018). Models and methods of auditing information security of integrated systems for managing complex industrial facilities: diss. Doctor of Technology. Sciences: 05.13.19/SPIIRAN. - St. Petersburg. 2018.-407 p.

Ordabaeva G.K. (2020). Development of network topology security model and methods. Informatics and Applied Mathematics: materials of the V International Scientific Conference (September 29 - October 1, 2020). Almaty, 2020. - b. 367-373.

Sagar Rahalkar Pune. (2019). Penetration Test Brief Guide (using NNMAP, OpenVAS and Metasploit), Maharashtra, India. ISBN-13 (pbk): 978-1-4842-4269-8\ ISBN-13 (electronic): 978-1-4842-4270-4 https://doi.org/10.1007/978-1-4842-4270-4. https: //www.baikalctf.ru/data/documents/Kratkoe_rukovodstvo_po_testirovaniyu_na_proniknovenie.pdf, 03.07.2022.

ST RK GOST R ISO/IEC 15408-1-2006 (2006). Information technology. Methods and means of safety assurance. Information Technology Security Assessment Criteria. Part 1. Introduction and general model.//- Committee on Technical Regulation and Metrology of the Ministry of Industry and Trade of the Republic of Kazakhstan. - 2006. - 100 c. [Electronic resource]//http://db4.sbras.ru/elbib/data/show_page.phtml?22+82. 23.06.2022.

ST RK 1698-2007 (2007). Information protection. Protection of information from technical intelligence and from leakage by technical computer equipment. Protection methods. [Electronic resource]// https://online.zakon.kz/Document/?doc_id=30374214&pos=6; -108 # pos = 6; -108. 23.06.2022.

Tokeev U.A., B.B. Akhmetov (2011). Information Security Management: Textbook. - Al-Farabi Kazakh National University, 2011. - Art. 161.

The concept of the development of the digital ecosystem for 2022-2027 (Cyberkalkan-2) (2022). [Electronic Resource] https://www.gov.kz/memleket/entities/mdai/documents/details/320322?lang=kk. 23.06.2022.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 117-135
https://doi.org/10.32014/2022.2518-1726.142
УДК 004.49

Ж.С. Каженова[1*], Ж.Е. Кенжебаева[1], А.М. Прудник[2]

[1]С. Сейфуллин атындағы Қазақ агротехникалық университеті,
Қазақстан, Астана;

[2]Беларусь мемлекеттік информатика және радиоэлектроника
университеті, Белорусь, Минск.
E-mail: *zhkazhenova75@gmail.com*

MQTT (ТЕЛЕМЕТРИЯ ХАБАРЛАМАЛАРЫ КЕЗЕГІН ТАСЫМАЛДАУ) ХАТТАМАСЫНЫҢ ҚАУІПСІЗДІК МЕХАНИЗМДЕРІ

Аннотация. Заттар интернеті (IoT) соңғы жылдары зерттеулердегі өте өзекті тақырыпқа айналды. IoT желісіне жеңілдетілген болуы тиіс жаңа байланыс хаттамалары қосылды, мысалы, MQTT. Мақалада заттар интернетінің MQTT (Message Queue Telemetry Transport) хаттамасының ерекшеліктері, қолданылу нұсқалары және өзіндік сипаттағы процедуралары талқыланады. «Жариялаушы-жазылушы» принципі қарастырылады. MQTT хаттамасына әуел бастан тиімді қауіпсіздік функ циялары жетіспейді, себебі ол қарапайым мәтін түрінде пайда-ланушы аты мен құпия сөзге негізделген аутентификацияны орындайды. Сонымен қатар, тасымалдау деңгейінде толық шифрлау үшін SSL/TLS пайдаланылады, ал ол шектеулі ресурстары бар құрыл ғылар үшін жеңіл хаттама болып саналмайды. Егер IOT құрылғылары байланыстың үстеме шығындарын көтере алса, TLS әр уақытта қолданылуы керек, себебі, қауіпсіздік IOT әзірлемесінің ажырамас бөлігі болып табылады. Бұл мақалада біз кейінірек MQTT хаттамасына негізделген заттар интернеті (IoT) үшін жаңа жеңілдетілген аутентификация механизмін ұсыну мақсатында MQTT хаттамасының қауіпсіздік механизмдерін зерттейміз. Ол үшін MQTT хаттамасы арқылы екі IoT құрылғысы (Raspberry pi) арасында деректер алмасуды жүзеге асырдық. Құрылғы-

лардың бірі брокер және жазылушы, екіншісі жариялаушы рөлін атқарады. MQTT хаттамасында пайдаланушы аты және құпия сөзді қолданумен қорғалған арна және TLS хаттамасымен қорғалған арна жүзеге асырылуы мүмкін. MQTT хаттамасының қауіпсіздік механизмдеріне сәйкес, екі IoT құрылғысы арасында қорғалмаған арна арқылы ақпарат алмасу, пайдаланушы аты және құпия сөзді қолданумен қорғалған арна арқылы және TLS хаттамасымен қорғалған арна арқылы ақпарат алмасу кезіндегі арнаның өткізу қабілеттері өлшенді және салыстырылды.

Түйін сөздер: жариялаушы, жазылушы, заттар интернеті, қауіпсіздік, хаттама, MQTT, IoT құрылғылары.

Ж.С. Каженова[1*], Ж.Е. Кенжебаева[1], А.М. Прудник[2]

[1]Казахский агротехнический университет имени С. Сейфуллина,
Казахстан, Астана;
[2]Белорусский государственный университет информатики
и радиоэлектроники, Беларусь, Минск.
E-mail: *zhkazhenova75@gmail.com*

МЕХАНИЗМЫ БЕЗОПАСНОСТИ ПРОТОКОЛА MQTT (ТРАНСПОРТ ТЕЛЕМЕТРИИ ОЧЕРЕДИ СООБЩЕНИЙ)

Аннотация. Интернет вещей (IoT) в последние годы стал очень актуальной темой в исследованиях. В сети IoT включены новые коммуникационные протоколы, которые должны быть легкими, как, например, протокол MQTT. В статье рассматривается протокол MQTT (Message Queue Telemetry Transport) Интернета вещей, его особенности, варианты применения, характерные процедуры. Рассматривается принцип «издатель-подписчик». MQTT изначально не хватает эффективных функций безопасности, поскольку он выполняет аутентификацию на основе имени пользователя и пароля в виде простого текста. Кроме того, для полного шифрования на транспортном уровне используется SSL/TSL, который не считается облегченным протоколом для устройств с ограниченными ресурсами. Если устройства IOT могут нести накладные расходы на связь, TLS следует использовать каждый раз, так как безопасность является неотъемлемой частью разработки IOT.

В этом документе мы исследуем механизмы безопасности протокола MQTT с целью в дальнейшем представить новый облегченный механизм аутентификации для Интернета вещей (IoT) на основе протокола MQTT. Для этого мы реализовали обмен данными между двумя устройствами IoT (Raspberry pi) по протоколу MQTT. Один из усройств выступит в роли брокера и подписчика, другое будет выступать в роли издателя. Протокол MQTT может реализовать канал, защищенный именем пользователя и паролем, и канал, защищенный протоколом TLS. Согласно механизмам безопасности протокола MQTT, измерялась и сравнивалась пропускная способность канала при обмене информацией между двумя устройствами IoT по незащищенному каналу, по каналу, защищенному с помощью имени пользователя и пароля, и по каналу, защищенному протоколом TLS.

Ключевые слова: издатель, подписчик, Интернет вещей, безопасность, протокол, MQTT, устройства IoT.

Zh.S. Kazhenova[1*], Zh.E. Kenzhebayeva[1], A.M. Prudnik[2]

[1]S. Seifullin Kazakh Agrotechnical University, Kazakhstan, Astana;
[2]Belarusian State University of Informatics and Radioelectronics, Belarus, Minsk.
E-mail: *zhkazhenova75@gmail.com*

SECURITY MECHANISMS OF PROTOCOL MQTT (MESSAGE QUEUING TELEMETRY TRANSPORT)

Abstract. The Internet of Things (IoT) has become a very hot topic in recent years. The IoT network includes new communication protocols that should be lightweight, such as the MQTT protocol. The article discusses the MQTT (Message Queue Telemetry Transport) protocol of the Internet of Things, its features, applications, and characteristic procedures. The principle of “publisher-subscriber” is considered. MQTT initially lacks effective security features as it performs authentication based on a username and password in plain text. In addition, SSL/TSL is used for full transport layer encryption, which is not considered a lightweight protocol for devices with limited resources. If IOT devices can incur communication overhead, TLS should be used every time, as security is an integral part of IOT development.

In this paper, we explore the security mechanism of the MQTT protocol, with the aim of further introducing a new lightweight authentication mechanism for the Internet of Things (IoT) based on the MQTT protocol. To do this, we implemented data exchange between two IoT devices (Raspberry pi) using the MQTT protocol. One of the devices will act as a broker and subscriber, the other will act as a publisher. The MQTT protocol can implement a username and password protected channel and a TLS protected channel. According to the security mechanisms of the MQTT protocol, the channel throughput was measured and compared when exchanging information between two IoT devices over an insecure channel, over a channel protected with a username and password, and over a channel protected by the TLS protocol.

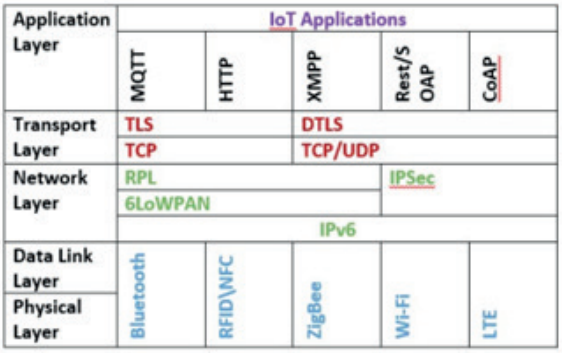
Key words: publisher, subscriber, Internet of Things, security, protocol, MQTT, IoT devices.

Кіріспе. Қазіргі кезде интернетке қосылған құрылғылардың және осы құрылғылар жасайтын деректер көлемінің жылдам өсуі байқалады. Кіріктірілген датчиктерді, сымсыз байланыстарды, процессорларды біріктіру арқылы заттар интернеті (IoT) желісін құруға болады. Заттар интернеті екі терминнен тұрады. Бірінші термин – бұл Интернет, ол миллиардтаған пайдаланушыларды, құрылғыларды, жүйелерді байланыстырады. Екінші термин – бұл интеллектуалды объектілерге жататын зат (Goyal т.б., 2018).

Заттар интернеті серпілісі деректер алмасу көрінісін өзгертті. Заттар интернетінің арқасында машинадан - машинаға (M2M) байланыс түрі пайда болды. Дегенмен, күнделікті заттарды интернетке қосу маңызды қауіпсіздік мәселелерін тудырады. Қауіпсіздік - заттар интернеті желілеріндегі басты мәселелердің бірі. IoT құрылғыларының көпшілігі ресурстарды және қуат тұтынуда шектеулі болғандықтан, сенімді қауіпсіздік механизмдерін енгізу оңай емес. Біздің алдыңғы мақаламызда IoT желілерінде кеңінен қолданылатын технологиялар мен стандарттарға сипаттама жасалған, сондай-ақ, қазіргі уақытта IoT жүйесінде қабылдау үшін қол жетімді ең танымал қауіпсіздік хаттамалары мен технологияларына шолу жасалған (Каженова т.б., 2022).

Ресурстарды тиімді пайдаланатын қолданбалы хаттамалар IoT ортасында хабар алмасу мен деректерді жеткізудің құрылыс блоктары болып табылады. Шектеулі қосымшалар хаттамасы (CoAP), кеңейтілетін хабар алмасу және қатысу хаттамасы (XMPP) және телеметрия

хабарламалары кезегін тасымалдау (MQTT) сияқты қосымша деңгей хаттамалары хабар алмасу мақсатында әзірленді. IoT хаттамалары стегінің жалпы тізімін 1-суретте көруге болады (Andy т.б., 2017).



1-сурет. IoT хаттамаларының стегі.

Қолданбалы деңгей соңғы түйіндер (IOT құрылғылары) мен желі арасындағы интерфейсті қамтамасыз етеді. Заттар интернетінде қолданылатын танымал хаттамалардың бірі - MQTT. MQTT (Message Queue Telemetry Transport) - бұл қосымша деңгей хаттамасы. Компьютерлер, ноутбуктер және мобильді құрылғылар жағдайында қосымша деңгей ролін әдетте браузер атқарады. IoT құрылғылары жағдайында қосымшалар деңгейі іске қосылған операциялық жүйе арқылы (егер ол ендірілген ОЖ жұмыс істеп тұрса) немесе микробағдарлама арқылы жүзеге асырылуы мүмкін.

MQTT соңғы жылдары IoT үшін жеңіл құрылымы мен пайдаланудың қарапайымдылығына байланысты IoT сахнасына шықты. MQTT хаттамасы Facebook Messenger-де қолданылуына және оның IoT-ке тікелей сілтемелеріне байланысты көпшіліктің назарында ілікті (Bryce т.б., 2018). 2011 жылы Люси Чжан және оның командасы Facebook-ке қосылып, Facebook мессенджерін әзірлей бастады. MQTT көмегімен олар дербес жұмыс істеу уақытын төмендетпей, Facebook серверлерімен тұрақты байланыс орната алды.

Қауіпсіздік – заттар интернеті қосымшаларын дайындаудың ажырамас бөлігі. Мұнда негізгі мақсат қарапайым күнделікті заттарды Интернетке қосу ғана емес, сонымен қатар әртүрлі соңғы нүктелер арасында деректерді қауіпсіз тасымалдау болып табылады. Осылайша интеллектуалды заттар интернеті қосымшалары әртүрлі талаптарды қанағаттандыруда тиімді және табысты болып қана қоймай, сонымен қатар жоғары сенімділікке ие болуы керек, яғни соңғы пайда-

ланушылардың құпиялылығын сақтау кез келген веб-қызметтің міндеті болып табылады.

Бізге MQTT қосымша деңгей хаттамасын пайдалана отырып, практикалық IoT әзірлеуге көшкен кезде, хаттамалар стегінде қол жетімді қауіпсіздік функцияларын түсіну маңызды. Бұл қауіпсіздік функцияларын IOT қосымшаларын әзірлеудің басынан бастап қолдану қажет. MQTT – шектеулі ресурстары бар IoT құрылғыларын қолдануына және желінің шектеулі өткізу қабілеттілігі кезінде деректер алмасуға арналған жеңілдетілген хаттама. IoT қосымшаларының масштабтылығының салдарынан құрылғылардың және жүйелердің осалдығын, сондай-ақ, олардың интернетте үнемі қол жетімділігін ескере отырып, MQTT хаттамасында желі деңгейінде, тасымалдау деңгейінде, сондай-ақ қосымша деңгейінде қосылған қауіпсіздік мүмкіндіктері бар. OSI деңгейлерінде әртүрлі қауіпсіздік мүмкіндіктерін қосу арқылы әртүрлі шабуыл түрлерін болдырмауға болады. Осы қауіпсіздік мүмкіндіктерінің барлығы әртүрлі жеткізушілер мен стандартты ұйымдар мойындаған стандартты механизмдер.

Бұл мақалада біз MQTT хаттамасының қауіпсіздік механизмдерін зерттейміз. Осы зерттеудің нәтижесі одан әрі MQTT хаттамасының негізіндезаттаринтернеті(IoT)үшінжаңа,жеңілдетілгенаутентификация механизмін жүзеге асыру барысында қажет болады. Зерттеу барысында MQTT хаттамасы арқылы екі IoT құрылғысы (Raspberry pi) арасында деректер алмасу жүзеге асырылды. Құрылғылардың бірі MQTT брокері және жазылушы-клиент, екіншісі жариялаушы-клиент рөлін атқарады. Брокер мен клиент арасында ақпарат алмасу үш түрлі тәсілмен жүзеге асырылды:

1. Ашық порт арқылы, яғни қорғаусыз қосылу. Бұл әзірлеу және практикалық қолданыстан бұрын MQTT-құрылғыларын сынау кезінде қолайлы. Брокер мен клиент арасында деректер шифрлаусыз қарапайым мәтін түрінде берілді.

2. Пайдаланушының аты мен құпия сөзін пайдалану арқылы қосылу. Бұл әдіс арқылы брокер клиенттің жеке басын тексере алады. Дегенмен, жіберу кезінде деректер шифрланбайды.

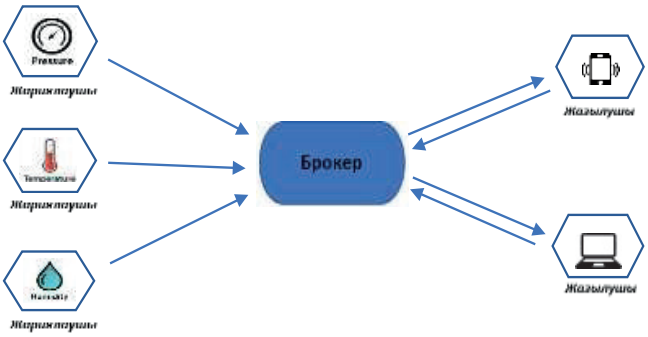
3. SSL шифрлауын пайдалану арқылы қосылу. Бұл Mosquitto MQTT брокерінде қол жетімді ең жетілдірілген шифрлау және аутентификациялау схемасы. SSL пайдалану кезінде сенімді сертификаттау орталығынан (Certificate Authority, CA) сервер сертификатын алу керек. Тестілеу үшін көбінесе өзбетімен жасалған сертификат дайындалады, оны сервер сертификатына қол қою үшін пайдалануға болады.

Зерттеу барысында Mosquitto MQTT брокері, Paho MQTT клиенті, OpenSSL утилитасы, iPerf құралы қолданылды.

Материалдар мен әдістер. 1.MQTT хаттамасы. MQTT (ағылш. message queuing telemetry transport) - TCP/IP үстінен жұмыс істейтін, жариялаушы-жазылушы принципі бойынша құрылғылар арасында хабарламалар алмасуға бағытталған жеңілдетілген желілік хаттама. MQTT немесе телеметрия хабарламалары кезегін тасымалдау хаттамасын 1999 жылы Энди Стэнфорд-Кларк (IBM) және Арлен Ниппер (Arcom, қазіргі Eurotech) ойлап тапты. 15 жылдан соң, яғни, 2014 жылы OASIS консорциумы MQTT 3.1.1 нұсқасын стандарттады және ол қазір ашық ISO (ISO / IEC 20922: 2016) стандарты болып табылады (Cornel – Cristian т.б., 2019). Бұл хаттама ең алғашқы рет мұнай құбырларын спутниктік қосылыстар арқылы қосу үшін дайындалды және оның жобалық параметрлері батареяны минималды шығындауды және өткізу жолағын минималды пайдалануды қамтыды (Katsikeas т.б., 2017) MQTT хаттамасының үстеме шығындары төмен, яғни ол қосымша деректердің аз көлемі мен нақты хабарлама мазмұнын ғана жібереді. MQTT тақырыпшасы HTTP немесе CoAP (Azzawi т.б., 2016.) сияқты басқа хаттамалармен салыстырғанда өте кішкентай (небәрі 2 байт) (Yassein т.б., 2017).

MQTT жариялау/жазылу архитектурасын қолдайды. MQTT хаттамасында екі клиент бар, бірі – жариялаушы (Publisher), екіншісі – жазылушы (Subscriber). Жариялаушы-клиент хабарламаны немесе пайдалы жүктемені (Payload) белгілі бір тақырыпта (Topic) жариялайды, ал жазылушы-клиент нақты тақырыпқа жазылып, хабарламаны алады да, сәйкес әрекеттерді орындайды. Тақырыптар – деректер тасымалданатын жолды анықтау үшін пайдаланылатын виртуалды арналар. Тақырыптың максималды ұзындығы MQTT сипаттамаларына сәйкес 65535 байтқа дейін болуы мүмкін. Тақырыптарды жариялаушылар жасайды және брокерге хабарламамен бірге жіберіледі, содан кейін жазылушы сол тақырыптарға жазыла алады.

MQTT хаттамасы арқылы IoT құрылғылары брокермен байланыса алады. IOT құрылғысын жариялаушы ретінде де, жазылушы ретінде де немесе екеуі ретінде MQTT брокерімен баптауға болады. MQTT хабарламаға бағытталған хаттама, онда клиенттер деректерді белгілі бір тақырыппен хабарлама ретінде жібереді және қабылдайды. 2-суретте MQTT хаттамасы арқылы «жариялау-жазылу» байланыс үлгісі көрсетілген. Бұл үлгіде брокер байланысудың орталық нүктесі болып табылады және ол арқылы клиенттер хабарламалармен алмасады.



2-сурет. Жариялау/жазылу байланыс үлгісі

Жариялаушы хабарламаларды жариялайды, ал жазылушы өзіне қатысты белгілі бір тақырыптарға жазылады және сол тақырыптар бойынша жарияланған әрбір хабарламаны алып отырады. Жарияланған хабарда брокер үшін бағыттау ақпаратын қамтитын тақырып бар. Әрбір хабарламаның тақырыбы бар және клиенттер бірнеше тақырыптарға жазыла алады. Брокер тақырыптарды орналастырады және жазылушыларға хабарламаларды қайта жіберуге жауапты. Белгілі бір тақырыпқа қатысты хабарларды алуға мүдделі кез келген клиент сол тақырыпқа жазылуы керек. Бұл тақырыптар компьютердегі файл жолдарына ұқсас иерархиялық жүйеде категорияларға бөлінген, мысалы, «Пәтер/қонақ_бөлме/кондиционер/күй» (Triawan т.б., 2016).

Пайдалы жүктеме – жариялаушы-клиент жібергісі келетін нақты хабарлама болып табылады. MQTT хаттамасы арқылы жіберуге болатын максималды пайдалы жүктеме 268, 435, 456 байтпен шектелген, бұл пакетке ең көбі 256 МБ болуы мүмкін. Ең үлкен өлшем MQTT спецификацияларымен шектелген.

Құрылғылар MQTT брокермен байланысу үшін белгілі бір хабарлама типтерін пайдаланады. Брокерден жіберілетін және алынатын хабарламалардың 14 типі бар. Негізгі хабарламаларға клиенттерді брокерге қосатын CONNECT/CONNACK, клиентке тақырыпқа жазылуға мүмкіндік беретін SUBSCRIBE/SUBACK және тақырып деректерін жариялаушыдан брокерге немесе брокерден жазылушыға жіберуге мүмкіндік беретін PUBLISH/PUBACK кіреді.

Әртүрлі IoT хаттамаларының арасында MQTT хаттамасының қызмет көрсету сапасы (QoS) оны бірегей етеді, және де хабарлардың жеткізілуі мен екі тарап арасында деректердің таралуына кепілдік береді (Kenzhebayeva т.б., 2021). MQTT хаттамасы байланыс үшін қызмет көрсету сапасының 3 түрін пайдаланады. MQTT қызмет көрсету

сапасы жариялаушы мен жазылушы арасында мәліметтердің түсуін растау туралы түсініктікті қамтамасыз етеді.

– QoS 0 – хабарлама бір рет ғана жіберіледі. Егер жеткізу үзілсе, онда хабарлама жоғалуы мүмкін – қайта жіберу болмайды.

– QoS 1 – хабарлама кемінде бір рет жіберіледі және алушы жеткізілгенін растайды. Бұл жағдайда хабарламаларды жіберу қайталануы мүмкін.

– QoS 2 – хабарлама мәселелер мен кедергілерге қарамастан хабарлама бір рет жеткізіледі. Сәтсіздікке байланысты жеткізу кешіктірілуі мүмкін, бірақ хабарлама бәрібір адресатқа жеткізіледі, мысалы, байланыс қалпына келтірілгеннен кейін.

2. MQTT хаттамасындағы қауіпсіздік механизмдері. MQTT мақсаты – Заттар интернеті үшін қолдануға жеңіл әрі қарапайым байланыс хаттамасын ұсыну. Хаттаманың өзінде тек бірнеше қауіпсіздік механизмдері анықталған. MQTT хаттамасында желі, транспорттық және қосымша деңгейлерінде қосылған қауіпсіздік функциялары бар. Әрбір деңгей әртүрлі шабуылдардан қорғайды.

Желілік деңгейде хаттама IoT құрылғыларын шлюз арқылы қосуды, содан кейін шлюзді VPN арқылы брокерге қосуды ұсынады. Құрылғылар Wi-Fi, Zigbee, Bluetooth және т.б. сияқты қауіпсіз физикалық деңгей хаттамалары арқылы шлюзге қосылуы керек.

Транспорттық деңгейде хаттама TLS/SSL қолданады. MQTT спецификациясында көрсетілгендей, TLS/SSL – стандартты транспорттық деңгей хаттамасы болып табылады, ол пакеттерді шифрлауға мүмкіндік береді, сонымен қатар сәйкесінше клиент пен сервер сертификаттарын пайдаланып клиент пен сервердің аутентификациясын қамтамасыз етеді.

Қолданба деңгейінде MQTT хаттамасы клиенттің аутентификациясы мен деректерді шифрлауды қамтамасыз етеді. Клиентті аутентификациялау әдетте брокер жағында клиент идентификаторы және пайдаланушы аты/құпия сөзі сияқты тіркелгі деректерінің көмегімен орындалады. MQTT хаттамасын қолдану кезінде транспорттық деңгейдегі толық шифрлауды немесе жай ғана пайдалы жүктемені шифрлауды орындауға болады.

MQTT спецификациясына сәйкес, MQTT хаттамасында келесі қауіпсіздік функциялары қарастырылған:

1) Аутентификация – бұл құрылғылар брокерге қосылып, басқа құрылғылармен әрекеттескенде түпнұсқалығын тексеру процесі. MQTT аутентификациясы пайдаланушы аты мен құпия сөз арқылы жүзеге асырылады. Аутентификацияға келгенде, MQTT хаттаманың

өзі CONNECT хабарында пайдаланушы аты мен құпия сөзі өрістерін береді. Сондықтан MQTT брокеріне қосылу кезінде клиентке пайдаланушы аты мен құпия сөзін жіберу мүмкіндігі болады. Клиентте пайдаланушы аты мен құпия сөзді орнатқаннан кейін олар брокерге мәтіндік форматта жіберіледі. Бұл шабуылдаушыға тыңдауға мүмкіндік береді. Осылайша, пайдаланушы аты мен құпия сөздің толық қауіпсіз берілуіне кепілдік берудің жалғыз жолы транспорттық шифрлауды пайдалану болып табылады.

Хаттама әрбір клиентке бірегей клиент идентификаторын беру арқылы кеңейтілген аутентификацияны қамтамасыз етеді. Бірегей клиент идентификаторын клиент MQTT CONNECT хабарламасында пайдаланады. Бірегей клиент идентификаторы ретінде 36 таңбадан тұратын UID кодын немесе желілік модульдің MAC мекен-жайы немесе құрылғының сериялық номері сияқты кез келген басқа бірегей ақпаратты пайдалануға болады.

2) Авторизация – ресурстарға қол жеткізу құқықтарын көрсету функциясы. Брокерге қосылғаннан кейін MQTT клиенті екі түрлі әрекет жасай алады: хабарламаларды жариялау және тақырыптарға жазылу. Клиентке ол рұқсат етілген тақырыптарды ғана жариялауға немесе жазылуға рұқсат беру үшін брокер жағында тақырыпқа рұқсат беруді жүзеге асыру қажет. Бұл рұқсаттар брокердің орындалуы кезінде конфигурацияланатын және реттелетін болуы керек. Тақырыпқа рұқсат берілу келесідей болуы мүмкін:

− Рұқсат етілген тақырып (нақты тақырып немесе қойылмалы таңбаларымен тақырып)

− Рұқсат етілген операциялар (жариялау, жазылу, екеуі де)

− Мүмкін қызмет көрсету деңгейі (0, 1, 2, барлығы)

Тақырыпқа рұқсат берудің бұл типі брокерге клиенттер үшін авторизация саясаттарын көрсетуге және олардың тақырыптарға жазылу және хабарларды жариялау мүмкіндігін шектеуге мүмкіндік береді.

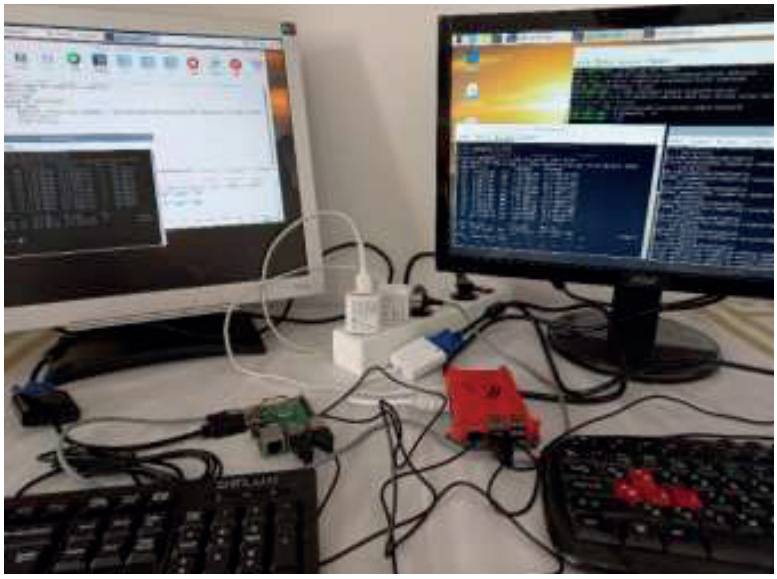
3) TLS/SSL қауіпсіздігі. MQTT хаттамасы үнсіз келісім бойынша шифрланған байланысты пайдаланбайтын TCP тасымалдау хаттамасына сүйенеді. Шифрлау мәселесін шешу үшін көптеген брокерлер қарапайым TCP орнына TLS пайдаланады, яғни TLS тасымалдау деңгейінде қауіпсіздікті қамтамасыз етеді. TLS көмегімен деректер пакеттерін шифрлау арқылы олардың мазмұнын үшінші тараптар оқуы немесе өзгертуі мүмкін емес. Қауіпсіз TLS/SSL MQTT байланысы үшін 8883 порты стандартталған. Көптеген брокерлер (мысалы, Mosquitto) TLS қауіпсіздігін қолдайды. MQTT TLS қауіпсіздік жүйесінде әрбір пакет шифрланатындықтан, бұл байланыс немесе қол алысудың

үстеме шығындарын арттырады. Ал шектеулі IoT құрылғылары үшін бұл мүмкін емес (Heer т.б., 2011). Әрине, егер IOT құрылғылары байланысқа кететін үстеме шығындарды көтере алатын болса, TLS әрқашан қолданылуы керек. Әзірлеушілер тасымалдау деңгейін толық шифрлау үшін TLS пайдаланады.

4) Пайдалы жүктемені шифрлау. MQTT хаттамасында пайдалы жүктемені шифрлау қосымша деңгейіндегі қауіпсіздікті қамтамасыз етеді. Әзірлеуші TLS қауіпсіздігін қолданбаған кезде, дегенмен, деректерді кәдімгі мәтін түрінде жібергісі келмесе, пайдалы жүктемені шифрлауды қолданады. Бұл қосымша қауіпсіздік деңгейін қамтамасыз етеді, себебі осылайша қосымшаның барлық деректері қорғалады және шифрланады. Пайдалы жүктемені шифрлау тасымалдау деңгейінде толық шифрлаудың орнына қосымшаның нақты деректерін шифрлауға мүмкіндік береді. Қосымшаның метадеректері (тақырып) шифрланбаған күйінде қалады. MQTT хаттамасында пайдалы жүктемені шифрлау кезінде PUBLISH пакеттерінің бөлігі ретінде пайдалы жүктемені шифрлауға болады. Пайдалы жүктеме әрқашан жариялаушы жағында шифрланады, ал дешифрлау процесі үздіксіз шифрлау жағдайында жазылушыда орындалуы мүмкін, бірақ оны брокерде де жасауға болады. Кез келген баптау барысында асимметриялық немесе симметриялық шифрлау схемаларын орнатуға болады. Сонымен қатар, кез келген сценарийде тек пайдалы жүктеме деректері шифрланады, ал барлық басқа деректер (клиент ақпараты, сертификаттар, тақырып туралы ақпарат) шифрланбаған күйде қалады. Қандай шифрлау/дешифрлау схемасы қолданылатынына байланысты, схеманың өзі көлемді ресурстарды қажет етеді, сондықтан бұл қайтадан шектеулі IoT құрылғылары үшін мәселе тудыруы мүмкін. Сондай-ақ барлық MQTT клиенттеріне кілттерді қауіпсіз жеткізу қажет болады. Жалпы, бұл шешімдер ортадағы адам шабуылдарына немесе қайталап ойнату шабуылдарының жолын кеспейді.

3. IoT құрылғылары арасында ақпарат алмасуды жүзеге асыру. IoT құрылғылары ретінде Raspberry Pi 3 model B микрокомпьютерлері қолданылды. Олар келесі сипаттамаларға ие: CPU (орталық процессор) – ARM Cortex-A53, процессор жиілігі - 1,2 ГГц, RAM (оперативті жады) – 1 ГБ.

Олардың бірі брокер және жазылушы ретінде, екіншісі жариялаушы ретінде бапталды. 4-суретте IoT құрылғылары арасында ақпарат алмасу процесі көрсетілген.



3-сурет. IoT құрылғылары арасында ақпарат алмасу процесі

Брокер ретіндегі Raspberry Pi 3 model B микрокомпьютеріне MQTT брокері Eclipse Mosquitto орнатылды. Mosquitto ресми веб-сайтындағы мәліметтерге сәйкес, Eclipse Mosquitto – Mosquito MQTT хаттамасын жүзеге асыратын, ашық бастапқы коды бар хабарлама брокері, жеңіл және қуаттылығы төмен бір тақталы компьютерлерден бастап толық серверлерге дейін барлық құрылғыларда пайдалануға жарамды.

Біз Mosquitto MQTT брокерін үш түрлі баптадық:

Бірінші қадамда қосылымды тексеру үшін қарапайым MQTT болды. Қарапайым MQTT-де шифрлау да немесе аутентификациялау да жоқ. Брокерді таба алатын кез келген нысан оған қосылып, оны қадағалай алады немесе оған хабарламалар жібере алады. Бұл қосылу негізгі MQTT қызметінің жұмыс істеп тұрғанын тексеру үшін уақытша, алдын ала орындалатын қадам болуы керек.

Екіншісі қадам логинді қажет ететін қарапайым, бірақ шифрланбаған аутентификация болды. MQTT қосылу кезінде пайдаланушы аты мен құпия сөзді талап ететіндей бапталды.

Үшінші қадам TLS шифрлауын және негізгі аутентификацияны қосу болды. TLS тек брокер үшін ғана емес, әрбір клиент үшін x509 сертификаттарын талап етеді, өйткені MQTT өзара аутентификацияны қажет етеді. Яғни, клиент серверге аутентификация жасайды, ал сервер клиентке аутентификация жасайды. Бұл қадамда жай ғана бөліктерді жеке сынауға болмайды және әрбір бөлік бір-біріне сәйкес келуі және

тұтас жұмыс істеуі үшін дұрыс болуы керек. X.509 сертификаты - бұл пайдаланушы немесе құрылғы туралы ақпаратты және олардың сәйкес ашық кілтін қамтитын стандартты өрістер жиынтығы. X.509 стандарты сертификатқа қандай ақпарат кіретінін және оның кодталуын (деректер пішімі) анықтайды (Forsby т.б., 2018).

Осы үш қадам бойынша бапталған MQTT брокері мен екінші IoT құрылғысы (жариялаушы-клиент) арасында ақпарат алмасуды орындадық және сәйкесінше әрбір қадам бойынша желінің өткізу қабілеті өлшенді. 4-суретте желінің өткізу қабілетін өлшеу процесінен үзінді көрсетілген.

```
Connecting to host 192.168.43.200, port 5201
[  5] local 192.168.43.89 port 46338 connected to 192.168.43.200 port 5201
[ ID] Interval           Transfer     Bitrate         Retr  Cwnd
[  5]   0.00-1.00   sec  1.70 MBytes  14.2 Mbits/sec    0    129 KBytes
[  5]   1.00-2.00   sec  1.80 MBytes  15.1 Mbits/sec    0    212 KBytes
[  5]   2.00-3.00   sec  1.74 MBytes  14.6 Mbits/sec    0    283 KBytes
[  5]   3.00-4.00   sec  1.43 MBytes  12.0 Mbits/sec    0    358 KBytes
[  5]   4.00-5.00   sec  1.68 MBytes  14.1 Mbits/sec    0    437 KBytes
[  5]   5.00-6.00   sec  1.49 MBytes  12.5 Mbits/sec    1    430 KBytes
[  5]   6.00-7.00   sec  1.43 MBytes  12.0 Mbits/sec   26    380 KBytes
```

4-сурет. Желінің өткізу қабілетін өлшеу процесінен үзінді

Желінің өткізу қабілетін өлшеу Iperf утилитасының көмегімен жүзеге асырылды. Iperf – желі қосылымының өнімділігін және екі құрылғы арасындағы деректерді берудің максималды жылдамдығын өлшеуге көмектесетін қарапайым және ыңғайлы желілік утилита.

Нәтижелер. Алдыңғы бөлімде IoT құрылғылары арасында ақпарат алмасуды жүзеге асырудың үш түрлі қадамын сипаттаған болатынбыз. Әр қадамға толығырақ тоқталып өтейік.

Бірінші қадам бойынша MQTT брокеріне клиент қарапайым қосылу жүзеге асырылды. Ол үшін Mosquito MQTT брокері іске қосылған соң жазылушы-клиент нақты тақырыпқа жазылды. Бірақ осы тақырып бойынша жариялаушы-клиент хабарлама жарияламайынша жазылушы-клиент күту күйінде болады. mosquitto_sub және mosquitto_pub командалары сәйкесінше тақырыпқа жазылуды және жариялауды жүзеге асырады. Бірінші қадам барысында 30 байт пайдалы жүктеме әрбір 3 секунд сайын жариялаушы-клиенттен брокерге жөнелтілді, және жазылушы-клиент терминалында көрініс тапты. Iperf құралымен 10 секунд бойы (үнсіз келісім бойынша) өлшеу 50 рет орындалды. Нәтижесінде орта есеппен желінің өткізу қабілеті - 13,4 мбит/с құрады.

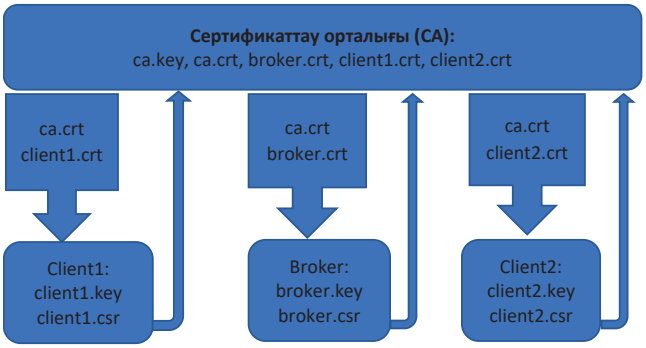
Екінші қадам бойынша MQTT брокеріне пайдаланушы аты мен

құпия сөзді талап ететін аутентификацияны орындау арқылы қосылу жүзеге асырылды. Жалпы жұмыс істеп тұрған MQTT брокеріне сервердің IP-адресін білетін кез-келген нысан қосыла алады. Бұл қауіпсіздік мәселесін шешу үшін хабарлама жіберуге рұқсат алған пайдаланушылардың аты мен құпия сөздерін көрсету керек. Mosquitto брокері қосылатын пайдаланушылар үшін құпия сөздерді шифрлау құралын қосады. sudo mosquitto_passwd -c /etc/mosquitto/passwd User1 командасы User1 пайдаланушыны қосады, осыдан соң екі рет құпия сөз енгізу ұсынылады. Пайдаланушыларды және олардың құпия сөздерін енгізген соң, Mosquitto брокерінің конфигурация файлына allow_ anonymous false және password_file /etc/mosquitto/passwd жолдарын енгізу керек. Осыдан кейін жариялаушы-клиент те, жазылушы-клиент те жүйеге ену үшін пайдаланушы атауы мен құпия сөзді енгізуі қажет болады. Екінші қадам барысында да 30 байт пайдалы жүктеме әрбір 3 секунд сайын жариялаушы-клиенттен брокерге жөнелтілді, және жазылушы-клиент терминалында көрініс тапты. Iperf құралымен 10 секунд бойы өлшеу жүргізілді. Бұл сынама да 50 рет орындалды. Нәтижесінде орта есеппен желінің өткізу қабілеті - 10,4 мбит/с құрады.

Үшінші қадам бойынша mosquitto MQTT брокерін TLS қауіпсіздігін қолдану үшін баптадық.

Біз openssl қосымшасын жеке сертификаттау орталығын (CA), сервер кілттері мен сертификаттарды жасау үшін қолдандық. 5-суретте Openssl құралымен сертификаттарды дайындау схемасы көрсетілген. Ең алдымен сертификаттау орталығының кілті (ca.key) дайындалады және осы кілт арқылы сертификаттау орталығының сертификаты (ca. crt) дайындалады. Сертификаттау орталығының кілті (ca.key) құпия сөзбен қорғалады. Осыдан соң брокердің және клиенттердің кілттері (broker.key, client1.key, client2.key) жеке-жеке дайындалады да, осы кілттердің негізінде сертификаттау орталығынан сертификаттар дайындауға сұраныстар (broker.csr, client1.csr, client2.csr) жеке-жеке дайындалып, сертификаттау орталығына жөнелтіледі. Сертификаттау орталығы осы сұраныстар (broker.csr, client1.csr, client2.csr) және өзінің кілті (ca.key) негізінде брокер мен клиенттерге сертификаттар (broker. crt, client1.crt, client2.crt) дайындап, жібереді. Әрбір сертификатпен қосып, сертификаттау орталығының сертификаты да (ca.crt) брокер мен клиенттердің қоймаларында сақталуы тиіс. Осылайша openssl құралының көмегімен біз TLS/SSL қосылысына қажетті сертификаттар жиынтығын дайындап, брокер мен әрбір клиенттердің қоймаларына орналастырдық.

5-сурет. Openssl құралымен сертификаттарды дайындау.



Осыдан кейін Mosquitto брокерінің конфигурация файлына қажетті ақпараттарды енгіземіз, яғни сертификаттардың орналасуы, порттар, сертификаттарды сұрау параметрлерін орнатамыз. Брокерді жұмысқа дайындаған соң mosquitto_sub және mosquitto_pub командалары арқылы, сертификаттардың орналасуын көрсете отырып, сәйкесінше тақырыпқа жазылуды және жариялауды жүзеге асырадық. Үшінші қадам барысында да 30 байт пайдалы жүктеме әрбір 3 секунд сайын жариялаушы-клиенттен брокерге жөнелтілді, және жазылушы-клиент терминалында көрініс тапты. Iperf құралымен 10 секунд бойы өлшеу 50 рет орындалды. Нәтижесінде орта есеппен желінің өткізу қабілеті - 8,48 мбит/с құрады.

Талқылау. Сынақтамада екі IoT құрылғысы (Raspberry Pi 3 model B микрокомпьютерлеі) арасында MQTT хаттамасы арқылы ақпарат алмасу процесі қауіпсіздік тұрғысынан зерттелді. Олардың бірі брокер және жазылушы-клиент ретінде, екіншісі жариялаушы-клиент ретінде бапталды. MQTT брокері ретінде Eclipse Mosquitto (https://mosquitto.org/download/) орнатылды. MQTT клиенті Mosquitto серверіне үш жолмен қосылды:

– қорғалмаған арна арқылы;

– пайдаланушы аты және құпия сөзді қолданумен қорғалған арна арқылы;

– TLS хаттамасымен қорғалған арна арқылы;

Сынақтамада өлшенген осы үш арнаның өткізу қабілеті 1-кестеде және 6-суретте көрсетілгендей болды.

1-кесте. Арналардың өткізу қабілеті

№	Арналар	Өткізу қабілеті, мбит/с
1	қорғалмаған арна арқылы;	13,4

2	пайдаланушы аты және құпия сөзді қолданумен қорғалған арна арқылы;	10,4
3	TLS хаттамасымен қорғалған арна арқылы;	8,48

6-сурет. Арналардың өткізу қабілеті.

Сынақтама жоғарыда айтылған қосылымдардың әрқайсысынан 50 рет орындалды. Нәтижесінде TLS хаттамасымен қорғалған арна арқылы қосылу желінің өткізу қабілетін едәуір азайтатыны көрінді. Демек IoT құрылғылары үшін жаңа жеңілдетілген аутентификация схемасы қажет.

Қорытынды. Соңғы кездері MQTT хаттамасының IoT қосымшалары үшін қолайлы, ең жақсы нұсқа болып табылатындығы белгілі болды. Сондықтан, MQTT хаттамасын қолданып IoT қосымшалары үшін жақсы шешімдер алу мақсатында көптеген зерттеулер жүргізілуде (Soni т.б., 2017). Осы зерттеулердің бір бағыты MQTT хаттамасының негізінде орындалған қосымшалар үшін жүзеге асырылуы тиіс қауіпсіздік шаралары. Ал MQTT хаттамасында тиімді қауіпсіздік функциялары жетіспейді, сондықтан тасымалдау деңгейнде SSL/TLS қолданылады. TLS жақсы қауіпсіз нұсқа болғанымен, үшінші тарапты жұмылдыру, сертификаттарды энергияға тәуелсіз жадыда сақтау және т.б. ресурсы шектеулі IoT құрылғылары үшін қосымша талаптар қояды. Сондықтан біздің мақсатымыз - болашақта MQTT хаттамасына негізделген заттар интернеті (IoT) үшін жаңа жеңілдетілген аутентификация механизмін ұсыну. Осы мақалада MQTT хаттамасының қауіпсіздік механизмдері зерттелді. MQTT хаттамасының спецификациясына сәйкес онда қарастырылған қауіпсіздік шаралары IoT құрылғысында (Raspberry pi) жүзеге асырылды және нәтижесі өлшенді. Сынақтама нәтижесінде қорғалмаған арна арқылы ақпарат алмасумен салыстырғанда пайдаланушы аты және құпия сөзді қолданумен қорғалған арна арқылы және TLS хаттамасымен қорғалған арна арқылы ақпарат алмасу кезінде

арнаның өткізу қабілетіне айтарлықтай әсер ететіні белгілі болды. Осы зерттеудің нәтижесі біздің алдағы зерттеулерде пайдалы болады деп есептейміз.

Information about authors:

Kazhenova Zhanar – Doctoral Student, Department of Computing and Software, S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; zhkazhenova75@gmail.com; ORCID ID: https://orcid.org/0000-0001-9022-7169;

Kenzhebayeva Zhanat – Candidate of technical sciences, acting associate professor, Department of Computing and Software, S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; kenzhebayeva.zh@gmail.com; ORCID ID: https://orcid.org/0000-0002-1942-4474;

Aleksander Prudnik – Candidate of technical sciences, Associate Professor, Department of Engineering Psychology and Ergonomics, Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus; aleksander.prudnik@bsuir.by; ORCID ID: *https://orcid.org/0000-0002-6180-1819*.

ӘДЕБИЕТТЕР

Andy S., Rahardjo B. and Hanindhito B., “Attack scenarios and security analysis of MQTT communication protocol in IoT system,” 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Yogyakarta, 2017, pp. 1-6.

Azzawi MA., Hassan R., Bakar KAA., “A Review on Internet of Things (IoT) in Healthcare” International Journal of Applied Engineering Research, November 2016.

Bryce R., Srivastava G.: The addition of geolocation to sensor networks. In: ICSOFT. pp. 796–802. SciTePress (2018).

Cornel-Cristian A., Gabriel T., Arhip-Calin M., Zamfirescu A., “Smart home automation with MQTT” 54th International Universities Power Engineering Conference (UPEC), 2019.

Filip Forsby et al. “Lightweight X.509 Digital Certificates for the Internet of Things”. In: (2018). Ed. by Giancarlo Fortino et al., pp. 123–133.

Goyal K.K., Garg A., Rastogi A., Singhal S., “A Literature Survey on Internet of Things (IoT),” Int. J. Advanced Networking and Applications. Volume: 09 Issue: 06 Pages: 3663-3668, 2018.

Heer T., Garcia-Morchon, O., Hummen, R., Keoh, S.L., Kumar, S.S., Wehrle, K.: Security challenges in the ip-based internet of things. Wireless Personal Communications 61(3), 527–542 (2011).

Каженова Ж.С., Кенжебаева Ж.Е. Безопасность в протоколах и технологиях IoT: обзор. International Journal of Open Information Technologies ISSN: 2307-8162 vol. 10, no. 3, 2022.

Katsikeas S., Fysarakis K., Miaoudakis A., Van Bemten A., Askoxylakis I., Papaefstathiou I., Plemenos A., "Lightweight secure industrial iot communications via the mq telemetry transport protocol," in Computers and Communications (ISCC), 2017 IEEE Symposium on, pp. 1193–1200, IEEE, 2017.

Kenzhebayeva Z., Akhmetova Z., Bainazarova R., Kazhenova Z., Sariyeva A. Simplified and secure authentication scheme for the internet of things(article) Journal of Theoretical and Applied Information TechnologyVolume 99, Issue 24, 5 December 2021, Pages 5774-5782.

Soni D., Makwana A., "A Suever on Mqtt: A Protocol of Internet Of Things (IOT)" April 2017. Conference: International conference on telecommunication, power analysis and computing techniques (ICTPACT - 2017).

Triawan M.A., Hindersah H., Yolanda D., Hadiatna F. "Internet of things using publish and subscribe method cloud-based application to NFT-based hydroponic system" 6th International Conference on System Engineering and Technology (ICSET, 2016.

Yassein M.B., Shatnawi M.Q., Aljwarneh S., Al-Hatmi R., "Internet of Things: Survey and open issues of MQTT protocol" International Conference on Engineering & MIS (ICEMIS), 2017.

REFERENCES

Andy S., Rahardjo B. and Hanindhito B., "Attack scenarios and security analysis of MQTT communication protocol in IoT system," 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Yogyakarta, 2017, pp. 1-6.

Azzawi M.A., Hassan R., Bakar K.A.A., "A Review on Internet of Things (IoT) in Healthcare" International Journal of Applied Engineering Research, November 2016.

Bryce R., Srivastava G.: The addition of geolocation to sensor networks. In: ICSOFT. pp. 796 – 802. SciTePress (2018).

Cornel-Cristian A., Gabriel T., Arhip-Calin M., Zamfirescu A., "Smart home automation with MQTT" 54th International Universities Power Engineering Conference (UPEC), 2019.

Filip Forsby et al. "Lightweight X.509 Digital Certificates for the Internet of Things". In: (2018). Ed. by Giancarlo Fortino et al., pp. 123–133.

Goyal K.K., Garg A., Rastogi A., Singhal S., "A Literature Survey on Internet of Things (IoT)," Int. J. Advanced Networking and Applications. Volume: 09 Issue: 06 Pages: 3663-3668, 2018.

Heer T., Garcia-Morchon O., Hummen R., Keoh S.L., Kumar S.S., Wehrle K.: Security challenges in the ip-based internet of things. Wireless Personal Communications 61(3), 527–542 (2011).

Kazhenova Z.S., Kenzhebayeva Z.E., Security in IoT protocols and technologies: an overview. International Journal of Open Information Technologies ISSN: 2307-8162 vol. 10, no. 3, 2022.

Katsikeas S., Fysarakis K., Miaoudakis A., Van Bemten A., Askoxylakis I., Papaefstathiou I., Plemenos A., "Lightweight secure industrial iot communications via the mq telemetry transport protocol," in Computers and Communications (ISCC), 2017 IEEE Symposium on, pp. 1193–1200, IEEE, 2017.

Kenzhebayeva Z., Akhmetova Z., Bainazarova R., Kazhenova Z., Sariyeva A. Simplified and secure authentication scheme for the internet of things(article) Journal of Theoretical

and Applied Information TechnologyVolume 99, Issue 24, 5 December 2021, Pages 5774-5782.

Soni D., Makwana A., "A Suever on Mqtt: A Protocol of Internet Of Things (IOT)" April 2017. Conference: International conference on telecommunication, power analysis and computing techniques (ICTPACT - 2017).

Triawan M.A., Hindersah H., Yolanda D., Hadiatna F. "Internet of things using publish and subscribe method cloud-based application to NFT-based hydroponic system" 6th International Conference on System Engineering and Technology (ICSET, 2016.

Yassein M.B., Shatnawi M.Q., Aljwarneh S., Al-Hatmi R., "Internet of Things: Survey and open issues of MQTT protocol " International Conference on Engineering & MIS (ICEMIS), 2017.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 136-152
https://doi.org/10.32014/2022.2518-1726.143
МРНТИ 28.23.29
УДК 004.89

А.Ж. Картбаев[1], Г.С. Ыбытаева[2*], О.Ж. Мамырбаев[1], К.Ж. Мухсина[1], Б.Ж. Жумажанов[1]

[1]Институт информационных и вычислительных технологий, Казахстан, Алматы;

[2]Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Казахстан, Алматы.
E-mail: ybytayeva.galiya@gmail.com

МЕТОДЫ ФОРМАЛЬНОГО ПРЕДСТАВЛЕНИЯ СУЩНОСТЕЙ В КРИМИНАЛЬНЫХ НОВОСТЯХ ДЛЯ АВТОМАТИЧЕСКОГО ПОСТРОЕНИЯ ОНТОЛОГИИ ПРЕСТУПЛЕНИЙ

Аннотация. В данной работе мы изучаем методы формального представления сущностей в криминальных новостях в виде онтологии путем определения их свойств и взаимосвязей между ними. Таким образом, разработанная нами онтология может быть использована правоохранительными органами для отслеживания и предотвращения преступной деятельности. Сейчас доступные нам открытые данные из социальных сетей и газетные статьи могут дать много полезной информации о формах преступной деятельности в конкретном месте и личной информации подозреваемых. А также мы проводим анализ наших данных и определяем их сложность, реализуем основные функций нашей системы, проверяем, какие цели и задачи решаются нашей системой. В результате анализа были выделены ключевые классы представления знаний о предметной области, составляющие основную структуру разработанной онтологии, которые мы используем, как ее словарь. Таким образом, была построена вложенная иерархия классов онтологии, представляющая собой совокупность иерархии терминов. В данном проекте мы предлагаем прикладной метод

разработки онтологии преступлений, который сначала использует сбор текста и изображений из новостных статей, затем мы расширяем и обогащаем онтологию, используя соответствующую информацию из известных социальных сетей. Получение семантически обоснованной информации играет важную роль в данном контексте, так как помогает должностным лицам понимать и эффективно работать со своими онтологиями и, в частности, использовать информацию в оптимальных масштабах. После этого нами были предложены способы практического применения построенной предметной онтологии, сформулированы направления последующих исследований.

Ключевые слова: онтология, извлечение информации, семантический анализ, граф знаний, криминальные данные.

А.Ж. Картбаев[1], Г.С. Ыбытаева[2*], О.Ж. Мамырбаев[1], К.Ж. Мухсина[1], Б.Ж. Жумажанов[1]

[1]Ақпараттық және есептеуіш технологиялар институты, Қазақстан, Алматы;
[2]Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Қазақстан, Алматы.
E-mail: *ybytayeva.galiya@gmail.com*

ҚЫЛМЫСТЫҚ ОНТОЛОГИЯНЫ АВТОМАТТЫ ТҮРДЕ ҚҰРУ ҮШІН ҚЫЛМЫСТЫҚ ЖАҢАЛЫҚТАРДАҒЫ СУБЪЕКТІЛЕРДІ РЕСМИ ТҮРДЕ ҰСЫНУ ӘДІСТЕРІ

Аннотация. Бұл мақалада онтология түрінде қылмыстық жаңалықтардағы субъектілерді ресми түрде ұсыну әдістері қасиеттері мен қатынастарын анықтау арқылы зерттеледі. Бұл онтологияны құқық қорғау органдары қылмыстық әрекетті бақылау және алдын-алу үшін қолдана алады. Бүгінде әлеуметтік желілер мен газет мақалаларынан алынған ашық деректер белгілі бір жерде қылмыстық әрекеттің сипаты туралы және күдіктілердің жеке мәліметтері туралы көптеген пайдалы ақпарат бере алады. Сондай-ақ деректер талданып олардың күрделілігі анықталды, жүйедегі негізгі функциялар іске асырылып, қандай мақсаттар мен міндеттерді шешетіні тексерілді. Талдау нәтижесінде дамыған онтологияның негізгі құрылымы болып табылатын пәндік сала туралы білімді ұсынудың негізгі кластары

анықталды. Терминдердің иерархияларының жиынтығы болып табылатын онтология кластарының кірістірілген иерархиясы құрылды. Онтологияны дамыту әдісін ұсынылды. Ол өз кезегінде, қылмыстық жаңалықтар мақалаларынан мәтіндер мен суреттер жиынтығын пайдаланып, танымал әлеуметтік желілердегі тиісті ақпаратты қолдана отырып онтологияны кеңейтеді. Семантикалық бай ақпаратты алу маңызды рөл атқарады, өйткені бұл лауазымды тұлғаларға онтологияны түсінуге және онымен тиімді жұмыс істеуге, сонымен қатар ақпаратты оңтайлы масштабта пайдалануға көмектеседі.

Түйін сөздер: онтология; ақпарат алу; семантикалық талдау; білім графигі; қылмыстық деректер.

A.Zh. Kartbayev[1], G.S. Ybytayeva[2*], O.Zh. Mamyrbayev[1], K.Zh. Mukhsina[1], B. Zh. Zhumazhanov[1]

[1]Institute of Information and Computer Technologies, Kazakhstan, Almaty;
[2]Kazakh National Research Technical University named after K.I. Satpayev, Kazakhstan, Almaty.
E-mail: ybytayeva.galiya@gmail.com

METHODS FOR FORMAL REPRESENTATION OF ENTITIES IN CRIME NEWS FOR AUTOMATIC CRIME ONTOLOGY CONSTRUCTION

Abstract. In this paper, we study methods for formally representing entities in crime news in the form of an ontology by identifying their properties and relationships. The ontology we have developed can be used by law enforcement agencies to track and prevent criminal activity. Todays the open data available to us from social media and newspaper articles can provide a lot of useful information about the patterns of criminal activity in a particular location and personal information of suspects. Also we analyze data and determine its complexity, implement basic functions of our system, and check what goals and objectives our system solves. As a result of the analysis we identified the key classes of knowledge representation about the subject area, which represents basic structure of the developed ontology, which we use as its vocabulary. Then a nested hierarchy of ontology classes was built, which is a set of hierarchy of terms. We propose a method for developing an ontology that uses the collection of text and images from criminal news

articles, then we extend the ontology using relevant information from popular social networks. Extracting semantically rich information plays an important role, because it helps officials understand and work effectively with the ontology and, use the information at optimal scale. After that we suggested cases of practical application of the constructed subject ontology and formulated directions for further research.

Key words: ontology; information extraction; semantic analysis; knowledge graph; criminal data.

Введение. Часто в реальном мире существует огромное количество необработанных данных о преступлениях. Обнаружение знаний в сложных областях может стать проблемой для методов добычи данных, которые обычно ограничиваются представлениями данных, не имея возможности получить доступ к их контексту и значению. Анализировать такой огромный объем данных простым взглядом довольно обременительно, и даже если потратить достаточно времени на их понимание, это не всегда приводит к семантически корректным результатам. Мы выбрали исходные данные криминальных новостей за 2015-2018 года, которые мы сохранили в формате RDF. Следовательно, в нашем исследовании мы реализуем систему поиска с использованием языка запросов RDF, который является одновременно языком запросов и протоколом доступа к данным, используемым для извлечения семантической информации из RDF.

Наша система предоставляет пользовательский интерфейс, в котором сотрудник правоохранительных органов может ввести строку поиска, а базовый API помогает получить и отобразить информацию для сотрудника. Концептуальные структуры, определяющие базовую онтологию, имеют отношение к идее машинного понимания данных в Семантической паутине.

Онтология – это схемы метаданных, предоставляющие контролируемый словарь понятий, каждое из которых имеет четко определенную и обрабатываемую машиной семантику. Определяя общие теории домена, онтологий помогают людям и машинам общаться лаконично, поддерживая обмен семантикой, а не только синтаксисом.

Краткий обзор исследований. В области извлечения информации было проведено достаточно много исследований. Как упоминалось в исследовательской работе (Euzenat и др., 2004) о сходстве между двумя сущностями, определенными между двумя узлами категории X графа, следует двум принципам: оно зависит от рассматриваемой категории;

оно учитывает все признаки этой категории (например, свойства). Пара сущностей, сходство которой оценивается называется якорной парой сравнения, а все пары, которые вносят индивидуальный вклад в расчет общего сходства называются участниками. Агрегирование сходства всех вкладчиков происходит с помощью взвешенной суммы, которая помогает контролировать вклад каждого признака.

Был оценен широкий спектр метрик сходства строк, а также стратегии предварительной обработки строк, такие как удаление стоп-слов и учет синонимов в различных типах онтологий (Gruber, 1995). Представлен набор рекомендаций о том, когда использовать ту или иную метрику, и показать, что оптимальные метрики сходства строк могут сами по себе производить выравнивания, конкурентоспособные с современными подходами в системах выравнивания онтологий. В обзоре по методам сходства текстов (Ristoski и др., 2016) обсуждается несколько подходов для поиска сходства слов, таких как лексическое сходство, семантическое сходство и так далее. Сходство на основе знаний (Cheatham и др., 2013; Qian и др., 2004) – это сходство на основе семантики, которое определяет степень сходства между словами, используя информацию, полученную из семантических сетей. Популярные семантические сети – это «Word Net», а также Natural Language Toolkit (NLTK) для измерения сходства между словами на основе знаний. Сходство на основе знаний также обеспечивает сходство на основе родства слов.

Термин онтология имеет долгую историю в философии, в которой он относится к предмету существования. В контексте управления знаниями онтологией называют общее понимание некоторых областей, которое часто представляется как набор сущностей, отношений, функций, аксиом и экземпляров. Онтологии – это структурированные представления области человеческих знаний, которые состоят из классов, описателей характеристик сущностей в этой области и набора отношений между этими классами.

Методика и материалы. Мы представляем коллекцию из более чем тысячи эталонного набора данных, которые необходимы для преодоления трудности в создании больших графов знаний путем использования сходства сущностей. Эти наборы данных включают данные из собранных нами криминальных газетных новостей, и исследуют сходство, рассчитанное на основе последовательностей данных и их семантических взаимодействий. Наборы данных имеют разный размер и охватывают как минимум три различных вида на разных уровнях полноты аннотации. Для каждого набора данных мы

также делаем расчеты семантического сходства с использованием самых современных репрезентативных мер.

Электронные газеты все чаще читаются пользователями из любого места и в любое время. Газеты являются источником достоверной и своевременной информации. Например, газетные статьи содержат информацию о преступлениях, несчастных случаях, политике, культурных и спортивных событиях. Несмотря на то, что ценная информация доступна в человеко-читаемой форме в газетах и архивах, но программных систем, которые могут извлекать соответствующую информацию и представлять ее, мало, и это представляет значительный интерес для исследователей в области извлечения информации. Таким образом, данный проект направлен на удовлетворение потребности в технологиях извлечения и обобщения информации путем создания концептуальной онтологии информации, собранной из онлайновых статей и социальных медиа. Извлеченная информация из газетных статей формирует базовую онтологию. Релевантная информация из социальных сетей была использована для обогащения онтологии.

Онтология газет создается путем соскабливания текстовых и графических данных из новостных статей в Интернете. Данные были предварительно обработаны и токенизированы для извлечения тегов, которые должны быть представлены в онтологии. Для получения онтологии газет были выполнены следующие шаги. Во-первых, онлайновые новостные статьи были обработаны для извлечения текстовой и визуальной информации. Создается резюме каждой статьи, которое затем подвергается дальнейшей обработке. Для этого был использован инструмент под названием Beautifulsoup, библиотека python для извлечения данных. Наш код далее выполняет автоматическое резюмирование заданной статьи. Автоматическое обобщение – это термин, который относится к извлечению сути документа с помощью программного обеспечения. Основная идея заключается в создании подмножества набора извлеченных данных, которое включает наиболее информативные предложения.

Далее, каждое предложение в резюме, извлеченное с помощью Data Scraping, рассматривается как событие в онтологии. Эти события включаются в онтологию с помощью уникального маркера, который их идентифицирует. Каждое предложение в резюме подвергается тегированию части речи (POS). Для POS-тегирования мы используем инструментарий естественного языка (NLTK). NLTK – это python-фреймворк, используемый для реализации обработки естественного

языка в программах на Python. Он предоставляет простые в использовании интерфейсы к более чем 50 корпорациям и лексическим ресурсам, таким как WordNet, а также набор библиотек обработки текста для классификации, токенизации, стеблирования, тегирования, синтаксического анализа и семантических рассуждений, а также обертки для промышленных библиотек NLP.

В конце мы также распознаем меру сходства изображения с места события с событием. Если в статье нет подписи к какому-либо изображению, заголовок статьи обрабатывается так же, как и подписи. Сходство рассчитывается с помощью синсета (набора синонимов) в WordNet, который дает оценку, основанную на семантическом сходстве различных слов в подписи к изображению и предложении. Все сущности связаны со своим типом сущности. Вот как формируется онтология путем извлечения текста и изображений. Объединенная онтология состоит из данных газет и социальных сетей с событиями, относящимися к любому из событий онтологии газет. Мы решаем, является ли конкретное событие онтологии социальных сетей релевантным или нет, используя методологию, аналогичную предложенной в исследовательской работе по выравниванию онтологий. Мы сравниваем каждый кортеж, полученный после выполнения POS-тегирования на соскобленных данных из социальных сетей, с кортежами из газетных статей для вычисления балла сходства. Для этого каждому типу сущностей были присвоены веса.

$$\{w_{relation} = 0.1;\ w_{person} = 0.25;\ w_{location} = 0.1;\ w_{organisation} = 0.25;\ w_{image} = 0.3\}$$

Пусть w_i представляет собой вес сущности i, а Sim_i ее функцию сходства. Мы вычисляем общий балл сходства для конкретного события онтологии социальных сетей как

$$Sim_t = \sum Sim_i * w_i \qquad (1)$$

который суммируется по всем сущностям, связанным с этим событием, и где $\sum w_i = 1$.

Мы сравниваем Sim_t с пороговым значением. При Sim_t > threshold, событие должно быть добавлено в объединенную онтологию. Порог 0,5 был установлен методом проб и ошибок.

Функции сходства для каждого из атрибутов зависят от типа атрибута. Атрибуты person, location и organization должны иметь сходство строк

(поскольку они являются существительными), в то время как relation должен быть схож семантически (поскольку это глагол). Сходство изображений рассчитывается с помощью сопоставления признаков. Это бинарные функции, т.е. они возвращают 1, если атрибуты совпадают, и 0 – в противном случае.

Мы вычислим сходство строк с помощью косинусного сходства. Косинусное сходство – это мера сходства между двумя векторами пространства внутренних произведений, которая измеряет косинус угла между ними. Косинус угла в 0 градусов равен 1, а для любого другого угла он меньше 1. Если значение больше 0,7 мы заключаем, что строки похожи или одинаковы, и, следовательно, функция сходства возвращает 1.

Мы рассчитываем семантическое сходство двух атрибутов, используя алгоритм Wu&Palmer (Wu и др., 1994). Алгоритм Wu&Palmer рассчитывает родство, учитывая глубину двух синсетов в таксономии WordNet, а также глубину LCS (Least Common Subsumer), используя формулу,

$$\text{Similarity\_score} = (2*\text{Depth(LCS)})/(\text{depth(s1)} + \text{depth(s2)}) \quad (2)$$

Это означает, что 0 < балл сходства <= 1. Оценка никогда не может быть нулевой, потому что глубина LCS никогда не бывает нулевой (глубина корня таксономии равна единице). Оценка равна единице, если два входных понятия одинаковы. Если балл сходства >= 0,6, то слова семантически совпадают.

Общая структура онтологии остается такой же, как и у онтологии газеты. Объединенная онтология содержит все экземпляры онтологии газеты, а также новые атрибуты, добавленные из сопоставленных экземпляров онтологии социальных медиа.

Результаты. Онтологии могут быть использованы для описания реальных объектов посредством процесса семантического аннотирования: объекты связываются с классами онтологии, наиболее подходящими для их описания. Набор сущностей, аннотированных с помощью данной онтологии, составляет граф знаний (Ristoski и др., 2016). Имея такое структурированное представление реальности, можно вычислительно рассуждать над сущностями, упрощая процесс, который был бы гораздо более дорогим и трудоемким, если бы его выполнял человек.

Одной из задач, которая стала возможной благодаря разработке онтологии, является расчет семантического сходства между сущнос-

тями. Мера семантического сходства – это функция, которая, задавая два класса онтологии или два набора классов, описывающих двух людей, возвращает числовое значение, отражающее близость смысла между ними (Seco и др., 2004). На рисунке 1 показано, как преступления представлены их классами и как их можно сравнивать с помощью мер семантического сходства. Точная оценка сходства между парой сущностей зависит от того, насколько хорошо они аннотированы, как в отношении широты (т.е. включения аннотаций для всех аспектов сущности, которые могут быть описаны в области онтологии), так и глубины аннотаций (т.е. выбора наиболее специфических классов онтологии, которые лучше всего описывают сущность) (Минский, 1979; Чень и др., 1983).

Подходы, используемые для количественной оценки семантического сходства, можно различать в зависимости от того, какие сущности они собираются сравнивать: существуют подходы для сравнения двух классов в онтологии и подходы для сравнения двух индивидуумов, каждый из которых связан со своим собственным набором классов. При сравнении классов эти меры могут быть узловыми, то есть изучающими свойства каждого класса, или краевыми, основанными на расстоянии между классами. Однако меры на основе ребер основаны на предположении, что узлы и ребра равномерно распределены по онтологии, что в основном неверно для криминальных онтологий, что делает меры на основе узлов более надежными.

Для расчета семантического сходства для двух сущностей, каждый из которых описывается набором классов, могут использоваться как парные, так и групповые подходы. Парные подходы оценивают сходство между двумя сущностями путем объединения семантического сходства между их аннотируемыми классами. Групповые подходы используют векторные или графовые меры, которые обрабатывают аннотации, взятые вместе как единое целое. Для создания эталонных наборов данных используется одна мера семантического сходства, характерная для каждого подхода.

Matching average «MA» – парный подход, основанный на парной мере, в которой сходство между двумя классами соответствует их среднему значению схожести характеристик. В MA для эвристического расчета парного сходства рассматривается только класс с наилучшим соответствием для каждого класса в каждом наборе классов, описывающих сущностей (т.е. наиболее похожий), который нами задается следующим образом

$$MA(A,B) = \frac{\sum_{c_1 \in C_A} \max sim(c_1,c_2)}{2|C_A|} + \frac{\sum_{c_2 \in C_B} \max sim(c_1,c_2)}{2|C_B|} \quad (3)$$

где уравнение A и уравнение B – сущности, уравнение C – набор классов уравнения, которыми описывается каждая сущность, а $sim(c_1, c_2)$ – наибольшие значение сходства, найденные для уравнения класса c_1, c_2.

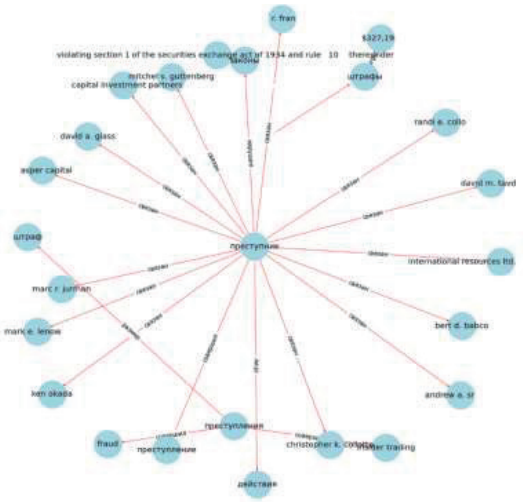


Рисунок 1 – Представление связей в онтологии, разбитые по классам и сущностям

Для создания этих эталонных наборов данных мы разработали общую методологию, разделенную на три этапа. Первый шаг состоит в выборе сущностей графа знаний, которые будут составлять пары в наборах данных. Эти сущности должны быть хорошо охарактеризованы в контексте онтологии, чтобы избежать предвзятости поверхностного аннотирования и иметь достаточно информации для вычисления сходства между ними. Следующий шаг – генерация пар сущностей. При этом мы старались гарантировать широкий диапазон сходства между парами сущностей, от нуля до полной идентичности, чтобы обеспечить репрезентативность пар сущностей. Наконец, отобрав пары сущностей для набора данных, необходимо рассчитать два типа мер сходства: семантическое сходство и приближенные показатели сходства, соответствующие типу сущностей и набору данных (Gruber, 2008).

После классификации методологий мы приступили к извлечению соответствующей информации из этих судебных разбирательств, такой как нарушения, нарушители, меры, принятые в отношении этих лиц, а

также наложенный штраф. Эти данные были сохранены в табличном формате. Вышеуказанные данные также были подготовлены для преобразования в граф знаний, который является вложенным по своей природе.

Для классификации преступлений было реализовано множество алгоритмов, позволяющих точно их идентифицировать. Были выбраны 4 основных класса, в которые были отнесены все документы, это – инсайдерская торговля, незаконное присвоение средств, незарегистрированные брокеры и мошенничество. Многие преступления относятся к нескольким классам, поэтому они были классифицированы соответствующим образом. Были опробованы три подхода. Первый подход заключался в использовании модели без наблюдения для классификации судебных релизов по различным классам. К сожалению, классификация была очень нестабильной и неточной. Кроме того, релизы не могли быть точно отнесены к нескольким классам. Этот подход был отброшен, и мы решили использовать контролируемую модель. Следующий подход заключался в использовании модели классификации текста под наблюдением для классификации релизов о судебных разбирательствах по различным категориям. Было замечено, что модель BERT не смогла правильно классифицировать документы по нужным классам. Модуль классификации текста BERT показал низкие результаты и имел точность 45,89%. Затем было решено увеличить размер обучающих данных, но точность снова осталась прежней. Причиной этого могло быть либо то, что обучающее и тестовое множество были недостаточно большими, либо то, что темы, по которым мы пытались классифицировать, были тесно связаны между собой, и различить их было сложно при недостаточном количестве данных. В конце концов мы решили прочесть достаточно значимое количество документов и выявили определенные закономерности в релизах судебных разбирательств. Это привело нас к использованию регулярных выражений (Kartbayev, 2016). Это был самый надежный подход, который оказался чрезвычайно точным. Релизы были успешно классифицированы на несколько классов с точностью 95% на том же наборе данных, который использовался для модели классификации тем BERT. В результате мы решили использовать regex-парсер для классификации релизов судебных разбирательств на различные преступления.

Для построения графа знаний мы попробовали несколько подходов к определению релевантных сущностей в корпусе. Наш первоначальный

подход состоял в том, чтобы определить субъект и объект документа и найти связь между ними (либо предикат, либо глагольное слово). Было замечено, что такое извлечение было не очень точным, и полученные результаты были неудовлетворительными. Это заставило нас улучшить алгоритм извлечения, и мы решили работать над извлечением SVO (Kartbayev и др., 2018; Kartbayev, 2015). Этот подход подразумевает идентификацию триплетных фраз. Мы определяли субъект и объект фразы, а связь между ними представляла собой глагольную фразу. Этот алгоритм извлечения показал себя значительно лучше, чем наш первоначальный подход. Однако в некоторых случаях связь между сущностями документа была потеряна. Этот недостаток побудил нас использовать концепцию онтологий для представления графа знаний. Мы решили, что поскольку эти документы имеют много сходств между собой, мы можем построить правила онтологии из имеющихся релизов судебных разбирательств. Наша онтология криминальных новостей на данный момент имеет 5 основных классов – Нарушитель, Нарушение, Преступление, Принятые меры, Штраф и Дата. Между различными классами были установлены связи, и это побудило нас использовать вложенную структуру графа знаний вместо стандартных триплетных связей. Затем мы классифицировали классы онтологии.

Полные наборы эталонных данных являются ключом к поиску наиболее эффективных инструментов для конкретного приложения. Существует ряд требований к хорошим эталонным наборам данных, а именно: актуальность, репрезентативность, отсутствие избыточности, масштабируемость и возможность повторного использования. В контексте этих эталонных наборов данных, мера семантического сходства означает, что наборы данных должны включать данные, релевантные для исследуемой области, иметь репрезентативные случаи как с точки зрения метрик сходства, так и их значений, или содержать как положительные, так и отрицательные примеры, чтобы сделать сравнительное исследование между ними более релевантным, должны поддерживать одно и то же исследование в наборах данных разного размера и имеют особую ценность, если они могут быть использованы для разных целей. Репрезентативность имеет особое значение для этих наборов данных, поскольку наборы данных должны обеспечивать сбалансированный срез криминальных сущностей. Коллекция из эталонного набора данных, которую мы представляем, направлена на поддержку крупномасштабной оценки мер семантического сходства на основе сходства графов. Она представляет собой эволюцию по

сравнению с предыдущими усилиями в этой области, как с точки зрения размера, так и разнообразия используемых данных (Akhmetov и др., 2022).

Обсуждение. Большой проблемой при оценке мер семантического сходства является разнообразие исследований, используемых для этого. Меры семантического сходства обычно тестируются на небольшом и контролируемом наборе данных, разработанном только для этого исследования. Такая несистематическая практика оценки может привести к смещению опубликованных результатов, особенно если не сравнивать их с результатами современных мер сходства в тех же условиях, т.е. с использованием точно такой же версии графа и тех же пар сущностей. Более того, отсутствие единой стратегии или, по крайней мере, одинаковых данных, делает результаты этих исследований несопоставимыми между собой.

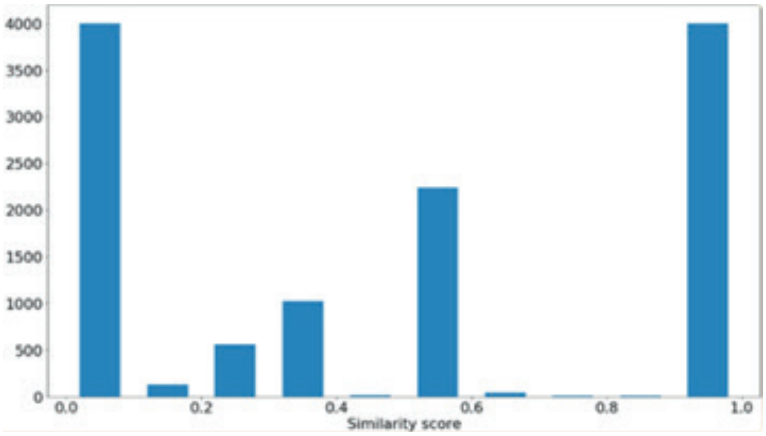


Рисунок 2 – Измерение сходства различных классов данных

Данная работа направлена на решение этих проблем путем предоставления наборов данных с парами сущностей разных видов, аннотированных различными онтологиями и обеспечивающих комбинацию различных проксисходства и нескольких современных мер семантического сходства. Чтобы гарантировать, что меры семантического сходства могут отразить функциональное сходство между сущностями, их значение должно быть хорошо передано в контексте онтологии (Рисунок 2). Это означало выбор сущностей, аннотированных более конкретными классами онтологии (классы с меньшим количеством дочерних классов), поскольку совместное использование одного или нескольких таких классов приведет к более

высокому и значительному семантическому сходству между двумя сущностями. Это было сделано для того, чтобы решить проблему неглубокого аннотирования для мер семантического сходства, в результате чего значения сходства не соответствуют человеческому восприятию из-за неглубоко описанных сущностей.

Кроме того, выбранные наборы данных соответствуют рекомендациям по качеству эталонных наборов данных, а именно: релевантность, репрезентативность, масштабируемость и возможность повторного использования. Хотя, мы предполагаем, что эталонные наборы данных должны быть нередуцируемыми, дублирование наборов данных одного и того же вида, но с разным уровнем заполнения аннотации, может быть использовано для оценки влияния более подробного описания преступлений на производительность мер семантического сходства. Наборы данных на каждом уровне полноты аннотации представляют собой компиляцию всех сущностей в каждом из классов данных.

Репрезентативность была особенно важной характеристикой при разработке этих наборов данных, например, при выборе наказаний, поскольку оценка мер семантического сходства должна проводиться как в сходных, так и в несходных парах сущностей. Кроме того, если эти наборы данных будут использоваться для приложений контролируемого обучения, то эти предикторы выиграют от обучения на более общем наборе данных. Если случаи, используемые для обучения, особенно предвзято относятся к одному признаку, производительность предиктора также будет предвзятой.

Разнообразие в структуре графа и мерах сходства, выбранных для построения этих наборов данных, позволяет предположить, что тестирование одного и того же семантического сходства в различных целевых наборах данных может быть хорошим показателем его способности к обобщению на различные графы, типы сущностей и их применения.

Заключение. В проекте реализован метод извлечения информации из новостных статей и страниц социальных сетей в Интернете и представления ее в интерактивной форме. Мы начинаем с введения различных важных понятий, связанных с извлечением информации и созданием онтологии. Затем мы предлагаем метод для достижения нашей цели. Основа нашего подхода состоит из сбора данных для анализа, затем извлечение сущностей и связей, и, наконец, визуализация информации. Каждый из этих основных этапов включает в себя множество шагов, которые мы объясняем в соответствующих разделах.

Использование данной онтологии было бы чрезвычайно полезно для правоохранительных органов и спецслужб для обнаружения и предотвращения радикализации, гражданских беспорядков и других антисоциальных действий в Интернете. Таким образом, в этом проекте мы обрабатываем множество информации из криминальных новостей, доступные в Интернете, и предлагаем метод разработки онтологии сущностей, и событий, которые связывают эти сущности, обеспечивая тем самым агрегированный обзор информации, представленной в многочисленных источниках.

Благодарность. *Работа выполнена при финансовой поддержке Комитета науки Министерства образования и науки Республики Казахстан (№AP09259309).*

Information about authors:

Kartbayev Amandyk Zhankozhauly – PhD, Institute of Information and Computational Technologies, Almaty, Kazakhstan, a.kartbayev@gmail.com; https://orcid.org/0000-0003-0592-5865;

Ybytayeva Galiya Seitkaliyevna – PhD student, specialty «Management information systems», Satbayev University, Almaty, Kazakhstan, ybytayeva.galiya@gmail.com; https://orcid.org/0000-0002-4243-0928;

Mamyrbayev Orken Zhumazhanovich – PhD, Institute of Information and Computational Technologies, Almaty, Kazakhstan, morkenj@mail.ru; https://orcid.org/0000-0002-8627-1949;

Mukhsina Kuralay Zhenisbekovna – PhD, Institute of Information and Computational Technologies, Almaty, Kazakhstan, kuka_ai@mail.ru; https://orcid.org/0000-0002-8627-1949;

Zhumazhanov Bagashar Zhumazhanovich – candidate of technical sciences, Institute of Information and Computational Technologies, Almaty, Kazakhstan, bagasharj@mail.ru; https://orcid.org/0000-0002-5035-9076.

ЛИТЕРАТУРА:

Akhmetov I., Gelbukh A., Mussabayev R. (2022) Topic-Aware Sentiment Analysis of News Articles. Computacion y Sistemas. – PP. 423-439. (in Eng.).

Euzenat J., Valtchev P. (2004) Similarity-based ontology alignment in OWL-Lite, Proc. 16th European conference on artificial intelligence (ECAI), Valencia, Spain. IOS press, – PP. 333-337. (in Eng.).

Gang Qian, Shamik Sural, Yuelong Gu, Sakti Pramanik. (2004) Similarity between euclidean and cosine angle distance for nearest neighbor queries, Proceedings of ACM Symposium on Applied Computing. – PP. 48-61. (in Eng.).

Gruber T. (1995) Toward principles for the design of ontologies used for knowledge sharing? Human-Computing Studies. – PP. 35-43. (in Eng.).

Gruber T. (2008) Collective knowledge systems: where the Social Web meets the Semantic Web. Journal of Web Semantics. – PP. 4-13. (in Eng.).

Kartbayev A. (2015) (Refining Kazakh Word Alignment Using Simulation Modeling Methods for Statistical Machine Translation. Lecture Notes in Computer Science, Springer. – PP. 421-427. (in Eng.).

Kartbayev A. (2016) Using Kazakh Morphology Information to Improve Word Alignment for SMT. Advances in Intelligent Systems and Computing, Springer. – PP. 351-359. (in Eng.).

Kartbayev A., Tukeyev U., Sheremeteva S., Kalizhanova A., Kalybek Uuly B. (2018) Experimental study of neural network-based Word alignment selection model trained with Fourier descriptors. Journal of Theoretical and Applied Information Technology. – PP. 4103-4113. (in Eng.).

Michelle Cheatham, Pascal Hitzler, Alani H. et al. (Eds.). (2013) String Similarity Metrics for Ontology Alignment, ISWC 2013. – PP. 263-285. (in Eng.).

Ristoski P., Paulheim H. (2016) Rdf2Vec: RDF graph embeddings for data mining. In: Groth P., Simperl E., Gray A., Sabou M., Krötzsch M., Lecue F., Flöck F., Gil Y., editors. The Semantic Web – ISWC 2016. Cham: Springer. – PP. 49-55. (in Eng.).

Ristoski P., Paulheim H. (2016) Semantic Web in data mining and knowledge discovery: A comprehensive survey. Journal of Web Semantics. – Vol. 36, PP. 12-22. (in Eng.).

Seco N., Veale T., Hayes J. (2004) An intrinsic information content metric for semantic similarity in WordNet. Proceedings of the 16th European Conference on Artificial Intelligence, ECAI'04. Amsterdam: IOS Press. – PP. 20-34. (in Eng.).

Wu Z. and Palmer M. (1994) Verb semantics and lexical selection. In Proceedings of the 32nd Annual meeting of the Associations for Computational Linguistics. – PP. 133-138. (in Eng.).

Минский М. (1979) Фреймы для представления знаний. – М.: Энергия. – 151 с.

Чень Ч. (1983) Математическая логика и автоматическое доказательство теорем, под ред. С.Ю. Маслова. – М.: Наука. – 360 с.

REFERENCES:

Akhmetov I., Gelbukh A., Mussabayev R. (2022) Topic-Aware Sentiment Analysis of News Articles. Computacion y Sistemas. – PP. 423-439. (in Eng.).

Chen Ch. (1983) Mathematical logic and automatic theorem proving, ed. by S.Y. Maslov. - Moscow: Nauka,. - 360 p. (in Rus).

Euzenat J., Valtchev P. (2004) Similarity-based ontology alignment in OWL-Lite, Proc. 16th European conference on artificial intelligence (ECAI), Valencia, Spain. IOS press, – PP. 333-337. (in Eng.).

Gang Qian, Shamik Sural, Yuelong Gu, Sakti Pramanik. (2004) Similarity between euclidean and cosine angle distance for nearest neighbor queries, Proceedings of ACM Symposium on Applied Computing. – PP. 48-61. (in Eng.).

Gruber T. (1995) Toward principles for the design of ontologies used for knowledge sharing? Human-Computing Studies. – PP. 35-43. (in Eng.).

Gruber T. (2008) Collective knowledge systems: where the Social Web meets the Semantic Web. Journal of Web Semantics. – PP. 4-13. (in Eng.).

Kartbayev A. (2015) (Refining Kazakh Word Alignment Using Simulation Modeling Methods for Statistical Machine Translation. Lecture Notes in Computer Science, Springer. – PP. 421-427. (in Eng.).

Kartbayev A. (2016) Using Kazakh Morphology Information to Improve Word Alignment for SMT. Advances in Intelligent Systems and Computing, Springer. – PP. 351-359. (in Eng.).

Kartbayev A., Tukeyev U., Sheremeteva S., Kalizhanova A., Kalybek Uuly B. (2018) Experimental study of neural network-based Word alignment selection model trained with Fourier descriptors. Journal of Theoretical and Applied Information Technology. – PP. 4103-4113. (in Eng.).

Michelle Cheatham, Pascal Hitzler, Alani H. et al. (Eds.). (2013) String Similarity Metrics for Ontology Alignment, ISWC 2013. – PP. 263-285. (in Eng.).

Minsky M. (1979) Frames for knowledge representation. - M.: Energia. - 151 p. (in Rus).

Ristoski P., Paulheim H. (2016) Rdf2Vec: RDF graph embeddings for data mining. In: Groth P., Simperl E., Gray A., Sabou M., Krötzsch M., Lecue F., Flöck F., Gil Y., editors. The Semantic Web – ISWC 2016. Cham: Springer. – PP. 49-55. (in Eng.).

Ristoski P., Paulheim H. (2016) Semantic Web in data mining and knowledge discovery: A comprehensive survey. Journal of Web Semantics. – Vol. 36, PP. 12-22. (in Eng.).

Seco N., Veale T., Hayes J. (2004) An intrinsic information content metric for semantic similarity in WordNet. Proceedings of the 16th European Conference on Artificial Intelligence, ECAI'04. Amsterdam: IOS Press. – PP. 20-34. (in Eng.).

Wu Z. and Palmer M. (1994) Verb semantics and lexical selection. In Proceedings of the 32nd Annual meeting of the Associations for Computational Linguistics. – PP. 133-138. (in Eng.).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
Volume 3, Number 343 (2022), 153-163
https://doi.org/10.32014/2022.2518-1726.144
IRSTI 20.20.19
UDC 004.943

A.T. Mazakova[1,2], K.B. Begaliyeva[2], T.Zh. Mazakov[1,2], Sh.A. Jomartova[2], G.Z. Ziyatbekova[1,2*]

[1]Institute of Information and Computational Technologies CS MES RK, Kazakhstan, Almaty;

[2]Al-Farabi Kazakh National University, Kazakhstan, Almaty.
E-mail: *ziyatbekova@mail.ru*

SOLUTION OF THE THERMAL CONDUCTIVITY EQUATION OF A ROD WITH A SQUARE SECTION BY CASTING TO A SYSTEM OF ORDINARY DIFFERENTIAL EQUATIONS

Abstract. The purpose of this paper is to study the thermophysical state of a rod of constant cross section and limited length. This work is devoted to automating the study of the thermophysical state of a rod of constant cross section and limited length. The research automation process is based on the laws of conservation of energy. A three-dimensional body is considered, the constant cross section of which has the shape of a square. It is assumed that the left end of the rod coincides with the origin of coordinates and the heat transfer coefficient is assumed to be constant over the entire surface of the rod. It is also assumed that the rod is subject to point temperature and surface heat transfer. The correctness of the problem under study is most often very difficult, and sometimes impossible. However, due to the complexity of the phenomena under study, solve analytically partial differential equations using modern mathematical methods. There are also many solution methods suitable for practical use, such as analytical, analog, numerical, graphical and experimental. Temperature is a parameter that characterizes the energy of the thermal motion of particles of a substance. Consequently, the process of heat propagation and its direction are inextricably linked with the temperature distribution inside the body. In this regard, the paper proposes a reduction of the heat equation to a system of ordinary differential equations.

The problem is solved by reducing to a system of linear ordinary differential equations, for the solution of which an appropriate algorithm has been developed. Therefore, the development of special methods and computational algorithms and a set of applied programs that make it possible to study the steady thermophysical state of rods of limited length under the simultaneous influence of heterogeneous types of heat sources is an urgent problem. A program has been developed for finding the temperature distribution along the rod, which places the results of numerical calculations in several files. The results of numerical calculations in dynamics (over time) are displayed in the form of a table and displayed in the form of one-dimensional graphs. They do not contradict the experimental data. A promising direction is the use of interval mathematics to study the heat equation.

Key words: thermal conductivity, thermal insulation, temperature, non-stationary thermophysical process, energy.

А.Т. Мазақова[1,2], Қ.Б. Бегалиева[2], Т.Ж. Мазаков[1,2], Ш.А. Жомартова[2], Г.З. Зиятбекова[1,2*]

[1]Ақпараттық және есептеуіш технологиялар институты, Қазақстан, Алматы;
[2]Әл-Фараби атындағы Қазақ ұлттық университеті, Казақстан, Алматы.
E-mail: *ziyatbekova@mail.ru*

КВАДРАТ ҚИМАСЫ БАР ӨЗЕКШЕНІҢ ЖЫЛУ ӨТКІЗГІШТІК ТЕҢДЕУІН ҚАРАПАЙЫМ ДИФФЕРЕНЦИАЛДЫҚ ТЕҢДЕУЛЕР ЖҮЙЕСІНЕ ҚОЮ АРҚЫЛЫ ШЕШУ

Аннотация. Бұл жұмыстың мақсаты – қимасы тұрақты және ұзындығы шектелген өзекшенің термофизикалық күйін зерттеу. Зерттеуді автоматтандыру процесі энергияның сақталу заңдарына негізделген. Үш өлшемді дене қарастырылады, оның тұрақты қимасы шаршы пішінді болып келеді. Өзекшенің сол жақ шеті бастапқы нүктемен сәйкес келеді деп болжанады және жылу беру коэффициенті өзекшенің бүкіл бетінде тұрақты деп қабылданады. Сондай-ақ, өзекше нүктелік температураға және беттік жылу алмасуға ұшырайды деп есептеледі. Зерттелетін есептің дұрыстығы көбінесе өте қиын жағдайда шешіледі. Дегенмен, зерттелетін құбылыстардың күрделілігіне байланысты қазіргі математикалық әдістерді қолдана отырып, аналитикалық дербес

дифференциалдық теңдеулерді шешуге тура келеді. Сонымен қатар аналитикалық, аналогтық, сандық, графикалық және эксперименттік сияқты практикалық қолдануға қолайлы көптеген шешу әдістері бар. Температура – зат бөлшектерінің жылулық қозғалысының энергиясын сипаттайтын параметр. Демек, жылудың таралу процесі және оның бағыты дене ішіндегі температураның таралуымен тығыз байланысты. Осыған байланысты аталмыш жұмыста жылу теңдеуін қарапайым дифференциалдық теңдеулер жүйесіне келтіру ұсынылған. Есеп сызықтық кәдімгі дифференциалдық теңдеулер жүйесіне келтіру арқылы шешіледі әрі оны шешу үшін де сәйкес алгоритм әзірленді. Сондықтан жылу көздерінің гетерогенді түрлерінің бір мезгілде шектелген ұзындықтарының әсерінен өзекшелердің тұрақты термофизикалық күйін зерттеуге мүмкіндік беретін арнайы әдістер мен есептеу алгоритмдерін және қолданбалы бағдарламалар кешенін жасау өзекті мәселе болып табылады. Температураның қарастырып отырған өзекше бойымен таралуын табуға арналған программа жасалды, ол сандық есептеулердің нәтижелерін бірнеше файлдарға орналастырады. Динамикадағы сандық есептеулердің нәтижелері (уақыт бойынша) кесте түрінде көрсетіледі және бір өлшемді график түрінде беріледі. Олар эксперименттік мәліметтерге қайшы келмейді. Жылуөткізгіштік теңдеуді зерттеу үшін аралық математиканы пайдалану перспективалы бағыт болып табылады.

Түйін сөздер: жылу өткізгіштік, жылу оқшаулау, температура, стационарлық емес термофизикалық процесс, энергия.

А.Т. Мазакова[1,2], К.Б. Бегалиева[2], Т.Ж. Мазаков[1,2], Ш.А. Жомартова[2], Г.З. Зиятбекова[1,2*]

[1]Институт информационных и вычислительных технологий, Казахстан, Алматы;
[2]Казахский национальный университет имени аль-Фараби, Казахстан, Алматы.
E-mail: *ziyatbekova@mail.ru*

РЕШЕНИЕ УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ СТЕРЖНЯ С КВАДРАТНЫМ СЕЧЕНИЕМ ПРИВИДЕНИЕМ К СИСТЕМЕ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Аннотация. Целью данной работы является исследование теплофизического состояния стержня постоянного сечения и ограниченной

длины. Данная работа посвящена автоматизации исследования теплофизического состояния стержня постоянного сечения и ограниченной длины. Процесс автоматизации исследования опирается на законы сохранения энергии. Рассматривается трехмерное тело, постоянное поперечное сечение которого имеет форму квадрата. Предполагается, что левый конец стержня совпадает с началом координат и коэффициент теплообмена считается постоянным по всей поверхности стержня. Также предполагается, что стержень находится под воздействием точечной температуры и поверхностного теплообмена. Корректность исследуемой задачи чаще всего бывает очень трудно, а иногда и невозможно. Однако ввиду сложности изучаемых явлений решить аналитически дифференциальные уравнения в частных производных современными математическими методами. Также существуют много методов решения, пригодных для практического использования, такие как аналитический, аналоговый, численный, графический и экспериментальный. Температура является параметром, характеризующим энергию теплового движения частиц вещества. Следовательно, процесс распространения теплоты и его направление неразрывно связаны с распределением температуры внутри тела. В этой связи в работе предлагается сведение уравнения теплопроводности к системе обыкновенных дифференциальных уравнений. Поставленная задача решается приведением к системе линейных обыкновенных дифференциальных уравнений, для решения которой разработан соответствующий алгоритм. Поэтому разработка специальных методов и вычислительных алгоритмов и комплекса прикладных программ, позволяющих исследовать установившегося теплофизического состояния стержней ограниченной длины находящихся под одновременным воздействием разнородных видов источников тепла, является актуальной проблемой. Разработана программа нахождения распространения температуры по стрежню, которая помещает результаты численных расчетов в несколько файлов. Результаты численных расчетов в динамике (по времени) выводится в виде таблицы и отображаются в виде одномерных графиков. Они не противоречат экспериментальным данным. Перспективным направлением является применение интервальной математики для исследования уравнения теплопроводности.

Ключевые слова: теплопроводность, теплоизоляция, температура, нестационарный теплофизический процесс, энергия.

Introduction. Rods of limited length are used as load-bearing elements of modern jet and hydrogen engines, gas-generating, nuclear and thermal power plants, technological lines of the processing industry, spacecraft power plants. The load-bearing elements of these installations operate under the simultaneous influence of heterogeneous types of heat sources. Therefore, the development of special methods and computational algorithms and a set of applied programs that make it possible to study the steady thermophysical state of rods of limited length that are under the simultaneous influence of heterogeneous types of heat sources is an urgent problem.

There are several methods for solving thermal conductivity problems: analytical, analog, numerical, graphical, and experimental. Four of them come directly from various forms of equations. The experimental method is used when other methods do not give results. It is used to determine thermophysical properties, such as thermal conductivity and specific thermal capacity (Karpovich et all, 2015).

Analytical and numerical methods are used to solve thermal conductivity problems in complex solids. Solutions are possible under known boundary conditions, including the initial temperature distribution in the body and boundary conditions on the surface of the body, which can be set in one of three ways: surface temperature, heat flow and heat-exchange coefficient (Voronenko et all, 2014).

Temperature is a parameter that characterizes the energy of the thermal motion of particles of a substance. Consequently, the process of heat propagation and its direction are inextricably linked with the temperature distribution inside the body. In general, the temperature is not the same at different points of the body and depends on time: $T = T(x, y, z, t)$.

After a significant period of time, the temperature of all parts of the body equalizes and becomes equal to the temperature of the medium (this is true for the case when the volume of the medium is much larger than the volume of the body and its temperature practically does not change with time) (Dede et all, 2020).

The work (Dede et all, 2020) contains new results on spectral methods for solving incorrectly set problems using the example of the Cauchy problem for the parabolic equation: a method for regularizing the solution of the inverse problem is proposed. The regularized equation is obtained by introducing a biquadratic Laplacian into the heat equation with a coefficient equal to the regularization parameter.

For the inverse boundary value problem of heat exchange, an approximate solution is constructed by the quasi-conversion method and an order-accurate estimate of the error of the constructed approximate solution is obtained

on one of the correctness classes of the inverse boundary value problem ([Visaria et all, 2020).

For an incorrectly set problem with the reverse time of the semi-linear differential-operator equation, a stable approximate solution was built and an estimate of its error was given (Jaffe et all, 2021).

Using the method of spectral analysis, a criterion for the uniqueness of the solution to the inverse problem for finding the initial condition is established. The theorems of uniqueness, existence and stability of the solution are proved for these problems (Amrit et all, 2021).

Materials and methods. Research analysis and problem statement. Consider a horizontal rod of limited length l_2 and constant cross section $S_{cs} = l_1 * l_1$. Build a global Cartesian coordinate system Oxyz (Figure 1).

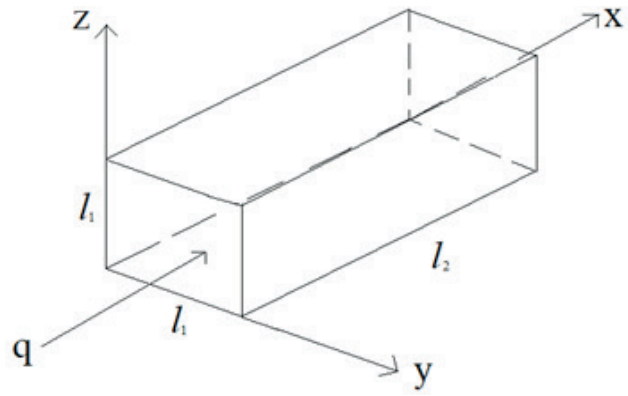


Figure 1 – General view of a metal rod with a square section

Heat propagation in a homogeneous rod in the absence of heat sources is described by the following three-dimensional thermal conductivity equation

$$\frac{\partial T}{\partial t} = a^2 \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right), \qquad (1)$$

where a^2 is the coefficient of thermal diffusivity, $t_0 < t < t_1$ – is the period of time during which the process of thermal conductivity of the rod is studied.

x, y, z are space variables $0 \le y, x \le l_1$, $0 \le y, x \le l_2$,

x_c, y_c, z_c – the center of the rod: $x_c - l_2/2$, $y_c - l_1/2$, $x_c - l_1/2$;

l_1 – width and height of the rod;

l_2 – rod length;

D – parallelepiped $\{0 \le y, x \le l_1, \ 0 \le y, x \le l_2\}$, Γ – border D,

Q – $\{x, y, z, t | (x, y, z) \in D, t \in t_0, t_1\}$.

The partial differential equation (1) is the differential equation of energy conservation for the isochoric heat transfer process, or the equation of

unsteady thermal conductivity. It establishes a relationship between temporal and spatial temperature change at any point in a solid, in which the thermal conductivity process occurs.

It is assumed that the left end of the rod coincides with the origin and the heat transfer coefficient is assumed to be constant over the entire surface of the rod. It is also assumed that the rod is under the influence of point temperature and surface heat exchange.

To isolate the unique solution of the thermal conductivity equation, it is necessary to add the initial and boundary conditions to equation (1).

In the general case, the initial condition can be analytically written as follows:

$$T|_{t=0} = q(M),\ M = (x, y, z) \in D. \qquad (2)$$

We set the boundary conditions in the form

$$\frac{\partial T}{\partial n}|_{\Gamma} = 0,\quad T(0, y_c, z_c, t) = q. \qquad (3)$$

The differential thermal conductivity equation together with the initial and boundary conditions fully determine the problem, that is, knowing the geometric shape of the body, the initial and boundary conditions, it is possible to solve the differential equation to the end and, consequently, find the temperature field in the body, T(x, y, z, t) - the temperature distribution function at any time t. The function T(x, y, z, t) must satisfy the differential equation (1), as well as the initial and boundary conditions.

This implies the correctness of the problem under study, however, due to the complexity of the phenomena under study, it is most often very difficult, and sometimes even impossible, to solve analytical partial differential equations using modern mathematical methods. However, there are many solution methods suitable for practical use. In this regard, it is proposed to reduce the thermal conductivity equation to a system of ordinary differential equations.

Result and discussion. Development of a computational algorithm. We cover the domain D with a uniform grid with steps Δx, Δy, and Δz along the x, y, and z axes respectively. We write the following difference-differential approximation of the equation (2)

$$\frac{dT_{i,j,k}}{dt} = a^2 \left(\frac{T_{i+1,j,k} - 2T_{i,j,k} + T_{i-1,j,k}}{\Delta x^2} + \frac{T_{i,j+1,k} - 2T_{i,j,k} + T_{i,j-1,k}}{\Delta y^2} + \right.$$

$$+\frac{T_{i,j,k+1}-2T_{i,j,k}+T_{i,j,k-1}}{\Delta z^2}\Bigg), \quad (4)$$

$$t \in [t_0,\ t_1],\ i = \overline{2, nx-1},\ j = \overline{2, ny-1}, k = \overline{2, nz-1},$$
$$\Delta x = l_1/nx,\quad \Delta y = l_1/ny,\quad \Delta x = l_2/nz,$$

nxyz=nx*ny*nz.

Here Δx is the step and nx is the number of split points along the Ox axis, Δy is the step and ny is the number of split points along the Oy axis, Δz is the step and nz is the number of split points along the Oz axis, the indices i, j, k are in x, y, and z coordinates, respectively (Sikovsky, 2013).

Initial conditions (2) take the form

$$T_{1,j,k}(t_0) = q_{i,j,k}\ ,\ i = \overline{1, nx},\ j = \overline{1, ny}, k = \overline{1, nz}. \quad (5)$$

Boundary conditions (3) are approximated by the following differential equations

$$\frac{dT_{1,j,k}}{dt} = 0, \frac{dT_{nx,j,k}}{dt} = 0,\ \frac{dT_{i,1,k}}{dt} = 0, \frac{dT_{i,ny,k}}{dt} = 0, \frac{dT_{i,j,1}}{dt} =$$
$$= 0, \frac{dT_{i,j,nz}}{dt} = 0, \quad (6)$$
$$i = \overline{1, nx},\ j = \overline{1, ny}, k = \overline{1, nz}$$

Introduce a vector x of dimension nxyz and a matrix A of dimension nxyz*nxyz.

We define the elements of the vector x as follows:

$$x_p\ (t) = T_{i,j,k}(t), p = (i-1) * ny * nz + (j-1) * nz + nz \quad (7)$$
$$i = \overline{1, nx},\ j = \overline{1, ny}, k = \overline{1, nz}.$$

The elements of the matrix A are determined through the coefficients of equation (4) and the given steps Δx, Δy, Δz along the corresponding axes.

Then the original problem (2)-(4) is reduced to the following Cauchy problem for a system of linear ordinary differential equations:

$$\frac{dx}{dt} = Ax, x(t_0) = x_0\ , t \in [t_0,\ t_1] \quad (8)$$

The following iterative solution algorithm is proposed:

1. Based on the data of the original problem (1)-(3), the matrix A is constructed

2. According to the initial conditions (5), the vector x_0 is calculated

3. Based on the Runge-Kutta method, the Cauchy problem (8) is solved

4. The resulting solution is output to Rezult.txt and to the GrafX.txt file for subsequent visualization using the MatLab software tool [Anufriev et all, 2005 – Dyakonov, 2005].

Numerical solution of problems for specific initial data. A program has been developed for finding the temperature distribution along the rod, which places the results of numerical calculations in several files. The results of numerical calculations in dynamics (over time) are displayed in the form of one-dimensional graphs. (Mazakov et all, 2021- Nurdaulet et all, 2018).

The calculations were carried out with the following initial data:

l_1 – 1.0; l_2 – 10,0; Δt – 0,01; nx =10; ny= 6; nz= 6; q– 200.

Figures 2-3 present the results of experimental calculations in graphical form. Figure 2 shows a graph of temperature distribution along the center of the rod in the X direction from the origin in dynamics.

In view of the large temperature difference from 0 to 200 degrees, Figure 3 shows a 2nd graph of the temperature distribution along the center of the rod in the X direction with a one-step offset from the origin in dynamics.

As can be seen from Figure 3, the temperature in the center of the rod at the end increases from 0 to 5 degrees in 100 seconds.

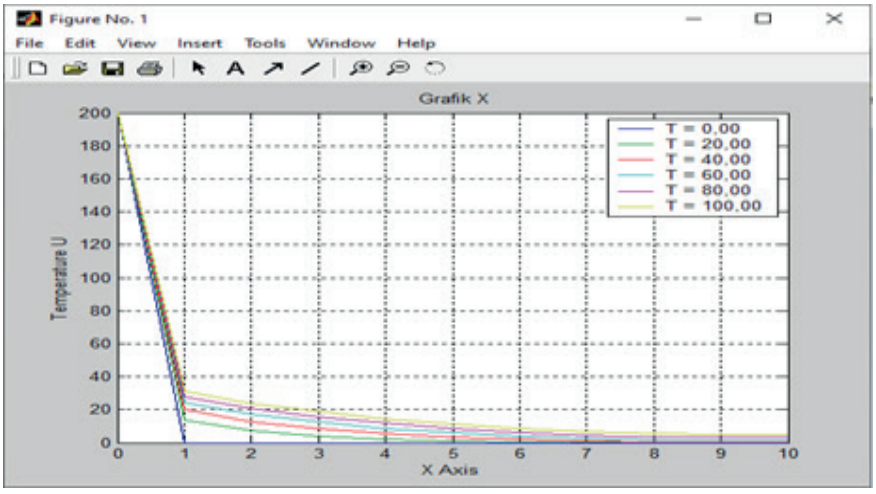


Figure 2 – Graph of temperature distribution along the center of the rod in the X direction from the origin

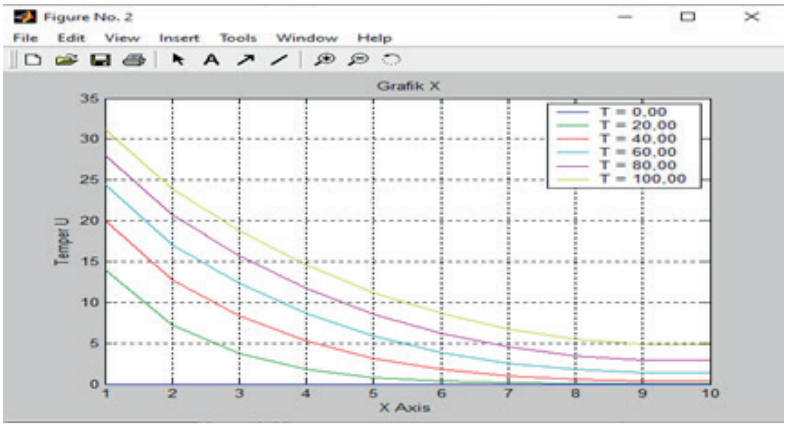


Figure 3 – Graph of temperature propagation along the center of the rod in the X direction with a one-step offset from the origin

Conclusion. The study of the equation of thermal conductivity of a rod with a square section is reduced to a system of linear ordinary differential equations, for the solution of which an appropriate algorithm has been developed. The results of numerical calculations do not contradict the experimental data. Additionally, the results are output to text files and provide the construction of one-dimensional images of temperature dynamics using the MatLab system, for which the corresponding program has been written. A promising direction is the use of interval mathematics to study the thermal conductivity equation.

Acknowledgements. *The work was carried out at the expense of program-targeted funding of scientific research for 2021-2022 under the IRN project OR11465437.*

Information about authors:

Mazakova Aigerim Talgatovna – PhD student of the Al-Farabi Kazakh National University; E-mail: *aigerym97@mail.ru*; ORCID: 0000-0003-3019-3352;

Begaliyeva Kalamkas Baltabekovna – PhD student of the Al-Farabi Kazakh National University; E-mail: *kalamkas_b@mail.ru*; ORCID: 0000-0002-4216-9184;

Mazakov Talgat Zhakupovich – Doctor of the Physical and Mathematical Sciences, Professor of the Al-Farabi Kazakh National University, CRS of the Institute of Information and Computational Technologies CS MES RK; E-mail: tmazakov@mail.ru; ORCID: 0000-0001-9345-5167;

Jomartova Sholpan Abdrazakovna – Doctor of Technical Sciences, Associate Professor of the Al-Farabi Kazakh National University; E-mail: jomartova@mail.ru; ORCID: 0000-0002-5882-5588;

Ziyatbekova Gulzat Ziyatbekkyzy – PhD, Al-Farabi Kazakh National University, Almaty, Kazakhstan; Senior Researcher of the Institute of Information and Computational Technologies CS MES RK; E-mail: ziyatbekova@mail.ru; ORCID: *0000-0002-9290-6074*.

REFERENCES

Amrit J., Nemchenko K., Vikhtinskaya T. (2021) Effect of diffuse phonon boundary scattering on heat flow // Journal of Applied Physics. – 129(8). DOI: 10.1063/5.0036935. (in Eng.).

Anufriev I.E., Smirnov A.B., Smirnova E.N. (2005) Matlab 7. – St. Petersburg: BHV-Petersburg. – 1104 p. (in Russ.).

Dede E.M., Yu Z., Schmalenberg P., Iizuka H. (2020) Thermal metamaterials for radiative plus conductive heat flow control // Applied Physics Letters. – 116(19). DOI: 10.1063/5.0007574. (in Eng.).

Dyakonov V.P. (2005) Matlab 6.0/6.1/6.5/6.5+SP1+Simulink 5/5. Processing of signals and images. – M.: SOLON-Press. – 592 p. (in Russ.).

Jaffe G.R., Brar V.W., Lagally M.G., Eriksson M.A. (2021) A simple numerical method for evaluating heat dissipation from curved wires with periodic applied heating // Applied Physics Letters. – 119(16). DOI: 10.1063/5.0059648. (in Eng.).

Karpovich D.S., Susha O.N., Korovkina N.P., Kobrinets V.P. (2015) Analytical and numerical methods for solving the heat equation // Proceedings of BSTU. Physical and Mathematical Sciences and Informatics. – No. 6. – Pp. 122-127. (in Russ.).

Mazakov T., Wójcik W., Jomartova Sh., Karymsakova N., Ziyatbekova G., Tursynbai A. (2021) The Stability Interval of the Set of Linear System // INTL Journal of Electronics and Telecommunications. – Vol. 67, N. 2. – Pp.155-161. DOI: 10.24425/ijet.2021.135958. (in Eng.).

T.Zh. Mazakov, Sh.A. Jomartova, T.S. Shormanov, G.Z. Ziyatbekova, B.S. Amirkhanov, P. Kisala. The image processing algorithms for biometric identification by fingerprints // News of the national academy of sciences of the Republic of Kazakhstan. Series of Geology and Technical Sciences, 2020. – Vol. 1, – No 439. – P. 14-22. ISSN 2518-170X (Online), ISSN 2224-5278 (Print). *https://doi.org/10.32014/2020.2518-170X.2* (in Eng.).

Nurdaulet I., Talgat M., Orken M., Ziyatbekova G. (2018) Application of fuzzy and interval analysis to the study of the prediction and control model of the epidemiologic situation // Journal of Theoretical and Applied Information Technology, Pakistan. – Vol. 96, – Issue 14, – Pp. 4358-4368. (in Eng.).

Sikovsky D.F. (2013) Methods of computational thermal physics. Novosibirsk: Novosib. state un-ty. – 98 p. (in Russ.).

Visaria D., Jain A. (2020) Machine-learning-assisted space-transformation accelerates discovery of high thermal conductivity alloys // Applied Physics Letters. – 117(20). DOI: 10.1063/5.0028241. (in Eng.).

Voronenko B.A., Krysin A.G., Pelenko V.V., Tsuranov O.A. (2014) Analytical description of the process of non-stationary heat conduction. – St. Petersburg: NRU ITMO; IKhiBT. – 48 p. (in Russ.).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 164-184
https://doi.org/10.32014/2022.2518-1726.145
УДК 519.816

Ж.Ж. Молдашева[1], Б.Б. Оразбаев[1], Б.У. Асанова[2], С.Ш. Искакова[3], К.Н. Оразбаева[4*]

[1]Л.Н. Гумилев атындағы Еуразиялық ұлттық университеті,
Қазақстан, Астана;
[2]Х. Досмухамедов атындағы Атырау университеті, Қазақстан, Атырау;
[3]С. Өтебаев атындағы Атырау мұнай және газ университеті,
Қазақстан, Атырау;
[4] Esil University, Қазақстан, Астана.
E-mail: *zhadira1985@mail.ru*

МҰНАЙ ҚҰБЫРЫ АГРЕГАТТАРЫНЫҢ ЖҰМЫС РЕЖИМДЕРІН БАСҚАРУ ҮШІН ЭВРИСТИКАЛЫҚ ТӘСІЛ ҚҰРУ

Аннотация. Қазіргі кезде Қазақстан Республикасының мол мұнай қорын игеру процесінде, өндірілген мұнайды тұтынушылар мен әлімдік нарыққа жеткізуде мұнай құбырлары аса маңызды орын алады. Практикада магистралды мұнай құбыры арқылы мұнай тасымалдауда оның технологиялық агрегаттарының жұмыс режимдерін тиімді басқару үшін айқынсыздықты ескере отырып шешім қабылдау есептерін шешуге тура келеді. Сондықтан мұнай құбырлары агрегаттарының жұмыс режимдерін басқару бойынша айқын емес ортада шешім қабылдау есептерін тиімді шешуге мүмкіндік беретін эвристикалық тәсілдерді құру аса өзекті ғылыми, практикалық мәселе болып табылады.

Жұмыста магистралды мұнай құбырлары агрегаттарының айқын емес ақпараттық ортада жұмыс режимдерін басқаруда шешім қабылдау есептері зерттеліп, оларды шешу үшін эвристикалық тәсілдер жасақталған. Зерттеу нысандарын басқару шынайы практикада көпкритерийлікпен және айқынсыздықпен сипатталатындықтан, олардың жұмыс режимдерін басқару есептері айқын емес көпкритерийлі оптимизациялау есептері түрінде қойылымдары алынған. Шешім

қабылдау есептерінің математикалық қойылымдары мен оларды шешудің эвристикалық тәсілдерін құру түрлі оптималдық принциптерін (компромисстік схемаларды) айқынсыздықта жұмыс істеуге модификациялау арқылы жүзеге асырылған.

Айқын емес есептерді шешудің белгілі тәсілдерінен оптималдық принциптерін модификациялау незінде қойылған есептерді шешуге ұсынылып отырған тәсілдердің ерекшеліктері мен жаңашылдығы, айқын емес есеп оны қою бырысында детерминделген есепке түрлендірілмей, айқын емес ортада қойылып, шешім қабылдаушы тұлға, эксперттердің білімін, тәжірибесін, интуициясы негізінде эвристикалық жолмен шешіледі. Бұл тәсілдеме жинақталған, қолжетімді айқын емес ақпаратты толықтай қолдану есебінен айқын емес ортада өндірістік есептердің тиімді және адекватты шешімін алуға мүмкіндік береді. Ұсынылған тәсілдеме Өзен-Атырау-Самара магистралдық мұнай құбырының Атырау пунктісіндегі мұнай қыздыру станциясының жұмыс режимдерін басқару бойынша шешім қабылдау есебін шешуде тексеріліп, сынақтан сәтті өткен.

Түйін сөздер: мұнай тасымалдау құбыры, мұнай қыздыру станциясы, айқын емес ақпарат, шешім қабылдау, шешім қабылдаушы тұлға (ШҚТ), эвристикалық тәсіл, оптималдық принциптері.

Ж.Ж. Молдашева[1], Б.Б.Оразбаев[1], Б.У. Асанова[2], С.Ш. Искакова[3], К.Н. Оразбаева[4]*

[1]Евразийский национальный университет имени Л.Н. Гумилева, Казахстан, Астана;
[2]Атырауский университет имени Х.Досмухамедова, Казахстан, Атырау;
[3]Атырауский университет нефти и газа имени С. Утебаева, Казахстан, Атырау;
[4] Esil University, Казахстан, Астана.
E-mail: *zhadira1985@mail.ru*

РАЗРАБОТКА ЭВРИСТИЧЕСКОГО МЕТОДА ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ УПРАВЛЕНИЯ РЕЖИМАМИ РАБОТЫ АГРЕГАТОВ НЕФТЕПРОВОДА

Аннотация. В настоящее время в процессе освоения богатых запасов нефти Республики Казахстан нефтепроводы играют очень важ-

ную роль в доставке добываемой нефти к потребителям и мировому рынку. На практике необходимо решать задачи принятия решений с учетом неопределенности и нечеткости, чтобы эффективно управлять режимами работы его технологических агрегатов при транспортировке нефти по магистральному нефтепроводу. Поэтому создание эвристических методов, позволяющих эффективно решать задачи принятия решений по управлению режимами работы объектов нефтепровода в нечеткой среде является в настоящее время весьма актуальной научной и практической задачей.

В работе исследуются проблемы принятия решений при управлении режимами работы агрегатов магистральных нефтепроводов в нечеткой информационной среде и разрабатываются эвристические подходы к их решению. Так как управление объектами исследования в реальной практике характеризуется многокритериальностью и нечеткостью, задачи управления режимами их работы представляются в виде нечетких задач многокритериальной оптимизации. Математические постановки задач принятия решений и эвристические методы их решения реализовались путем модификации различных оптимальных принципов (компромиссных схем) для работы в условиях нечеткости.

Особенности и новизна предлагаемых методов решения задач нечеткой среде от известных методов решения нечетких задач, заключаются в том, что нечеткая задача в момент постановки не преобразуется в набор детерминированных задач, а ставится и решается в нечеткой среде эвристическим путем на основе знаний, опыта и интуиции лица, принимающего решения, экспертов. Такой подход позволяет получить эффективное и адекватное решение производственных задач в нечеткой среде за счет полного использования имеющейся, собранной нечеткой информации. Предложенный подход успешно апробирован при решении задачи управления режимами работы станции подогрева нефти в пункте Атырау магистрального нефтепровода Узен-Атырау-Самара.

Ключевые слова: нефтеповод, станция подогрева нефти, нечеткая информация, принятия решений, лицо, принимающее решение (ЛПР), эвристический метод, принципы оптимальности.

Zh. Moldasheva[1], B. Orazbayev[1], B. Assanova[2], Sh. Iskakova[3], K. Orazbayeva[4*]

[1]L.N. Gumilyov Eurasian National University, Kazakhstan, Astana;
[2]H. Dosmukhamedov Atyrau University, Kazakhstan, Atyrau;
[3]S. Utebayev Atyrau Oil and Gas University, Kazakhstan, Atyrau;
[4] Esil University, Kazakhstan, Astana.
E-mail: *zhadira1985@mail.ru*

OPTIMIZATION OF OPERATION MODES OF REFORMING REACTORS OF A CATALYTIC REFORMING UNIT ON THE BASIS OF COMPUTER MODELING

Abstract. Currently, in the process of developing the rich oil reserves of the Republic of Kazakhstan, oil pipelines play a very important role in the delivery of produced oil to consumers and the world market. In practice, it is necessary to solve decision-making problems taking into account uncertainty and fuzziness in order to effectively control the operating modes of its technological units during oil transportation through the main oil pipeline. Therefore, the creation of heuristic methods that make it possible to effectively solve the problem of decision-making on managing the operation modes of oil pipeline facilities in a fuzzy environment is currently a very relevant scientific and practical task.

The paper studies the problems of decision-making when managing the operation modes of oil trunk pipeline units in a non-intrusive information environment and develops heuristic approaches to their solution. Since the management of research objects in real practice is characterized by multi-criteria and fuzziness, the tasks of controlling their operation modes are presented in the form of fuzzy problems of multi-criteria optimization. Mathematical formulations of decision-making problems and heuristic methods for their solution were implemented by modifying various optimal principles (compromise schemes) for operation in fuzzy conditions.

The features and novelty of the proposed methods for solving problems in a fuzzy environment from the known methods for solving fuzzy problems lie in the fact that a fuzzy problem at the time of setting is not converted into a set of deterministic problems, but is posed and solved in a fuzzy environment in a heuristic way based on the knowledge, experience and intuition of a person decision maker, experts. This approach makes it possible to obtain an effective and adequate solution of production problems

in a fuzzy environment through the full use of the available, collected fuzzy information. The proposed approach has been successfully tested in solving the problem of controlling the operating modes of an oil heating station at the Atyrau point of the Uzen-Atyrau-Samara main oil pipeline.

Key words: pipeline, oil heating station, fuzzy information, decision making, decision maker (DM), heuristic method, optimality principles.

Кіріспе. Практикада өндірістітиімдібасқаруүшінкөпкритерийлікпен, яғни экономикалық, технологиялық және экологиялық сипаттағы критерийлер векторымен, сондай-ақ бастапқы ақпараттың айқын еместігімен сипатталатын технологиялық нысандардың оптималды жұмыс режимдерін таңдау бойынша шешім қабылдау есептері жиі туындайды. үшін туындайды. Мұндай өндірістік нысандардың көптеген параметрлерігі мен көпкритерийлігі, олардың математикалық сипатталуына қажетті бастапқы ақпараттың тапшылығы мен айқын еместігі, аталған есептерді формализациялау, математикалық тұжырымдау және шешуді қиындатады (Алиев и др., 2017:378.; Kuz'min et all., 2012: 1649-1678).

Соңғы уақытта ғылыми әдебиеттер мен жарияланымдарда өндірісті оптималды басқару (Емельянов и др., 2016:88.; *Dimitriadi* et all., 2017:1322-1335.,), оның ішінде айқын емес бастапқы ақпарат жағдайында (Orazbayev et all., 2021: 147–162., Зайченко и др., 2019:355) көпкритерийлі шешім қабылдау есептерін шешудің мәселелері мен тәсілдері белсенді талқылануда. Айқын емес ортада бұл есептерді шешу тәсілдері айқын емес жиындар теориясы тәсілдерін қолдануға негізделген (Алиев и др., 2017:378., Орловский и др., 2018:287., Рыжов и др., 2017:115., *Markovskii* et all., 2018:1486-1495). Технологиялық нысандардың оптималды параметрлері мен жұмыс режимдерін көпкритерийлі таңдау есептерін қою және шешу мәселелері (Оразбаев и др., 2010:307., *Grebenyuk* et all., 2016:805-812., Orazbayev et all., 2021:147-162., Зайченко и др. 2019:355., Оразбаев и др., 2007:138., Оразбаев и др., 2022:71-82) қарастырылған.

Магистральдық мұнай құбырларының технологиялық объектілері жұмыс режимдерін айқынсыздықта басқару бойынша бұл жұмыста тұжырымдалып, эвристикалық шешу тәсілдері ұсынылатын шешім қабылдау есептері, айқын емес жиындар және көпкритерийлі оптимизациялау теорияларының маңызды ғылыми, практикалық сұрақтарымен байланысты. Сонымен қатар бұл бағыт мұнай айдау саласының аса өзекті мәндеттеінің біріне жатады. Сонымен бұл

мақалада айқын емес ортада мұнай құбырының технологиялық агрегаттарының модельдері негізінде олардың жұмыс режимдерін басқару бойынша шешім қабылдау есептерін формализациялап, математикалық қойлымдарын тұжырымдау, сондай-ақ оларды шешу тәсілдерін әзірлеу мәселелері зерттеледі. Аталған есептерді тұжырымдау және шешу кезінде айқынсыздыққа модификацияланған шешім қабылдау компромисстік схемалары пайдаланылады (Orazbayev et all., 2021:147-162., Оразбаев и др., 2007:138). Өндіріс жағдайында көптеген технологиялық объектілер көпкритерийлі және бастапқы ақпараттың айқынсыздығымен сипатталатындықтан, зерттелетін және шешілетін мәселелер шешім қабылдау және басқару теориясы мен практикасының маңызды міндеті болып табылады.

Жұмыстың мақсаты бастапқы ақпараттың айқын еместігін ескере отырып, мұнай қызжыдыру станциясы мысалында көпкритерийлі технологиялық объектілердің жұмыс режимдерін басқару бойынша шешім қабылдау мәселесін формализациялап, математикалық тұжырымдау және оларды шешудің интерактивті режимде жүзеге асырылатыын тиімді тәсілдерін әзірлеу. Көпкритерилік және бастапқы ақпараттың айқынсыздығы мәселелерін шешу үшін жасақталатын эвристикалық тәсілдер, бастапқы ақпараттың айқын емес еместігі шарттарына модификацияланып, бейімделген, детерминдік жағдайда белгілі оптималдық принциптерінің идеялары мен айқын емес жиындар теориясы математикалық аппаратының мүмкіндіктері пайдаланылады.

Есепті қою және зерттеу тәсілдері. Практикада мұнай айдау жүйесінің өндірістік есептеррі көбтеген жағдайларда көпкритерийлі және айқын емес бастапқы ақпаратпен сипатталатынды, ал бұл жағдайлар магистральдық мұнай құбырларының технологиялық агрегаттарының жұмыс режимдерін басқару үшін шешім қабылдау есептерін қою және шешу процедураларын күрделендіреді. Көпкритерийлі және қол жетімді ақпараттың анық еместігі жағдайында аталған объектідің жұмыс режимдерін басқару есептерін айқын емес ортада шешім қабылдау есептері түрінде тұжырымдап, шешу қажет. Бұл кезде шешім қабылдау берілген экономикалық және экологиялық критерийлерге сәйкес ықтимал шешімдер жиынын бағалап, олардың ішінен ең жақсы шешімді таңдау бролып табылады. Аталған есепті келесіден формализациялап, қоюға болады:

$\mathbf{f}(\mathbf{x}) = f_1(\mathbf{x}),..., f_m(\mathbf{x})$ – мұнай құбыры технологиялық жүйесінің жұмыс сапасын бағалайтын критерийлер векторы болсын. Мысалы, $f_1(\mathbf{x}), f_2(\mathbf{x}),..., f_k(\mathbf{x})$ – сәйкесінше, мұнай айдау көлемі, өндірістік шығындар,

пайда, т.б. сияқты экономикалық критерийлер; $f_{k+1}(\mathbf{x}), f_{k+2}(\mathbf{x}),..., f_m(\mathbf{x})$ – өндірістің экологиялық жағдайын бағалайтын локалды критерийлер, мысалы қоршаған ортаға тасталатын мұнай, түтін, тасымалдау қалдықтары сияқты зиянды заттар көлемі, табиғатты қорғау шараларының көлемі т.с.с. m критерийлердің әр қайсысы, нысан жұмысын басқару үшін қолданылатын нысанның n кіріс, режимдік параметрлі векторына $\mathbf{x} = (x_1,..., x_n)$ байланысты анықталады. Мұнай құбыры технологиялық агрегаттардың кіріс, режимдік параметрлеріне, мысалы: оолардың кірісінрдегі шикізат көлемі; температурасы; қысымы; мұнайдың реологиялық қасиеттері, отын, реагенттер көлемі т.б. жатады. Өндірістік практикада әрқашан түрлі (экономикалық, технологиялық, қаржылық, экологиялық) шектеулер болады, оларды жалпы түрде келесі шектеу функциялары арқылы $\varphi_q(\mathbf{x}) \geq b_q,\ q = \overline{1, L}$ арқылы сипаттауға болады. Режимдік, басқару параметрі де, агрегаттың технологиялық режимімен, табиғатты қорғау талаптырымен анықталатын, өздерінің өзгеру интервалдарында шектеледі: $x_i \in \Omega = [x_i^{\min}, x_i^{\max}]$, мұнда x_1 параметрінің өзгеруінің $x_i^{\min}$ – минималды және $x_i^{\max}$ – максималды мәндері. Келтірілген шектеулер айқынсыздықпен сипатталуы мүмкін.

Сонымен жоғарыда келтірілген есепті формализациялау негізінде көпкритерийлі пен айқынсыздық жағдайларында мұнай құбыры технологиялық агрегаттары жұмыс режимдерін тиімді басқару бойынша шешім қабылдау есебін келкесідей айқынсыздыққа бейімдеп, мате матикалық қойылымын алуға болады. Ол үшін шешім қпабылдау есебін төмендегідей түрлендіреміз:

$\mu_0(\mathbf{x}) = \left(\mu_0^1(\mathbf{x}),..., \mu_0^m(\mathbf{x})\right)$ – қарастырылған $f_1(\mathbf{x}),..., f_m(\mathbf{x})$ критерийлерінің нормализацияланған векторы болсын, ал L шектеулер $f_q(\mathbf{x}) \lesseqgtr b_q,\ \ q = \overline{1, L}$ – айқын емес инструкциялар арқылы сипатталсын. Әр айқын емес шектеудің орындалуын сипаттайтын тиістілік функциялары $\mu_q(\mathbf{x}),\ q = \overline{1, L}$ белгілі, немесе шешім қабылдаушы тұлға (ШҚТ), эксперртер көмегімен тұрғызылаты деп қабылдайық (Orazbayev et all., 2021:147-162., Баронец и др. 2019:156-160. Сондай-ақ критерийлердің $\boldsymbol{\gamma} = (\gamma_1,..., \gamma_m)$ және шектеулердің $\boldsymbol{\beta} = (\beta_1,..., \beta_L)$ маңыздылықтарын (салмақтарын) бейлелейтін салмақ векторлары белгілі нее анықталады деп санаймыз (Оразбаев и др., 2022:71-82., Larichev et all., 2002:304-315).

Сонда экономикалық және экологиялық критерийлері бойынша магистральдық мұнай құбырларының жұмыс режимдерін басқару есебін айқын емес ортада тиімді шешім қабылдау есебін жалпы түрде келесідей жазуға болады:

$$\max_{\mathbf{x} \in X} \mu_0^i(\mathbf{x}),\ i = \overline{1, m}\ , \qquad (1)$$

$$X = \left\{ \mathbf{x} : \arg \max_{\mathbf{x} \in \Omega} \mu_q(\mathbf{x}), q = \overline{1, L} \right\}. \quad (2)$$

Магистральдық мұнай құбырының мұнай жылыту станциясының технологиялық жүйесінің оптималды жұмысын қамтамасыз ететін басқарудың оптималды мәндерін, режим параметрлерін, көрсетілген шектеулер орындалған кезде критерийлер векторының экстремалды мәндерін, ШҚТ, өндіріс басшысы, өндірістік персоналдың қалауын ескере отырып, табу қажет. Біздің жағдайда ШҚТ – мұнай құбырының технологиялық нысанының, мысалы сорғы, мнай қыздыру станциялары, басқару критерийлері (айдау көлемі, режимнің қауіпсіздігі мен сенімділігі және т.б.) бойынша оптималды мәндерін қамтамасыз ететін, мұнай құбырлары арқылы айдау режимдерін басқару операторлары операторлары.

Жұмыста көпкритерийлік жағдайда қабылдау есептерін формализациялау, математикалық қойылымын тұжырымдау және эвристикалық шешу тәсілдерін жасақтау үшін оптималдық принциптері, көпкритерийлік оптитмизациялау (*Groppen* et all., 2018:660-675., *Dimitriadi* et all., 2017:1322-1335., Orazbayev et all., 2021:147-162., Зыков и др. 2018:208) және шешім қабылдау теориясы тәсілдері қолданылады (Емельянов и др., 2016:88., Rademaker et all., 2011:29-51). Ал айқынсыздық мәселелеріен шешу үшін эксперттік бағалай және айқын емес жиындар теорисының тәсілдері пайдаланылады (Алиев и др., 2017:378., Kuz'min et all., 2012:1649-1678., Орловский и др., 2018: 287., Liu et all., 2022:2368., Sabzi et all., 2017:145-163).

Зерттеу нәтижелері. Жалпы түрде алынған (1)–(2) айқынсыздықта шешім қабылдау есебін орын алған өндірістік жағдайға нақтылап, математикалық тұрғыдан дұрыствп қою сұрақтарын шешу нәтижелерін қарастырайық.

Өндірісте келесі жағдай туындасын делік: нысанды $\mu(\mathbf{x}) = (\mu_0^1(\mathbf{x}),..., \mu_0^m(\mathbf{x}))$ нормалданған m критерийлері бойынша және тиістілік функциялары $\mu_q(\mathbf{x}),\ q = \overline{1, L}$ бірнеше айқын емес шектеулер талаптарын орындай отырып тиімді басқару үшін шешім қабылдау қажет болсын. Критерийлер басымқылары (приоритеттері) қатары $I = \{1,...,m\}$ немесе локалды критерийлердің салмақ векторы $\boldsymbol{\gamma} = (\gamma_1,..., \gamma_m),\ \gamma_i \geq 0, i = 1, m$, m, $\gamma_1 + \gamma_2 + ... + \gamma_m = 1$, белгілі болсын.

Сонда жалпы түрдегі (1)–(2) есебін айқынсмыздықта нақты түрде келесідей жазуға болады:

$$\max_{x \in X} \mu_0^i(\mathbf{x}), \quad i = \overline{1, m}, \quad (3)$$

$$X = \left\{ \mathbf{x} : \arg\max_{\mathbf{x}\in\Omega} \sum_{q=1}^{L} \beta_q \mu_q(\mathbf{x}) \wedge \sum_{q=1}^{L} \beta_q = 1 \wedge \beta_q \geq 0 \; q = \overline{1,L} \right\}. \qquad (4)$$

Алынған (3) – (4) қойылымындағы есеп, m критерийлер бір нүктеде біруақытта максимумға жетуді талап ететіндіктен, шешімі күрделі және өте сирек табылады.

Бұл жағдайдан шығудың әмбебап тәсілі Парето жиынын анықтап, ол жиыннан ШҚТ көмегімен еі жақсы шешімді таңдау болып болып табылады. Сонымен Парето оптималдығы прицині негізінде соңғы есепті келесідей жазуға болады:

$$\max_{\mathbf{x}\in X} \mu_0(\mathbf{x}),\ \mu_0(\mathbf{x}) = \sum_{i=1}^{m} \gamma_i \mu_0^i(\mathbf{x}) \qquad (5)$$

$$X = \left\{ \mathbf{x} : \arg\max_{\mathbf{x}\in\Omega} \sum_{q=1}^{L} \beta_q \mu_q(\mathbf{x}) \wedge \sum_{q=1}^{L} \beta_q = 1 \wedge \beta_q \geq 0,\ q = \overline{1,L} \right\} \qquad (6)$$

Алнған шешім таңдау (5)–(6) есебін шеші үшін Парето оптисмалдық принц ипін айқынсыздыққа модификациялау арқылы жаңа және айқынсыздықта тиімді жұмыс жасайтын ПО+ПО эвристикалық тәсіл жасақталған. Ұсмынылған ПО+ПО эвристикалық тәсілін алгоритмизациялау нәтижесінде келесі негізгі қадамдардыан тұратын эвристикалық алгоритм алынған:

ПО+ПО эвристикалық алгоритмі:

Қадам 1. ШҚТ, эксперттер қатысуымен локалды критерийлер маңыздылықтарын басғалайтын салмақ векторын анықтау $\boldsymbol{\gamma} = (\gamma_1,...,\gamma_m)$, $\gamma_i \geq 0, i = \overline{1,m},\ \gamma_1 + \gamma_2 + ... + \gamma_m = 1$.

Қадам 2. Әр q-ші координат бойынша қадамдар санын анықтау: $p_q, q = \overline{1,L}$.

Қадам 3. Айқын емес шектеулердің $\boldsymbol{\beta} = (\beta_1,...,\beta_L)$ салмақ векторын өзгерту үшін әр қадамның шамасын келесі формуламен есептелеу: $h_q = 1/p_q, q = \overline{1,L}$.

Қадам 4. Алдыңғы қадамда анықталған $h_q, q = \overline{1,L}$ қадам шамасымен [0,1 интервалында өзгерте отырып, салмақ векторлары жиынын анықтау: $\beta^1, \beta^2, ..., \beta^N,\ N = (p_1 + 1)(p_2 + 1)...(p_L + 1)$.

Қадам 5. Айқын емес шектеулерді сипаттайтын терм-жиынды анықтау және олардың орындалу деңгейін бағалайтын тиістілік функцияларын тұрғызу: $\mu_q(\mathbf{x}),\ q = \overline{1,L}$.

Қадам 6. (5)–(6) қойлымындағы көпкритерийлі оптимизациялау $\max_{\mathbf{x}\in X} \mu_0(\mathbf{x}) = \max_{\mathbf{x}\in X} \sum_{i=1}^{m} \gamma_i \mu_0^i(\mathbf{x})$ есебін (6) өрнекпен анықталатын X рұқсат етілген жиынында шешіп: $\mathbf{x}(\boldsymbol{\gamma}, \boldsymbol{\beta})$ – кіріс, режимдік (басқару) параметрлері

векторын, бұл вектор қамтамасыз ететін $\mu_0^1(\mathbf{x}(\boldsymbol{\gamma},\boldsymbol{\beta})),...,\mu_0^m(\mathbf{x}(\boldsymbol{\gamma},\boldsymbol{\beta}))$ – локалды критерийлер мәндерін және $\mu_1(\mathbf{x}(\boldsymbol{\gamma},\boldsymbol{\beta})),...,\ \mu_L(\mathbf{x}(\boldsymbol{\gamma},\boldsymbol{\beta}))$ – айқын емес шектеулердің максималды орындюалу функцияларын анықтау.

Қадам 7. Алдыңғы қадамда алынған ағымдағы шешімдер талдау мен ең жақсы шешімді таңдау үшін ШҚТ-ға ұсынылады. Егер ұсынылған ағымдағы шешімдер ШҚТ-ны қанағаттандырса, онда келесі қадамға өту. Басқаша жағдайда, яғни ағымдағы шешімдер ШҚТ-ны қанағаттандырмаса, ол $\boldsymbol{\gamma}$ және/немесе $\boldsymbol{\beta}$ векторлары мәндерін шешімді жақсарту мақсатымен өзергетеді. Содан кейін шешімді жақсарту мақсатымен қадам 2-ге қайтып келу.

Қадам 8. Шешім іздеу тоқталап, ШҚТ-ны қанағаттандыратын ең жақсы шешім: $\mu_0^1(\mathbf{x}^*(\boldsymbol{\gamma},\boldsymbol{\beta})),...,\mu_0^m(\mathbf{x}^*(\boldsymbol{\gamma},\boldsymbol{\beta}))$ – локалды критерийлердің оптималды мәндері мен және $\mu_1(\mathbf{x}^*(\boldsymbol{\gamma},\boldsymbol{\beta})),...,\mu_L(\mathbf{x}^*(\boldsymbol{\gamma},\boldsymbol{\beta}))$ – айқын емес шектеелердің орындалу деңгейлерінің максималды мәндерін қамтамасыз ететін $\mathbf{x}^*(\boldsymbol{\gamma},\boldsymbol{\beta})$ – кіріс, режимдік параметрлерінің тиімді мәндерін шығару.

Осылайша шешім қабылдаудың түрлі оптималдық принциптерін комбинацияларын айқынсыздыққа модификациялау арқылы айқын емес ортада көпкритенийлі шешім қабылдау есептерінің басқа қойылымдарын тұжырымдап, оларды шешу тәсілдерін ұсынуға болады.

Мысалы, басты критерий (критерийлерге) мен идеалды нүкте (шектеулерге) принциптері комбинациясын айқынсыздыққа модефикациялау арқылы айқын,сыздықта шешім қабылдау есебінің келесідей қойылымын алуға болады:

$$\max_{\mathbf{x}\in X}\mu_0^1(\mathbf{x}), \tag{7}$$

$$X=\left\{\mathbf{x}:\mathbf{x}\in\Omega\wedge\arg\left(\max_{\mathbf{x}\in\Omega}\max\ \mu_0^i(\mathbf{x})\geq\mu_R^i\right)\wedge\arg\mu_q(\mathbf{x})\geq\min\left\|\mu(\mathbf{x})-\mu^u\right\|_D\right),i=\overline{2,m},q=\overline{1,L}\Big\}, \tag{8}$$

мұнда $\left\|\mu(\mathbf{x})-\mu^u\right\|_D$ – шектеулердің ағымдағы мәндері мен идеалды мәні ара қашықтығын бағалайтын метрика D; $\boldsymbol{\mu}(\mathbf{x})=(\boldsymbol{\mu}_1(\mathbf{x}),...,\mu_L(\mathbf{x}))$ – айқын емес шектеулердің орындалу деңгейін сипаттайтын тиістілік функциялары вектор; $\mu^u=(\max\mu_1(\mathbf{x}),...,\ \max\mu_L(\mathbf{x}))$ – идеалды нүкте координаттары, олар шектеулердің орындалу тиістілік функцияларының максималды мәндерімен анықталады. Егер аталған тиістілік функциялар нормалды болса, онда идеалды нүкте координаттары бірліктерге тең болады: $\mu^u=(1,...,1)$.

Бұл жағдайда басты критерий принципі негізінде ең маңызды, басты критерий анықталып, оны мәні оптимизацияланады, ал қалған локалды критерийлерге шектік мәндері тағайындалып, олар шектеулер

ретінде ескеріледі. Ал идеалды нүкте принципі, анықталған идеалды нүктеден, яғни шешімнен ағымдағы шешімнің ара қашықтықтын (мера) минимизациялауға негізделген оптималды шешімді табуға мүмкіндік береді.

Қойылған (7)–(8) көпкритерийлі шешім қабылдау есебін шешу үшін басты критерий мен идеалды нүкте принциптерін модификациялау негізінде бұл жұмыста БК+ИН эвристикалық тәсілі жасақталған.

Ұсынылған БК+ИН эвристикалық тәсілін алгоритмизациялау нәтижесінде, оның келесі негізгі қадамлдарын сипкаттауға болады:

БК+ИН эвристикалық алгоритмі:

Қадам 1. ШҚТ қатысуымен локалды критерийлердің басымқыларын енгізу $I_k = \{1,...,m\}$ (басты критерий 1-ші басымқыға ие болуы тиіс).

Қадам 2. ШҚТ, эксперттерден алынған ақпарат негізінде айқын емес параметрлердің терм-жиыны анықталады $T(X,Y)$ және әр айқын емес шектеулердің орындалу тиістілік функциялары тұрғызылады $\mu_q(\mathbf{x})$, $q = \overline{1,L}$.

Қадам 3. ШҚТ басқа локалды критерийлердің шектік мәндерін анықтап, ендіреді: $\mu_R^i(\mathbf{x}), i = \overline{2,m}$.

Қадам 4. Идеалды нүкте координаттарын анықтау (бұл нүкте координаттарын жалпы жағдайда айқын емес шектеулердің орындалу тиістілік функцияларының мачсималды мәндері $\mu^u = (\max \mu_1(\mathbf{x}),..., \max \mu_L(\mathbf{x}))$ немесе, ол тиістілік функциялар нормалды болса, бірліктерді $\mu^u = (1,...,1)$ алу қажет).

Қадам 5. Ағымдағы шешім $\mathbf{x}^* = (x_1^*,...,x_n^*)$ мен идеалды нүкте (шешім) μ^u арасындағы қашықтықты анықтайтын метрика түрін таңдау.

Қадам 6. (7)–(8) қойылымындағы шешім қабылдау есебін шешім, келесі ағымдағы шешімдерді анықтау: $\mu_0^1\left(\mathbf{x}\left(\mu_R^i, \left\|\mu(\mathbf{x}) - \mu^u\right\|_D\right)\right)$ – басты критерийдің ағымдағы мәнін, $\mu_0^2\left(\mathbf{x}\left(\mu_R^i, \left\|\mu(\mathbf{x}) - \mu^u\right\|_D\right)\right), ..., \mu_0^m\left(\mathbf{x}\left(\mu_R^i, \left\|\mu(\mathbf{x}) - \mu^u\right\|_D\right)\right)$, $i = \overline{2,m}$ – локалды критерийлердің және $\mu_1\left(\mathbf{x}\left(\mu_R^i, \left\|\mu(\mathbf{x}) - \mu^u\right\|_D\right)\right), ..., \mu_L\left(\mathbf{x}\left(\mu_R^i, \left\|\mu(\mathbf{x}) - \mu^u\right\|_D\right)\right)$ – шектеулердің тиістілік функцияларының ағымдағы мәндерін қамтамасыз ететін $\mathbf{x}\left(\mu_r^i, \left\|\mu(\mathbf{x}) - \mu^u\right\|_D\right)$ – кіріс, режимдік (басқару) параметрлері векторын.

Қадам 7. ШҚТ-ға алынған ағымдағы шешімдерді ұсыну. Егер алынған ағымдағы шешімдер ШҚТ-ны қанағаттандырмаса, онда ол шешімді жақсарту мақсатымен $\mu_R^i(x)$ мәнән өзгертеді және/немесе $\left\|\mu(x) - \mu^u\right\|_D$ метрикасының басқа түрін таңдайды да ең жавақсмы шешімді іздеу алғынғы қадамнан қайта басталады. Ал ШҚТ алынған ағымдағы шешімдермен қанағаттанса, келесі қадамға өту.

Қадам 8. ШҚТ таңдаған және оны қанағаттандыратын ең

жақсы соңғы шешімдерді шығару: $\mu_0^1\left(\mathbf{x}^*\left(\mu_R^i, \left\|\mu(\mathbf{x}^*)-\mu^u\right\|_D\right)\right)$ – басты критерийдің максималды мәнін, $\mu_0^2\left(\mathbf{x}^*\left(\mu_R^i, \left\|\mu(\mathbf{x}^*)-\mu^u\right\|_D\right)\right)$,..., $\mu_0^m\left(\mathbf{x}^*\left(\mu_R^i, \left\|\mu(\mathbf{x}^*)-\mu^u\right\|_D\right)\right)$, $i=\overline{2,m}$ – локалды критерийлердің шектіке мәндерін және $\mu_1\left(\mathbf{x}^*\left(\mu_R^i, \left\|\mu(\mathbf{x}^*)-\mu^u\right\|_D\right)\right)$,..., $\mu_L\left(\mathbf{x}^*\left(\mu_R^i, \left\|\mu(\mathbf{x}^*)-\mu^u\right\|_D\right)\right)$ – шектеулердің тиістілік функцияларының максималды мәндерін қамтамасыз ететін $\mathbf{x}\left(\mu_r^i, \left\|\mu(\mathbf{x})-\mu^u\right\|_D\right)$ – кіріс, режимдік (басқару) параметрлері оптисмалды векторын.

Технологиялық нысандардың жұмыс режимдерін басқаруда айқынсыздықта көпкритерийлі шешім қабылдау есептерінің келтірілген жаңа тұжырымдары және оларды шешуге ұсынылған эвристикалық тәсілдері айқын емес жиындар теориясы тәсілдеріне және көпкритерийлі оптимизациялаудың детерминирді әдістерін модификациялауға негізделген. Алынған шешімдер аталған тәсілдердің бастапқы ақпараттың ақын еместігі жағдайларында жалпылануы мен дамуы болып табылады.

Нәтижелерді талқылау. Зерттеу нәтижелерін практикалық қолдану, белгілі нәтижелермен салыстыру және талқылау нәтижелері қарастырайық

Технологиялық объектілердің жұмыс режимдерін басқарудың ұсынылған тәсілін жүзеге асырудың мысалы ретінде Өзен-Самара мұнай құбырының Атырау пунктісіндегі мұнай қыздыру станциясының жұмыс режимдерін басқару үшін шешім қабылдау есебін тұжырымдап, шешу нәтижелерін келтіреміз. Мұнай қыздыру станциясының негізгі міндеті жылыту пештері мен олармен байланысқан агрегаттардың апатсыз және үздіксіз жұмысын қамтамасыз ету және «ыстық» мұнай құбырының оптималды технологиялық жұмыс режимін қамтамасыз ету болып табылады. Бұл кезде келесі критерийлерді оптимизациялау есептерін шешу қажет:

– мұнайды қыздыру мен айдау өзіндік құнын минимизациялау;
– отын мен эксплуатациялақ шығындарды минимизациялау;
– мұнай айдау көлемі мен өнімділікті максимизациялау;
– нысан мен лоның механизмдерінің сенімділігін арттыру;
– нысанның экологиялық қауіпсіздігін арттыру т.б.

Айдалатынмұнайкөлемінтүрліаспаптар,өлшеуішқұралдары(шығын өлшегіштер және т.б.) көрсеткіштерімен анықтауға болады. Біздің жағдайда айдалатын мұнай көлемі [705÷ 725] т/сағ бірлікпен өлшенеді. Нсыанның жұмыс сапасы мен экологиялық қауіпсіздігін бағалауға келетін болсақ, мұндағы жағдай әлдеқайда күрделі болып келеді.

Мұнай құбырының технологиялық-өндірістік кешені жұмысының сапасын, нысанның жұмысының экологиялық қауіпсіздігін бір санмен бағалау өте қиын және әрқашан мүмкін бола бермейді. Көбінесе бұл көрсеткіштер анықсыздықпен, айқынсыздықпен сипатталатыққтан, оларды өлшеу, сандық тұрғыдан анықтау күрделі немесе мүмкін емес болады. Шынайы жағдайда өндірістік нысандардың көптеген жұмыс сапасы көрсеткіштьері, шектелері, экологиялық қауіпсіздіктігі көрсеткіштері көбінесе «кем емес», «артық емес» және «шамамен» сияқты айқын емес шектеулермен сипатталады.

Практикада өндірістік нысанның экономикалық критерийлері (өнімділік, пайда, айдау көлемі және т.б.) және сапа көрсеткіштері максималды, ал оның қоршаған орта экологиясына тигізетін кері әсері минималды болғаны қажет. Алайда бұл экономикалық, экологиялық критерийлер жиі қарама-қайшы келетіні және оларды бір уақытта жақсарту мүмкін емес екендігңі белгілі. Мұндай жағдайда компромисстер облысында тиімді шешімді өндірістік жағдай мен жоспарды және ШҚТ-ны қанағаттандыратын шеші м қабылдвау керек.

Сонымен магистральдық мұнай құбыры қыздыру станциясын тиімді басқару бойынша шешім қабылдау есебін келесідей нақтылауға болады:

$f(\mathbf{x}) = F(f(\mathbf{x})) = \mu_0^i(\mathbf{x}), i = \overline{1,3}$ – мұнай айдау көлемін $\mu_0^1(\mathbf{x})$ мұнай қыздыру пешінің шысындағы температура $\mu_0^2(\mathbf{x})$ мен қысымды $\mu_0^3(\mathbf{x})$ бағалайтын нормалданған локалды критерийлер болсын. Нсыанның жұмыс сапасын, экологиялық көрсеткіштерінсипаттайтын әр айқын емес шектеулерге $\varphi_q(x) \tilde{>} b_q, q = 1,2$ олардың орындалу деңгейін бағалайтын тиістілік функциялары $\mu_{\mathbf{q}}(\mathbf{x}), q = 1,2$ тұрғызылсын. Сондай-ақ локалды критерийлердің басымқылары қатары $I_k = \{1,2,3\}$ мен айқын емес шектелердің салмақ коэффициенттері векторы $\boldsymbol{\beta} = (\beta_1, \beta_2)$ белгілі не анықталады деп қабылданады.

Критерийлер меен шектеулер кіріс, режимдік параметрлеріне $x_i, i = \overline{1,4}$ (x_1– температура, x_2 – қысым, x_3– отын шығыны, x_4– пештің кірісіндегі мұнай көлемі) тәуелді болады. Бұл тәуелділіктер (Оразбаев и др., 2022:71-82., Зыков и др., 2018:208) жұмыстарында келтірілген математикалық модельдер негізінде анықталады.

Бастапқы ақпараттың кейбір бөлігінің айқын еместігі жағдайында формализацияланған мұнай қыздыру станциясының жұмыс режимдерін процесті басқару бойынша шешім қабылдау есебін басты критерий және идеалды нүкте принциптері негізінде (7)–(8) қойылымына сәйкес келесі көпкритерийлі оптимизациялау есебі түрінде жазылуы мүмкін.

$$\max_{x\in X} \mu_0^1(\mathbf{x}), \qquad (7^*)$$

$$X = \left\{\mathbf{x} : \mathbf{x} \in \Omega \wedge (f_i(\mathbf{x}) \tilde{\geq} b_i) \wedge \arg\left(\mu_q(\mathbf{x}) \geq \min_{\mathbf{x}\in\Omega} \left\|\mu(\mathbf{x}) - \mu^u\right\|_D\right) i = 2,3, q = 1,2\right\} \quad (8^*)$$

мұндағы $f_i(\mathbf{x}), i = 2,3$ – мұнай қыздыру станциясы шығысындағы температура мен қысымға қойылған айқын емес шектеулер функциялары; $\left\|\mu(\mathbf{x}) - \mu^u\right\|_D$ – қолданалатын метрика D; $\mu(\mathbf{x}) = (\mu_1(\mathbf{x}), \mu_2(\mathbf{x}))$, $\mu^u = (\max \mu_1(\mathbf{x}), \max \mu_2(\mathbf{x}))$.

Бұл есептің шешімі болып, критерийлердің экстремалдыы мәндерін (басты критерийлердің максималды мәнін, қалған критерийлердің шектеу шарттарын) және айқын емес шектелердің орындалу тиістілік функцияларының максималды мәндерін қамтамасыз ететін, басқару үшін қолданылатын, кіріс, режимдік параметрлерінің оптималды мәндері $\mathbf{x}^* = (x_1^*, x_2^*, x_3^*, x_4^*)$, табылады. Сонымен қатар алынғаен шешім ШҚТ-ны қанағаттандыруы тиіс.

Нақты қойлымы алыынған (7*)–(8*) шешім қабылдау есебін бейімделгшен БК+ИН алгоритмін пайдалана отырып шешеміз.

1) Локалды критерийлердің басымқылары қатары анықталды: $I_k = \{1,2,3\}$ (басты критерий ретінде айдалатын мұнай көлемі таңдалған, сәйкесінше оған 1-басымқы берілген, 2-ші бассымқы пеш шығысындағы температураға, ал 3-ші басымқы пеш шығысындағы қысымға берілген).

2) ШҚТ, эксперттер алынған ақпараттар негізінде терм-жиын анықталып, ір айқын емес шектеге, оның орындалу деңгейін бағалайтын тиістілік функциясы $\mu_q(\mathbf{x})$, $q = 1,2$ тұрғызылған:

$$\mu_1(x) = \exp(0.20 \mid a_1 - 50.0 \mid \cdot 0.5),$$
$$\mu_2(x) = \exp(0.10 \mid a_2 - 80.0 \mid \cdot 0.7),$$

мұндағы a_1, a_2 – мұнай қыздыру станциясының шығысындағы температурасы мен қысымның орташа мәндері.

3) шектеуге ендірілген шектеу функцииялырынын $f_i(x), i = 2,3$ анықтау және $b_i, i = 2,3$ мәндеріен анықтау. (Оразбаев и др., 2022:71-82., Зыков и др., 2018:208) жұмыстары мен зерттеулер нәтижесінде олар келесідей анықталды:

$$f_1(x) = 7 + 1{,}2 \cdot x_1 - 0.25 \cdot x_2 + 5.7 \cdot x_3 - 1.3 \cdot x_4 + 1.8 \cdot x_1^2 + 8.3 \cdot x_3^2;\ b_1 = 55,$$
$$f_2(x) = 0.25 - 1.31 \cdot x_1 + 7.35 \cdot x_2 - 3.1 \cdot x_3 + 2.25 \cdot x_4 + 9.85 \cdot x_2^2 + 8.7 \cdot x_3^2;\ b_2 = 8.5.$$

4) Идеалды нүкте координаттарын анықтау. Бұл нүктелер координаттары ретінде шектеулердің тиістілік функцияларының максималды

мәндері анықталады. Бұл есеп жағдайында тиістілік функциялары нормалды болғандықтан, олар келесідей анылған:$\mu^u = (1,1)$.

5) Выбирается вид метрики Ағымдағы шешім $\mu(\mathbf{x})$ мен идевалды шешім μ^u арысындағы қашықтықты анықтайтын $\|\mu(x)-\mu^u\|_D$ метрикасы түрін таңдау. Қойылған есеп жағдайында метриканың келесі түрі таңдалған:

$$\|\mu(\mathbf{x})-\mu^u\|_E^2 = \sum_{q=1}^{2}\left(\beta_q\left(\max_{\mathbf{x}\in\Omega}\mu_{\mathbf{q}}(\mathbf{x})-\mu_q(\mathbf{x})\right)\right)^2,$$

мұнда β_q – q-ші айқын емес шектеудің салмақ коэффициенті.

6) Модификацияланған математикалық программалау тәсілі негізінде (7*)–(8*) оптимизациялау есебі шешіп, келесі ағымдағы шешімдер анықталған: $\mathbf{x}\left(\sum_{q=1}^{2}\left(\mu_R^i,\beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)$, $i=2,3$ – кіріс, режимдік (басқару) параметрлері векторы; бұл вектор қамтамыз ететін локалды критерийлер мәндері: $\mu_0^1\left(\mathbf{x}\left(\left(\sum_{q=1}^{2}\left(\mu_R^i,\beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)\right)\right)$, $\mu_0^2\left(\mathbf{x}\left(\left(\sum_{q=1}^{2}\left(\mu_R^i,\beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)\right)\right)$, $\mu_0^3\left(\mathbf{x}\left(\left(\sum_{q=1}^{2}\left(\mu_R^i,\beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)\right)\right)$, $i=2,3$ және айқын емес шектеулер тиістілік функциялары әндері $\mu_1\left(\mathbf{x}\left(\sum_{q=1}^{2}\left(\mu_R^i,\beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)\right)$, $\mu_2\left(\mathbf{x}\left(\sum_{q=1}^{2}\left(\mu_R^i,\beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\right.\right.\right.$ $\left.\left.\left.-\mu_q(\mathbf{x})\right)^2\right)\right)$, $i=2,3-$ значения функции принадлежности выполнения ограничении.

7) ШҚТ-ға алынған ағымдағы шешімдерді ұсыну. Егер қсынылған ағымдағы шешімдер ШҚТ-ны қанағаттандырмаса, онда ол щещімді жақсарту мақсатымен $\mu_R^1(\mathbf{x})$, $\mu_R^2(\mathbf{x})$ шектік мәндерін өзгертеді немесе/және $\|\mu(\mathbf{x})-\mu^u\|_D$ метрикасының жаңа түрін таңдайды. Содан кейін жаңа, жақсартылған шешімді анықтау үшін алдыңғы қадамнан бастап шешім қайта есептеледі. Егер алынған ағымдағы шешімдер ШҚТ-ны қанағаттандырса, онда келесі қадамға өту.

8) ШҚТ-ны қанағаттандыратын соңғы еі жақсы шешімдерді шығару: $\mu_0^1\left(\mathbf{x}^*\left(\left(\sum_{q=1}^{2}\left(\mu_R^i,\beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)\right)\right)$, – басты критерийдің максималды мәнін қамтамасыз ететін, ал қалған критерийлердің $\mu_0^2\left(\mathbf{x}^*\left(\left(\sum_{q=1}^{2}\left(\mu_R^i,\beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)\right)\right)$, $\mu_0^3\left(\mathbf{x}^*\left(\left(\sum_{q=1}^{2}\left(\mu_R^i,\beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)\right)\right)$, шектік мән-

дерін қанағаттандыратын және айқын емес шектеулердің тиістілік функцияларының $\mu_1\left(\mathbf{x}^*\left(\sum_{q=1}^{2}\left(\mu_R^i, \beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)\right)$, $\mu_2\left(\mathbf{x}^*\left(\sum_{q=1}^{2}\left(\mu_R^i, \beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)\right)\right)$, максималды мәндерін қамтамасыз ететін кіріс, режимдік параметрлері векторы $\mathbf{x}^*\left(\left(\sum_{q=1}^{2}\left(\mu_R^i, \beta_q(\max_{\mathbf{x}\in\Omega}\mu_q(\mathbf{x})-\mu_q(\mathbf{x})\right)^2\right)\right)\right)$.

Ең жақсы шешімді іздеудің 5-циклінен кейін алвнған ШҚТ-ны қанағаттандыратын соңғы шешімдер 1-ші кестеге ендірілген.

Детерминді тәсілмен ұсынылған тәсілдің 1-кестеде келтірілген нәтижелерін эәрне өндірістік шынайы деректерді салыстыру және талқылау арқылы айқынсыздықта шешім қабылдау есебін шешудің ұсынылған тәсілінің артықшылықтары жайында келесідеуйқорытынды жасауға боладыі:

1) Ұсынылған эвричстикалық тәсіл детерминді тәсілге қарағанда тиімдірек және шынайы эксперименталдық деректерге сәйкестігі жоғары.

2) Ұсынылған эвристикалық алгоритм негізінде оптимизациялау есептерін шешу кезінде идеализациясыз нақты айқынсыз жағдайды толық сипаттайтын қосымша айқын емес ақпарат (ШҚТ, эксперт тәжірибесі, білімі, интеллектісі) қолданылатындықтан, айқын емеса ортада өндірістік есептің шешімінің адекваттылығы артады.

3) Тиімді шешім қабылдаудың көпкритерийлі есебін шешуде қолданылған БК+ИН эвристикалық алгорритмі айқын емес шектеулердің орындалу деңгейлдерін бағалайтын тиістілік функциялары негізінде айқын емес шектеулердің орындалуын қамтамасыз ете алады. Ал лдетермиенді тәсілдәі айқын емес шектеулердің орындалуын бақылау, қамтамасыз ету мүмкін емес.

Кесте 1 – Қойылған шешім қабылдау есебін ұсынылған эвристикалық алгоритм (БК+ИН), детерминді тәсіл (Shumsky et all., 2019:380) арқылы шешу нәтижелерін және өндірістен алынған шынайы деректермен слыстыру

№	Критерийлер мен шектеулердің мәндері	Детерминді тәсіл	Ұсынылған БК+ИН алгоритмі	Өндірістік эксперименталдық мәліметтер
	Айдалған мұнай көлемі (өнімділік), т/сағ, $\widetilde{y}_1$	707	≈ 710	709
	Мұнай қыздыру станциясы шығысындағы температура, y_2, °С	48	50	50
	Мұнай қыздыру станциясы шығысындағы қысым, y_3 кгс/см^2	8.5	8	8.1

	1-айқын емес шектеу орындалуының тиістілік функциясы $\mu_1\left(\mathbf{x}^*(\mu_R^2), \left\|\mu(\mathbf{x}^*)-\mu^u\right\|_D\right)$	–	1.0	–
	2-айқын емес шектеу орындалуының тиістілік функциясы $\mu_2\left(\mathbf{x}^*(\mu_R^2), \left\|\mu(\mathbf{x}^*)-\mu^u\right\|_D\right)$	–	0.98	–
	Кіріс, режимдік (басқару) параметрлері оптималдық мәндері $\mathbf{x}^*=(x_1^*, x_2^*, x_3^*, x_4^*)$: x_1^* – пеш кірісіндегі температура, °С;	35	33	34
	x_2^* – пеш кірісіндегі қысым, кгс/см^2;	10,5	9.8	10
	x_3^* – пеш кірісіндегі отын көлемі, кг/сағ;	27	25	26
	x_4^* – пеш кірісіндегі мұнай көлемі, т/сағ;	710	710	710

Ескертпе: (–) сәкес тәсілмен бұл көрсеткіштің анывқталмайтындығын білдіреді; Салыстырылған тәсілдерде шешімді іздеу уақыты жуықша бірдей.

Алынған нәтижелер мен қорытындылардың сенімділігі мыналармен расталады: шешім қабылдау және оптимизациялау теорияларының, айқын емес жиындар теориясының, эксперттік бағалау тәсілдерінің ғылыми ережелеріне негізделген пайдаланылған зерттеу тәсілдерінің дұрыстығымен; есептеу-модельдік (теориялық) және эксперименттік (пилоттық-өнеркәсіптік) зерттеу нәтижелерінің жеткілікті деңгейде сәйкес келулерімен.

Қорытынды. Ғылыми мақалада түрлі оптималдық принциптерін комбинациялау және айқынсыздыққа модификациялау негізінде айқын емес бастапқы ақпарат жағдайында магистральдық мұнай құбырының технологиялық агрегаттарының жұмыс режимдерін басқаруғ бойынша шешім қабылдау есептерінің жаңа тұжырымдары алынған және қойылымдары келтірілген есептерді айқын емес орталда шешу үшін тиімді эвристикалық тәсілдері жасақталған. Ұсынылған эвристикалық тәсілдер айқын емес ортада жұмыс істеу үшін түрлі оптималдық пр инциптерін (Парето оптималдық, басты критерий, идеалды нүкте) модификациялауға және комбинациялауға негізделген. Өзен-Самара мұнай құбырының Атырау пунктісіндегі мұнай қыздыру станциясының жұмыс режимдерін басқару және оптимизациялауда ұсынылған айқын емес тәсілдеме (басты критерий – критерийлер үшін, және

идеалды нүкте – критерийлер үшін, принциптерін модификациялау арқылы) практикада жүзеге асырылған. Детерминделген тәсілгмен салыстырғанда ұсынылып, қолданылған эвристикалық алгоритмнің тиімділігі мен адекваттығы жоғары екені көрсетілген.

Зерттеу нәтижелерінің ғылыми жаңашылдығы – айқын емес есеп, оларды шешудің белгілі тәсілдеріндегідей алдын ала детерминирленген есепкеа түрлендірілмей, айқын емес ортада қойылып, шешіледі. Бұл жинақталған айқын емес ақпаратты толық пайдалануды және айқын емес бастапқы ақпаратпен сипатталатын күрделі өндірістік есептің неғұрлым адекватты шешімін алуды қамтамасыз етеді. Алынған нәтижелер көпкритерийлі және айқын емес бастапқы ақпаратпен сипатталатын күрделі өндірістік объектілерді оптимизациялау және олардың жұмыс режимдерін басқару тәсілдерін дамытуға, қолдану аясын кеңейтуге мүмкіндік береді.

Жұмыстың практикалық маңыздылығы дәстүрлі детерминирді немесе стохастикалық математикалық тәсілдермен шешілуі күрделі немесе мүмкін емес көпкритерийлікпен және айқынсыздықпен сипатталатын күрделі өндірістік есептерді тиімді шешумен анықталады. Сонымен қатар, айқын емес ортада шешім қабылдау есебінің шешудің ұсынылған тәсілдемесі өндірістік жағдайға және әртүрлі сипаттағы бастапқы ақпараттың болуына байланысты ШҚТ-ға ұсынылған алгоритмдер жиынтығынан есепті шешудің неғұрлым қолайлы алгоритмін таңдау мүмкіндігі беріледі.

Бұл бағыттағы әрі қарай ғылыми әзірлемелердің перспективті бағыты өндірістік процестерді автоматтандырудың әртүрлі жүйелеріне, мысалы шешім қабылдауды қолдау интеллектуалды жүйелері, компьютерлік басқару жүйелері және т.б., программалық қамтамасыз етуді әзірлеумен байланысты.

Зерттеуді Қазақстан Республикасы Білім және ғылым министрлігінің Ғылым комитеті қаржыландырады (грант № AP08855680 – Каталитикалық риформинг қондырғысы жұмыс режимдерін басқару үшін шешім қабылдаудың интеллектуалдандырылған жүйесі).

Information about the authors:

Moldasheva Zh.Zh. – doctoral student of the Department of Information Systems, L.N. Gumilyov Eurasian National University, st. Satpaeva 2A, Nur-Sultan, Kazakhstan. E-mail: *zhadira1985@mail.ru*;

Orazbayev B.B. – doctor of technical sciences, academician of the

Engineering academy of the Republic of Kazakhstan, professor of the department of System analysis and Control, L.N. Gumilyov Eurasian National University, Satpayev str. 2A, Nur-Sultan, Kazakhstan. E-mail: *batyr_o@mail.ru*;

Assanova B.U. – PhD, Dean of the Faculty of Physics and Mathematics, Kh. Dosmukhamedov Atyrau University, st. Students 112, Atyrau, Kazakhstan. E-mail: *baha1981_13@mail.ru*;

Iskakova S.Sh. – Candidate of Technical Sciences, Associate Professor, Dean of the Faculty of Information Technology, S. Utebaev Atyrau University of Oil and Gas, st. Baymukhanov 45a, Atyrau, Kazakhstan. E-mail: iskakova_sh@mail.ru;

Orazbayeva K.N. – doctor of technical sciences, professor of the department of management, Esil University, Zhubanov str. 7, Nur-Sultan, Kazakhstan. E-mail:kulman_o@mail.ru.

ӘДЕБИЕТТЕР:

Алиев Р.А., Церковный А.Э., Мамедова Г.А. (2017) Управление производством при нечеткой исходной информации. М.: Энергоатомиздат, 2-изд. 378 с.

Оразбаев Б.Б. (2010) Математические методы оптимального планирования и управления производством. -Алматы: Ғылым, 307 с.

Kuz'min V.B., Travkin S.I. (2012) The theory of fuzzy sets in control problems and in principles for organizing fuzzy processors. A survey of foreign literature // Automation and Remote Control, 53(11). P. 1649–1678. (in Eng.).

Емельянов С.В., Ларичев О.И. (2016) Многокритериальные методы принятия решений. -М.: Знание, 88 с.

Groppen V.O. (2018) Principles of reference-aided decision making // *Automation and Remote Control*. 67(4). P. 660-675. (in Eng.).

Grebenyuk G.G. (2016) Mathematical modeling as a decision tool in the control of urban heat supply //*Automation and Remote Control*. 58(5). P. 805-812. (in Eng.).

Dimitriadi G.G., *Larichev* O.I. (2017) Decision Support System and the ZAPROS-III Method for Ranking the Multiattribute Alternatives with Verbal Quality Estimates // *Automation and Remote Control*. 63(8). P. 1322-1335. (in Eng.).

Orazbayev B., Moldasheva Zh., Orazbayeva K., Makhatova V., Kurmangaziyeva L., Gabdulova A. (2021) Development of mathematical models and optimization of operation modes of the oil heating station of main oil pipelines under conditions of fuzzy initial information. Eastern-European Journal of Enterprise Technologies, 6: 2(114)), 147–162. (in Eng.).

Орловский С.А. (2018) Проблемы принятия решений при нечеткой исходной информации. -М.: Наука 2-изд. 287с.

Зайченко Ю.П. (2019) Исследование операций: нечеткая оптимизация. -Киев: Выща школа, 3-изд. 355 с.

Rademaker M., Bernard B. (2011) Aggregation of monotone reciprocal relations with

application to group decision making // Fuzzy Sets and Systems. 184(3). P. 29–51. (in Eng.).

Рыжов А.П. (2017) Теория нечетких множеств и ее приложений. – М.: МГУ. 115 с.

Liu Y., Rodríguez R.M., Martínez L. (2022) Interval Type-2 Fuzzy Envelope of Proportional Hesitant Fuzzy Linguistic Term Set: Application to Large-Scale Group Decision Making. Mathematics, 10, 2368. https://doi.org/10.3390/math10142368. (in Eng.).

Markovskii A.V. (2018) Solution of Fuzzy Equations with Max-Product Composition in Inverse Control and Decision Making Problems // *Automation and Remote Control*. 64(9). P. 1486-1495. (in Eng.).

Оразбаев Б.Б., Мухамбеткалиева А.К. (2007) Задачи и методы многокритериального выбора оптимальных режимов работы объектов нефтепровода. Алматы: Эверо, 138 с.

Оразбаев Б.Б., Молдашева Ж.Ж, Ла Л.Л., Оразбаева К.Н. Тулеуов Ж.Н., Утенова Б.Е. (2022) Разработка моделей станции подогрева нефти магистральных нефтепроводов в условиях нечеткости исходной информации. // Вестник Национальной инженерной академии Республики Казахстан. 1(83). -С.71-82.

Баронец В.Д., Гречихин М.А. (2019) Построения и представления функции принадлежности в экспертных системах // Техническая кибернетика 6(3). 156 –160.

Larichev O.I. (2002) Properties of the Decision Methods in the Multicriteria Problems of Individual Choice //*Automation and Remote Control*. 63 (2). P. 304-315. (in Eng.).

Sabzi H.Z. (2017) Developing an intelligent expert system for streamflow prediction, integrated in a dynamic decision support system for managing multiple reservoirs: a case study // Expert systems with applications. 82(3) – P. 145–163. (in Eng.).

Муханбеткалиева А.К., Оразбаев Б.Б. (2018) Проблемы математического моделирования технологического комплекса магистральных нефтепроводов и подходы к их решению // Поиск. 4. С. 229 – 235.

Зыков В.В. (2018) Математическое моделирование и оптимизации процессов сбора, подготовки и транспортировки нефти и газа. Тюмень: ТПУ. 2-изд 208 с.

Шумский В.М., Зырянова Л.А. (2019) Инженерные задачи в нефтепереработке и нефтехимии. -М.: Химия, 3-изд. 380 с.

REFERENCES:

Aliev R.A., Tserkovny A.E., Mamedova G.A. (2017) Production management with fuzzy initial information. Moscow: Energoatomizdat, 2nd ed. 378 p. (in Russ.).

Orazbaev B.B. (2010) Mathematical methods for optimal planning and production management. -Almaty: Gylym, 307 p. (in Russ.).

Kuz'min V.B., Travkin S.I. (2012) The theory of fuzzy sets in control problems and in principles for organizing fuzzy processors. A survey of foreign literature // Automation and Remote Control, 53(11). P. 1649–1678. (in Eng.).

Emelyanov S.V., Larichev O.I. (2016) Multicriteria decision making methods. -M.: Knowledge, 88 p. (in Russ.).

Groppen V.O. (2018) Principles of reference-aided decision making // *Automation and Remote Control*. 67(4). P. 660-675. (in Eng.).

Grebenyuk G.G. (2016) Mathematical modeling as a decision tool in the control of urban heat supply //*Automation and Remote Control*. 58(5). P. 805-812. (in Eng.).

Dimitriadi G.G., *Larichev* O.I. (2017) Decision Support System and the ZAPROS-

III Method for Ranking the Multiattribute Alternatives with Verbal Quality Estimates // *Automation and Remote Control*. 63(8). P. 1322-1335. (in Eng.).

Orazbayev B., Moldasheva Zh., Orazbayeva K., Makhatova V., Kurmangaziyeva L., Gabdulova A. (2021) Development of mathematical models and optimization of operation modes of the oil heating station of main oil pipelines under conditions of fuzzy initial information. Eastern-European Journal of Enterprise Technologies, 6: 2(114)), 147–162. (in Eng.).

Orlovsky S.A. (2018) Problems of decision making with fuzzy initial information. -M.: Science 2nd ed. 287 p. (in Russ.).

Zaichenko Yu.P. (2019) Operations research: fuzzy optimization. -Kiev: Vyscha school, 3rd ed. 355 p. (in Russ.).

Rademaker M., Bernard B. (2011) Aggregation of monotone reciprocal relations with application to group decision making // Fuzzy Sets and Systems. 184(3). P. 29–51. (in Eng.).

Ryzhov A.P. (2017) Fuzzy set theory and its applications. – Moscow State University. 115 p. . (in Russ.).

Liu Y., Rodríguez R.M., Martínez L. (2022) Interval Type-2 Fuzzy Envelope of Proportional Hesitant Fuzzy Linguistic Term Set: Application to Large-Scale Group Decision Making. Mathematics, 10, 2368. https://doi.org/10.3390/math10142368. (in Eng.).

Markovskii A.V. (2018) Solution of Fuzzy Equations with Max-Product Composition in Inverse Control and Decision Making Problems // *Automation and Remote Control*. 64(9). P. 1486-1495. (in Eng.).

Orazbaev B.B., Mukhambetkalieva A.K. (2007) Tasks and methods of multi-criteria selection of optimal operating modes for oil pipeline facilities. Almaty: Evero, 138. (in Russ.).

Orazbaev B.B., Moldasheva Zh.Zh., La L.L., Orazbaeva K.N. Tuleuov Zh.N., Utenova B.E. (2022) Development of models of an oil heating station for main oil pipelines in conditions of fuzzy initial information. // Bulletin of the National Engineering Academy of the Republic of Kazakhstan. 1(83). -P.71-82. (in Russ.).

Baronets V.D., Grechikhin M.A. (2019) Constructions and representations of the membership function in expert systems // Technical Cybernetics 6(3). P. 156–160. (in Russ.).

Larichev O.I. (2002) Properties of the Decision Methods in the Multicriteria Problems of Individual Choice //*Automation and Remote Control*. 63 (2). P. 304-315. (in Eng.).

Sabzi H.Z. (2017) Developing an intelligent expert system for streamflow prediction, integrated in a dynamic decision support system for managing multiple reservoirs: a case study // Expert systems with applications. 82(3) – P. 145–163. (in Eng.).

Mukhanbetkalieva A.K., Orazbaev B.B. (2018) Problems of mathematical modeling of the technological complex of main oil pipelines and approaches to their solution // Search. 4. P. 229–235. (in Russ.).

Zykov V.V. (2018) Mathematical modeling and optimization of oil and gas gathering, treatment and transportation processes. Tyumen: TPU. 2nd edition 208 p. (in Russ.).

Shumsky V.M., Zyryanova L.A. (2019) Engineering tasks in oil refining and petrochemistry. -M.: Chemistry, 3rd ed. 380 p. (in Russ.).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 185-197
https://doi.org/10.32014/2022.2518-1726.146
УДК 621.7.08

А.Б. Миненбаева[1*], А.С. Аканова[2]

[1]Астана халықаралық университеті, Қазахстан, Астана;
[2]С.Сейфуллин атындағы Қазақ агротехникалық университеті,
Қазахстан, Астана.
E-mail: *aigulka79_79@mail.ru*

СОЛТҮСТІК ҚАЗАҚСТАН ОБЛЫСЫНЫҢ АУЫЛШАРУАШЫЛЫҒЫ ДАҚЫЛДАРЫНЫҢ КҮЙІН NDVI СЫЗЫҚТЫҚ ТРЕНДТЕРІ АРҚЫЛЫ ЗЕРТТЕУ

Аннотация. Мақалада Солтүстік Қазақстан облысында өсірілетін ауылшаруашылығы дақылдарының үш жыл аралығындағы NDVI (Normalized Difference Vegetation Index) маусымдық вегетациялық индекс уақыт қатарларының трендтері зерттелген. Солтүстік Қазақстан республиканың агроөнеркәсіп кешенінің экономикалық маңызды аймақтарының бірі болып табылады. Мұнда ауылшаруашылық дақыл дарының негізгі тауарлық өндірісі шоғырланған. Дақылдардың өнімділігіне анықтайтын жағдайларды білу оны басқаруға, әрі оңтай ландыруға мүмкіндік береді. Мақалада қарастырылған аймақ-тағы 2018-2020 жылдар аралығындағы ауылшаруашылығы дақыл-дарының өнімділік кестесі, EOS Land Viewer геоақпараттық жүйесі арқылы алынған NDVI тарату карталары берілген. Сонымен қатар, осы жылдардағы вегетативті кезеңдеріндегі ауыл шаруашылығы дақылдарының NDVI динамикасы зерттелген.

Ғылыми зерттеудің жарияланған тақырыбының өзектілігі Солтүстік Қазақстан облысында ауылшаруашылығы дақылдарының өнімділігін арттыру қажеттілігінен туындайды. Ауылшаруашылық дақылдарының күйін анықтайтын негізгі факторларының бірі әртүрлі ұзындықты толқын дардың сәулеленуімен сипатталатын спектрлі шағылысу мүмкіндігі болады.

Вегетациялық индекс мәндері Landsat 8 жерсерігі кескіндері негізінде алынған. NDVI өсімдік жамылғысының индексінің уақыт қатарындағы сызықтық тенденциялардың коэффициенттерін талдау негізінде өзгерістерді бағалаудың әдістемесі мен нәтижелері ұсынылған. Өсімдіктің күйін қашықтықтан бағалау әдетте, NDVI (Normalized Difference Vegetation Index) вегетациялық индекстері арқылы жүзеге асады. Бұл индекс өсімдіктің қызыл және инфрақызыл аралығында сәулеленуін бағалау негізінде есептелінеді. NDVI шамасы өсімдіктің тығыздығына, оның даму фазасына, жапырақ бетінің ауданына және т.б. факторларға тәуелді болады.

Жүргізілген зерттеулерге сәйкес, егістік алқаптарына тән көрсеткіштің маусымдық динамикасы мамыр айының басынан маусымның бірінші жартысына дейін созылатын үздіксіз өсу кезеңімен сипатталатыны анықталды. Сонымен қатар, тамыз айынан қыркүйек айына дейінгі кезеңде күздік егістік алқаптарында индекс мәндерінің төмендеу қарқыны байқалады. NDVI коэффициенттерінің минималды вариация коэффициенттерінің мәндері (5-6%) 16 күн аралықпен маусымның ортасында болған. Солтүстік Қазақстан облысының 2018-2020 жылдар аралығындағы NDVI мәндерін зерттеу нәтижесінде вегетациялық кезеңдегі мәндерінің ауа-райына, ылғалдылыққа және егіс алқаптарының өсімдік жамылғысындағы антропогендік факторларға тәуелді болатыны анықталды.

Түйін сөздер: жерді қашықтықтан барлау**,** NDVI вегетациялық индексі, EOS Land Viewer, ауылшаруашылығы дақылдарының өнімділігі, уақыт қатарлары, геоақпараттық жүйе.

А.Б. Миненбаева[1*], А.С. Аканова[2]

[1]Международный университет Астана, , Казахстан, Астана;
[2]Казахский агротехнический университет им. С.Сейфуллина,
Казахстан, Астана.
E-mail: *aigulka79_79@mail.ru*

ИССЛЕДОВАНИЕ СОСТОЯНИЯ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР СЕВЕРО-КАЗАХСТАНСКОЙ ОБЛАСТИ ПО ЛИНЕЙНЫМ ТРЕНДАМ NDVI

Аннотация. В статье рассмотрены тренды временных рядов сезонного вегетационного индекса NDVI (Normalized Difference Vegetation

Index) за три года сельскохозяйственных культур, выращиваемых в Северо-Казахстанской области. Северный Казахстан – один из наиболее экономически важных регионов агропромышленного комплекса Республики. Здесь сосредоточено основное товарное производство зерна яровой мягкой пшеницы. Знание теории процессов формирования урожая дает возможность управлять ими и, в конечном счете, оптимизировать.

Определяющим признаком сельскохозяйственной культуры и ее состояния является спектральная отражательная способность, характеризующаяся широким диапазоном в отражении излучения разных длин волн. Значения индекса растительности основаны на спутниковых снимках Landsat 8. Представлены методика и результаты оценки изменений на основе анализа коэффициентов линейных трендов временных рядов вегетационного индекса NDVI.

Дистанционную оценку состояния растений обычно проводят с использованием индексов вегетации NDVI (Normalized Difference Vegetation Index). Этот индекс рассчитывается на основе оценки излучения растений между красным и инфракрасным. Величина NDVI зависит от густоты растения, фазы его развития, площади листовой поверхности и др. зависит от факторов. Установлено, что сезонная динамика пашни характеризуется периодом непрерывного роста с начала мая до первой половины июня. Кроме того, в период с августа по сентябрь наблюдается темп снижения значений индекса по озимым посевам. Значения минимальных коэффициентов вариации коэффициентов NDVI (5-6%) наблюдаются в середине июня с интервалом в 16 дней. Изучение значений NDVI в Северо-Казахстанской области за 2018-2020 годы выявило, что значения вегетационного периода зависят от погодных, влажностных и антропогенных факторов в растительности пашни.

Ключевые слова: вегетационный индекс NDVI, дистанционное зондирование земли, Landsat, урожайность, EOS Land Viewer.

A.B. Mimenbayeva[1*], A.C. Akanova[2]

[1]Astana International University, Kazakstan, Astana;
[2]S. Seifullin Kazakh Agrotechnical University, Kazakstan, Astana.
E-mail: *aigulka79_79@mail.ru*

RESEARCH OF THE STATE OF AGRICULTURAL CROPS NORTH KAZAKHSTAN REGION ACCORDING TO LINEAR NDVI TRENDS

Abstract. The article considers the trends in the time series of the seasonal vegetative index NDVI (Normalized Difference Vegetation Index) for three years of crops grown in the North Kazakhstan region. Northern Kazakhstan is one of the most economically important regions of the agro-industrial complex of the Republic. The main commodity production of spring soft wheat is concentrated here. Knowledge of the theory of crop formation processes makes it possible to manage them and, ultimately, optimize them. The defining feature of an agricultural crop and its condition is the spectral reflectivity, which is characterized by a wide range in the reflection of radiation of different wavelengths. Vegetation index values are based on Landsat 8 satellite images. Methods and results of changes assessment based on the analysis of coefficients of linear trends in the NDVI vegetation index time series are presented.

Remote assessment of the state of plants is usually carried out using the NDVI (Normalized Difference Vegetation Index) vegetation indices. This index is calculated on the basis of an estimate of plant radiation between red and infrared. The value of NDVI depends on the density of the plant, the phase of its development, the area of the leaf surface, etc. depends on factors. It has been established that the seasonal dynamics of arable land is characterized by a period of continuous growth from the beginning of May to the first half of June. In addition, in the period from August to September, there is a rate of decline in the values of the index for winter crops.

The values of the minimum coefficients of variation of the NDVI coefficients (5-6%) are observed in mid-June with an interval of 16 days. The study of NDVI values in the North Kazakhstan region for 2018-2020 revealed that the values of the growing season depend on weather, humidity and anthropogenic factors in arable land vegetation.

Key words: time series trends, NDVI vegetation index, remote sensing data, Landsat, yield, EOS Land Viewer.

Кіріспе. Жерді қашықтықтан барлау (ЖҚБ) деректерінің қарқынды дамуы соңғы онжылдықтарда ауыл шаруашылығы дақылдары егістерін жедел мониторингіне жаңа мүмкіндіктер ашты.

Спутниктік өлшеу құралдарының қарқынды дамуы және ЖҚБ жерсеріктері топтамасының кеңеуі, ішкі шаруашылық жерге орналастыру схемаларын құру және нақтылау, ауыл шаруашылығы дақылдарын және пайдаланылмайтын жерлерді анықтау тәрізді ауыл шаруашылығы саласындағы алуан түрлі міндеттерді шешуге мүмкіндік береді. Егістіктегі дәнді-дақылдардың жай-күйін және олардың өнімділіктерін бағалауды спутниктік кескіндерді талдау арқылы алынатын вегетациялық индекстер арқылы жүзеге асыруға болады.

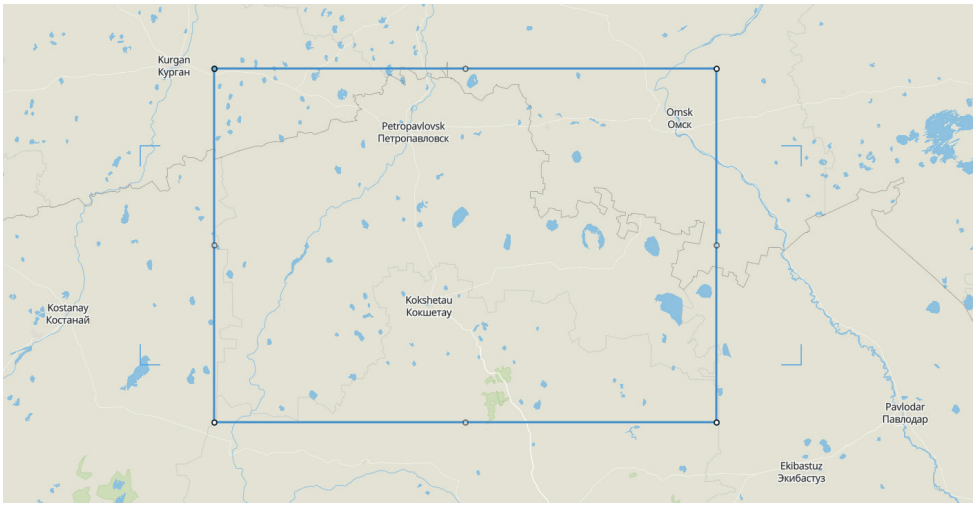
Өсімдік жамылғысының фитомассасын, өнімділігін, динамикасын бағалау үшін өсімдік жамылғысының индекс көрсеткіштері және оның әртүрлі модификациялары қолданылады. Индекстердің дерекқорында (Jinru және т.б., 2017) вегетациялық көрсеткіштің 250-ден астам сорттары бар. Ауылшаруашылық есептерін шешуде ең қарапайымы, әрі қолдануға ыңғайлысы индекстің нормаланған айырмашылық индексі (NDVI) болып табылады. Маусымдық NDVI деректерінің уақытша қатарын талдау ауылшаруашылығы дақылдарының әртүрлі түрлері үшін фитомассаны бағалауға, жердің өсімдік жамылғысының фенологиялық заңдылықтарын бақылауға, вегетациялық және құрғақшылық кезеңдерінің ұзақтығын бағалауға мүмкіндік береді (Gao және т.б., 2020).

Жалпы, өсімдік жамылғысының қалыптасуына әсер ететін басты факторлардың бірі-ауа райы. Жаһандық және аймақтық деңгейлерде жиырма жылдық кезеңдегі қыс пен көктемгі температураның жоғарылауы өсімдік жамылғысының өнімділігін арттыруда да, вегетациялық кезеңнің ұзаруына да оң әсерін тигізетіні дәлелденген. Құрғақ аймақтардағы өсімдік жағдайының жыл аралық динамикасын анықтайтын екінші маңызды фактор – жауын-шашын факторы. Жарияланған зерттеулер вегетациялық жағдайдың ағымы мен вегетациялық кезеңдегі гидротермиялық фактордың ағымы арасындағы тығыз байланысты көрсетеді (Igor және т.б., 2016).

Бұл зерттеудің негізгі мақсаты - 2018-2020 жылдар аралығындағы Солтүстік Қазақстан облысының ауылшаруашылығы дақылдарының NDVI индексінің маусымдық және жыл аралық динамикасын зерттеу және талдау.

Индекстік кескіндерді алу және талдау үшін, тесттілік аймақ ретінде Көкшетау облысының 132,5 м^2 полигоны алынды.

Материалдар және әдістеме. Солтүстік Қазақстан облысы Қазақстан Республикасының солтүстігінде, Батыс Сібір жазығының оңтүстік шетінде орналасқан, Ресей Федерациясымен шекаралас жерді алып жатыр. Облыстың ауданы 98 мың км², оның 58,8 мың км² ауыл шаруашылығы жерлері болып табылады, бұл облыс аумағының 60% құрайды (Dzhalankuzov және т.б., 2016). Солтүстік Қазақстан облысының ауыл шаруашылығы облыстың жетекші саласы болып табылады. Облыс Еуразия астық белдеуінің орталығында орналасқан және кең егістік және жайылымдық жерлерге ие. Егістік жерлерінде жаздық бидай, арпа, сұлы, жүгері, бұршақ тәрізді ауылшаруашылығы дақылдары өседі. Осы дақылдардың 2018-2020 жж. аралығындағы күйін зерттеу үшін, EOS Land Viewer жүйесі арқылы алынған 08 мамырдан 01 тамызға дейінгі аралықтағы Landsat 8 деректерінің бұлтсыз композиттері және NDVI вегетациялық индекстері пайдаланылды (Nigam, т.б., 2012).

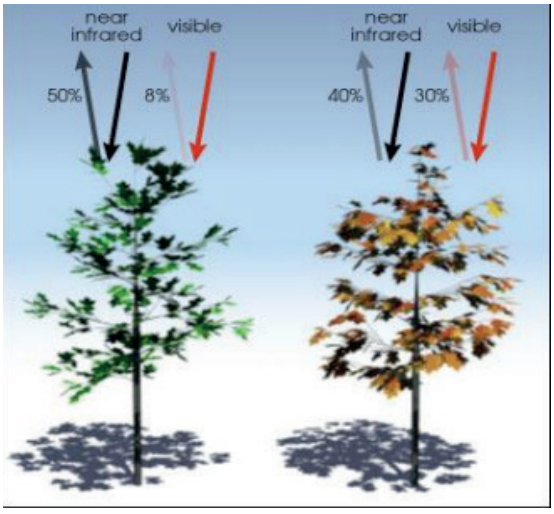


1 сурет - Land Viewer жүйесінде Солтүстік Қазақстан облысының тәжірибе аймағы

Landsat 8 жерсерігі бір нүктеде 15-тен 100 метрге дейінгі кескіндердің кеңістіктік рұқсатымен кескіндерді қабылдайды. Бұл жерсерікте екі құралдар жиынтығы бар: операциялық жерді бейнелеу құрылғысы (OLI) және жылу инфрақызыл сенсоры (TIRS). Бірінші жиынтық 9 көрінетін жарық диапазонында және жақын-ИҚ диапазонында, екінші жиын - алыс (жылу) ИҚ-ның екі диапазонында суретке түсіреді. Осы арналар бойынша Солтүстік Қазақстан облысының белгіленген

аймағына NDVI индекстері EOS Land Viewer жүйесінде есептелді (1 сурет).

NDVI (Normalized Difference Vegetation Index) – фотосинтетикалық белсенді биомасса мөлшерінің қарапайым сандық өлшемі (Pradhan және т.б., 2018). Бұл индекс -1-ден 1-ге дейінгі мәндерді қабылдай алады. Өсімдік үшін NDVI индексі 0,2-ден 0,9-ға дейін оң мәндерді қабылдайды. Өсімдіктердің пайда болуымен (вегетациялық кезеңде) өсімдік биомассасының өсуі NDVI мәндерінің жоғарылауына сәйкес келеді. Белсенді вегетациялық кезеңде NDVI мәндерінің төмендеуі дақылдардың күйзеліс жағдайын көрсетеді. NDVI индекстері көбінесе құрғақшылықты бақылау, өнімділікті болжау үшін қолданылады (Plotnikov, т.б., 2018).



2 сурет- Жасыл және қуарған өсімдіктің шағылысып көріну және жақын инфрақызыл сәулелерін салыстыру.

NDVI индексі төмендегі формула арқылы есептеледі:

$$NDVI = \frac{NIR - RED}{NIR + RED} \qquad (1)$$

мұндағы NIR – спектрдің жақын инфрақызыл аймағының шағылысы, ал RED спектрдің қызыл аймағының шағылысы (2 сурет). Формула бойынша кескіннің белгілі бір нүктесінде өсімдіктердің тығыздығы қызыл және инфрақызыл диапазондардағы шағылысқан жарық қарқындылығының айырмашылығына олардың қарқындылық қосындысына бөлгенге тең.

Маусымның басында NDVI индексі бойынша, өсімдіктің қалай қыстағанын түсінуге болады (Kataev және т.б., 2016:а).

1. Егер NDVI 0,15-тен төмен болса, аймақтағы барлық өсімдіктер өліп қалған болуы мүмкін. Әдетте мұндай көрсеткіштер вегетациясыз жыртылған топыраққа сәйкес келеді.

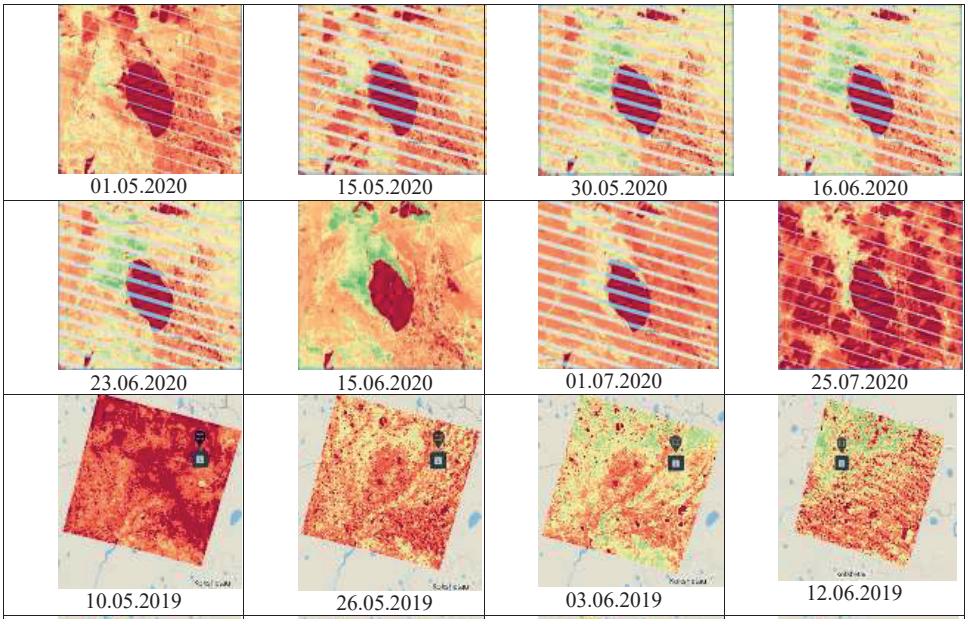
2. 0,15−0,2 - төмен көрсеткіш. Бұл өсімдіктердің қыстауға ерте фенологиялық кезеңінде, өңдеуден бұрын кіргенін көрсетуі мүмкін.

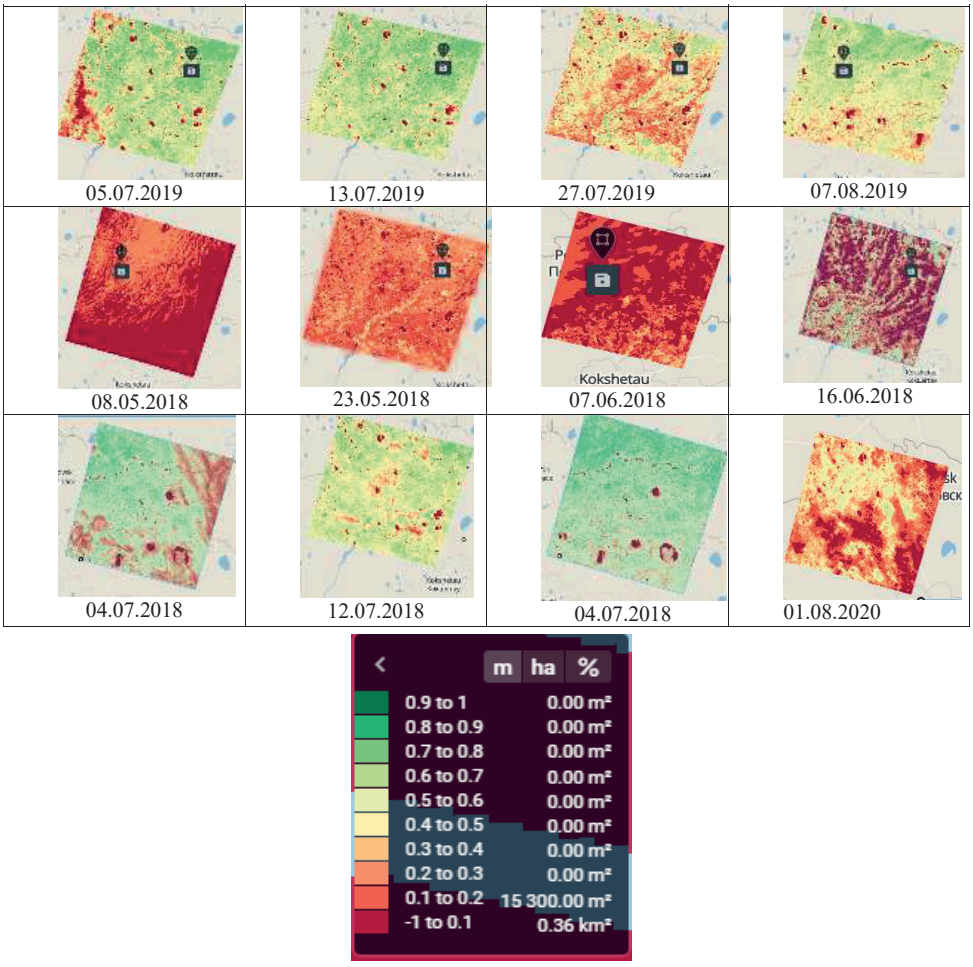
3. 0,2−0,3 - салыстырмалы жақсы көрсеткіш. Өсімдіктер вегетациялық өңдеу кезеңіне өткенін білдіреді.

4. 0,3–0,5- жақсы көрсеткіш. Егер спутниктік сурет өсімдіктердің қалпына келуіне дейін алынған болса, онда өсімдіктердің өсуі мен дамуы кезеңі басталғаннан кейін аймақтың жағдайын қайтадан талдау қажет болады.

Вегетациялық кезеңдегі NDVI индексі мәндерінің өзгеруін зерттеу үшін Солтүстік Қазақстан облысының тәжірибе полигоны белгіленді. Осы полигондағы егістік жерлердің Landsat бұлтсыз композиттері EOS Land Viewer геоақпараттық жүйесі арқылы алынып, 12-15 күндік NDVI мәндері қарастырылды. Нәтижесінде 2018-2020 жылдарға сәйкес келетін NDVI индекстерінің 24 динамикалық сериясы құрылды.

Солтүстік Қазақстан облысының сынақ полигонына жүргізілген зерттеулер барысында түсірілімнің әрбір күніне NDVI тарату карталары алынды (3 сурет).





3 сурет - Вегетациялық индекстердің мәндеріне сәйкес өсімдіктердің жай-күйі.

Талқылау және нәтижелер. Вегетациялық кезең ауылшаруашылығы дақылдарының өсуі мен дамуын көрсететететін басты факторлардың бірі болып табылады. Солтүстік Қазақстан облысында вегетациялық кезең мамыр айының ортасынан басталып, тамыз айының екінші онкүндігіне дейін созылады. NDVI индекстері арқылы бағаланған дақылдардың күйі вегетациялық даму кезінде айтарлықтай өзгереді. Сонымен, вегетациялық кезең басталған сәттен бастап, ауылшаруашылығы дақылдары жасыл биомасса жинайды және индекс мәні артады; маусым айының аяғында - шілдеде биомасса максимумға жетеді, содан кейін NDVI мәндерінің тұрақтануы және тіпті төмендеуі байқалады (Kataev, т.б., 2018:b).

(Plotnikov және т.б., 2018), (Voronina және т.б., 2018), (Shurr және

т.б., 2018), (Kataev және т.б., 2016:a) және (Kataev және т.б., 2018:b) зерттеулерінде, ауыл шаруашылығы дақылдарының өнімділігінің жылдық ауытқулары вегетация кезеңінде NDVI индекстері көмегімен дәл болжанатыны дәлелденген.

NDVI арқылы бағаланған өсімдіктердің күйі өсімдіктердің вегетативті дамуы кезінде айтарлықтай өзгереді . Landsat-8 спутнигінен алынған жерді қашықтықтан зондтау деректері бойынша есептелген вегетациялық индексті пайдалана отырып, Солтүстік қазақстан облысы аумағындағы өсімдіктер жай-күйінің уақытша-аумақтық динамикасының заңдылықтары анықталды (1 кесте).

1 кесте - Солтүстік Қазақстан облысындағы 2018-2020 жж. аралығында өскен ауылшаруашылық дақылдарының NDVI нормаланған индекс мәндері (1 формула бойынша)

Дата	2018	2019	2020
10.05	0,17	0,20	0,29
24.05	0,24	0,35	0,48
07.06	0,33	0,43	0,52
16.06	0,42	0,46	0,51
05.07	0,47	0,51	0,51
15.07	0,45	0,51	0,50
27.07	0,48	0,50	0,46
07.08	0,50	0,46	0,46
14.08	0,49	0,44	0,50
30.08	0,46	0,38	0,32

Уақыт бойынша трендтерге қарасақ, NDVI индекстері қарастырылып отырған кезеңде мамырдан шілдеге дейін артқанын, содан кейін төмендеу үрдісін ұстанғанын көреміз (4-сурет).

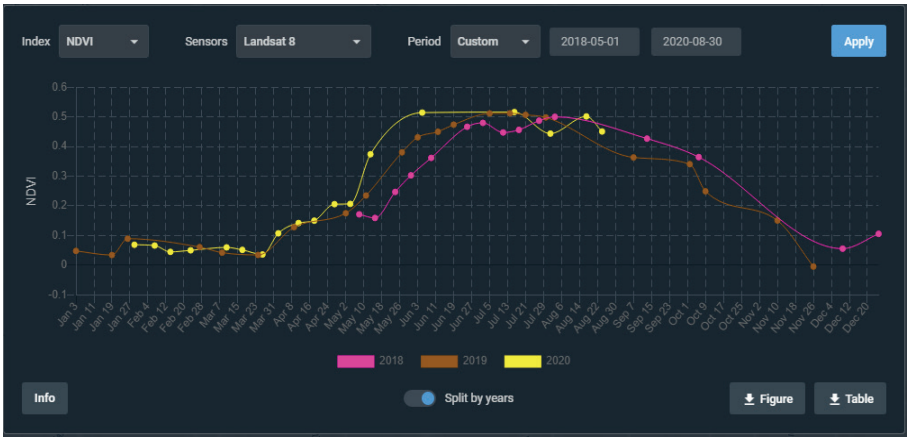
Жоғарыда айтылғандай, NDVI нормаланған индекс мәндері өсімдіктің жамылғысының тығыздығына, оның даму фазасына, ауа-райы және т.б. факторларға тәуелді болады (Bashirova және т.б., 2019).

Ауыл шаруашылығы дақылдары үшін олардың белсенді вегетациялық кезеңіндегі жауын-шашынның (егуден пісіп-жетілуіне дейін) маңызы зор. Жылдың суық мезгіліндегі жауын-шашынның да маңызы зор, ол топырақтағы ауыл шаруашылығы дақылдарын себу кезеңіндегі көктемгі ылғал қорының мөлшерін анықтайды.

Зерттеу барысында қарастырылған жылдардағы Солтүстік Қазақстан облысындағы ауа-райын саралайық.

2018 жылдың мамыр айында 47,7 мм жауын-шашын түсті, бұл орташа көп жылдық норманың 170% құрайды. Маусым айында 52,6 мм, шілде айында 67,9 мм жауын-шашын түскен, бұл сәйкесінше,

көпжылдық орташа көрсеткіштің 120% және 96,0% құрайды. Тамыз айында 148,8 мм жауын-шашын түсті, бұл орташа жылдық норманың 314%-ын құрайды. Осы жылы ауылшаруашылық дақылдарының NDVI коэффиценттері 0,17-ден 0,50-ге дейін өсіп отырған.



4 сурет - 2018-2020 жылдар аралығындағы Солтүстік Қазақстан облысындағы ауыл шаруашылығы дақылдарының NDVI динамикасы

2019-2020 жж. қысында Қазақстан аумағының басым бөлігінде жауын – шашын нормадан көп түскен. Кейбір солтүстік – батыс аймақтарда жауын – шашын нормадан 70-80%-ға асқан, ал Қазақстанның солтүстік аймақтарында жауын – шашынның мөлшері нормадан 1,5-3,0 есе артық түсті.

2020 ж. Қазақстан территориясы бойынша орталағанда жауын – шашын мөлшері 270,7 мм (норманың 85%) болды. Солтүстіктің кейбір аймақтарында жылдық жауын – шашын мөлшері нормадан ең көбі максимум 40-45% жоғары болды. Көктемдегі жауын – шашын мөлшері Қазақстан бойынша орташалағанда 82% болды. Қостанай (норманың 142%-ы) және Солтүстік Қазақстан облыстарында (норманың 121%-ы) көктем айтарлықтай «ылғалды» болды. Сәйкесінше, алдыңғы екі жылға қарағанда, 2020 жылғы NDVI көрсеткіштері жоғары, яғни мамыр айының өзінде 0,29-дан 0,48-ге дейін көтерілген.

Қорытынды. Қашықтықтан зондтау деректерінің қолжетімділігі артқан сайын, оларды өңдеу әдістері де көбейе бастады. Қарастырылған мақалада NDVI вегетациялық индекстерінің сызықтық трендтерінің коэффициенттерінің өзгеруін талдау нәтижесі көрсетілген. Сызықтық тренд коэффициенттерінің графигінен вегетациялық коэффиценттердің өзгеру уақыты анық көрінеді, одан кейін вегетациялық индекстердің

төмендеу тенденциясы байқалады, бұл өсімдік биомассасының әлдеқайда азаюын білдіреді (Tokareva және т.б., 2020) .

Визуалды талдау мен сызықтық тренд коэффициенттерінің мәндерінен вегетациялық индекстердің қарастырылған жылдарда шілде айының ортасы, тамыз айының басына дейін өсіп, одан кейін төмендегенін байқаймыз. NDVI абсолютті максимумға 2018 жылы тамыздың басында (0,50), 2019 жылы шілденің басында (0,51) және 2020 жылы мамырдың басында (0,52) жеткен, содан кейін индекс мәндері төмендей бастайды. NDVI орташа мәні 0,41 (мамырдың ортасы) мен 0,45 (шілденің ортасы) аралығында болған. Бұл Солтүстік Қазақстан облысында ауылшаруашылығы дақылдарының биомассасының мамыр айынан бастап вегетациялық кезең барысында ауа-райының өзгерісіне тәуелді екені білдіреді. Сонымен қатар, Қазақстанның солтүстік аймағының климаты бидай, арпа, сұлы, т.б. тәрізді дақылдарды өсіруге айтарлықтай қолайлы екені, алайда (-1 -2°C) –дан бастап төмен температураға және құрғақ желмен + 35-40°C-дан жоғары температураға шыдамайтыны айқындалды.

Осылайша, NDVI коэффициенттерінің сызықтық тенденцияларын саралау Солтүстік Қазақстан облысының ауылшаруашылық дақылдарының 2018-2020 жылдардағы зерттелетін пиксельдегі жай-күйін толығырақ қарастыруға мүмкіндік береді.

Information about authors:

Mimenbayeva Aigul Bilialovna – PhD student of specialty 8D094 – «Information technology», Astana International University, Nur-Sultan, Kazakstan, e-mail: *aigulka79_79@mail.ru*, *https://orcid.org/0000-0003-4652-470X*;

Akanova Akerke Saparovna – PhD, Senior Lecturer Department of "Computers and Software", S.Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakstan, e-mail: *a.akanova@kazatu.kz*, *https://orcid.org/0000-0002-7178-2121*.

REFERENCES:

According To 2009, The Population Was 114 People. Classification of thematic tasks of agricultural monitoring using Modis remote sensing data. Computer technologies. 3: 76–102.

Bashirova ch. F. (2019) NDVI indicators for remote plant monitoring. Young scientist, 31: 30-31. https://moluch.ru/archive/269/61895/

Enebish B., Dashhuu D., Renchin M. et al. (2020).Climate impact on the NDVI of

northern Mongolia. Indian remote sensing society, 333-340. https://doi.org/10.1007/s12524-019-01080-9.

Gao B., Gun H., Zhou J., Liu Yu, Tsui Yu (2020) reconstruction of the spatial-time continuous modulated reflection band in East and South Asia from 2012 to 2015. Remote sensing of the environment, 12 (21): 1-19.

Igor E., Victoria V., Richard D., Martin V., Kurchatov A. (2016) trends in the normal plant difference index (NDVI) associated with Urban Development in northern western Siberia. Chemistry and physics of the atmosphere, 16 (15): 9563-9577.

Jinru H., Baofen S. (2017) important plant indexes for remote sensing: an overview of developments and applications. Journal of sensors, 1-17. https://doi.org/10.1155/2017/1353691.

Kataev M. (2016 a) method for equalizing the time series of the NDVI vegetative index obtained from the MODIS spectrum radiometer. Tusur Reports. 1: 35-39.

Kataev M., Bekerov A., Shal P. (2018 B) analysis of trends in the time series of the NDVI vegetative index. TUSURA reports, 1: 81-84.

Lo H., Dai S., Li M. et al. analysis of the impact of climate change and human activity on plant change on Hainan Island based on NDVI. Remote sensing of Indian society, 49, 1755-1767 (2021). https://doi.org/10.1007/s12524-021-01357-2.

Myrzatai A. (2022). Implementation and use of a local area network monitoring system to systematize the input data of incident forecasting systems. News of NAS RK. Computer Science Series, (2), 54-63. https://doi.org/10.32014/2022.2518-1726.129.

Nigam R., Bhattacharya B.K., Ganjal K.R., etc. creating a time series plant index from an Indian geostationary satellite and comparing it with a global product. Remote sensing of the Indian society, 40, 1-9 (2012) https://doi.org/10.1007/s12524-011-0122-2

Plotnikov D.E., Ponyatikov S.A., Bartalev S.A. (2018) a method of automated mapping of types of agricultural crops using remote sensing data and a simulation model of plant growth. Modern problems of remote sensing of the Earth from space, 15: 131-141.

Pradhan S., Sehgal V.K., Bandiopadhyay K.K. et al. (2018) comparison of plant indicators with two ground sensors. Indian remote sensing society, 46, 321-326 .https://doi.org/10.1007/s12524-017-0671-0.

Sandra E., Fabia H., Hanspeter L., Elias H. (2015) analysis of MODIS NDVI time series trends to determine land degradation and recovery in Mongolia. 113: 16-28. https://doi.org/10.1016/j.jaridenv.2014.09.001.

Shurr A. (2013) agro-industrial complex of the North Kazakhstan region of the Republic of Kazakhstan. Economics, 4: 135-139.

Stefan E., Tobias K., Christian L., Matthias B., Patrick H. (2016). Demonstrate the intensity of arable land use throughout Europe using the MODIS NDVI time series. Letters on Environmental Research. 11(2): 2-10.

Tokareva O., Pasko O., Majid S., Kabral P. (2020) monitoring the state of vegetation cover of the territory of central Iraq using Landsat-8 satellite data. News of Tomsk Polytechnic University. Georesources engineer, 6: 19-31.

Yang W., Zou H., Wang H. (2014) analysis of changes in land use in Boao City based on remote sensing images. Jiangxi Agriculture Act, 26 (6): 87-91.

Zhalankuzov T., Muller L., Saparov A. (2014) concept and results of soil monitoring in northern Kazakhstan. New measurement and evaluation tools for monitoring and managing land and water resources in agricultural landscapes of Central Asia. Environmental Science and engineering, 653-666. Springer, my friend. https://doi.org/10.1007/978-3-319-01017-5_42.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 198-213
https://doi.org/10.32014/2022.2518-1726.147

М.О. Ногайбаева[1*], Б. Ахметов[1], Дж.Дж. Расулзаде[2], Е.А. Максум[1], С. Рустамов[2]

[1]Сатпаев Университет, Казахстан, Алматы;
[2]Университет АДА, Азербайджан, Баку.
E-mail: mnogaibayeva@gmail.com

УСКОРЕНИЕ ВЫЧИСЛИТЕЛЬНОГО ПРОЦЕССА ТОПОЛОГИЧЕСКОЙ ОПТИМИЗАЦИИ НА ОСНОВЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ U-NET

Аннотация. На каждом этапе истории одной из актуальных проблем в инженерии и строительной отрасли было снижение стоимости материалов или веса для любого решения. Более легкие изделия (мечи, доспехи и т.д.) означали эффективность для пользователя, а также снижение стоимости означало удешевление производства. На протяжении всей инженерной истории топологическая оптимизация разрабатывалась для достижения желаемых результатов и была научно внедрена примерно в 18 веке. В настоящее время используются два наиболее известных алгоритма топологической оптимизации: Упрощенный изотропный материал с пенализацией (SIMP) или двунаправленная эволюционная структурная оптимизация (BESO). Поскольку SIMP обычно используется исследователями и имеет реализации в готовых инструментах, таких как SolidWorks, в этой статье основное внимание будет уделено этой методологии. Цель этой статьи - попытаться повысить общую эффективность, поскольку даже для простой 2D-модели обработка алгоритма оптимизации SIMP может занять более 10 минут. Было предложено два способа ускорить алгоритм при сохранении его точности: частично и полностью использовать модели глубокого обучения. Для первой модели алгоритм SIMP использовался для более ранних итераций, а затем задание было передано моделям глубокого обучения. В качестве второго предложения была предпринята попытка использовать тот же алгоритм, только с

использованием моделей глубокого обучения. В обоих случаях общее время выполнения было успешно сокращено при сохранении точности алгоритма.

Ключевые слова: топологическая оптимизация, метод конечных элементов, ускорение вычислений, SIMP, сверточная нейронная сеть.

М.О. Ногайбаева[1*], Б. Ахметов[1], Дж.Дж. Расулзаде[2], Е.А. Максум[1], С. Рустамов[2]

[1]Сәтбаев университеті, Қазақстан, Алматы;
[2]АДА университеті, Әзірбайжан, Баку.
E-mail: mnogaibayeva@gmail.com

U-NET КОНВОЛЮЦИЯЛЫҚ НЕЙРОНДЫҚ ЖЕЛІ НЕГІЗІНДЕ ТОПОЛОГИЯЛЫҚ ОҢТАЙЛАНДЫРУДЫҢ ЕСЕПТЕУ ПРОЦЕСІН ЖЕДЕЛДЕТУ

Аннотация. Тарихтың әр кезеңінде кез-келген шешім үшін материалдардың құнын немесе салмағын төмендету инженерия және құрылыс саласындағы өзекті мәселелердің бірі болды. Әлдеқайда жеңіл бұйымдар (қылыштар, қару-жарақтар және т.б.) пайдаланушы үшін тиімділігін, сондай-ақ құнын төмендету өндірісті арзандатуды білдіреді. Инженерлік тарих бойында топологиялық оңтайландыру қажетті нәтижелерге қол жеткізу үшін жасалды және шамамен 18 ғасырда ғылыми түрде енгізілді. Қазіргі уақытта топологиялық оңтайландырудың әлдеқайда танымал екі алгоритмі қолданылады: жеңілдетілген изотропты материал (SIMP) немесе екі бағытты эволюциялық құрылымдық оңтайландыру (BESO). SIMP әдетте зерттеушілермен қолданылып және SolidWorks сияқты дайын құралдарда жүзеге асырылатын болғандықтан, осы мақалада осы әдіснамаға басты назар аударылады. Бұл мақаланың мақсаты - қарапайым 2D моделі үшін SIMP оңтайландыру алгоритмін өңдеу 10 минуттан астам уақытты алатындықтан, жалпы тиімділікті арттыруға тырысу болып табылады. Алгоритмнің дәлдігін сақтай отырып, оны жеделдетудің екі әдісі ұсынылды: терең оқыту модельдерін ішінара және толық пайдалану. Бірінші модель үшін SIMP алгоритмі бұрынғы итерациялар үшін қолданылды, содан кейін тапсырма терең оқыту модельдеріне берілді. Екінші ұсыныс ретінде тек терең оқыту модельдерін қолдана отырып

сол алгоритмді қолдануға әрекет жасалды. Екі жағдайда да алгоритмнің дәлдігін сақтай отырып, жалпы жұмыс уақыты сәтті қысқартылды.

Түйін сөздер: топологиялық оңтайландыру, соңғы элементтер әдісі, есептеулерді жеделдету, SIMP, конвульсиялық нейрондық желі.

M. Nogaibayeva[1]*, B. Akhmetov[1], J. Rasulzade[2], Y. Maksum[1], S. Rustamov[2]

[1]Satbayev University, Kazakhstan, Almaty;
[2]ADA University, Azerbaijan, Baku.
E-mail: *mnogaibayeva@gmail.com*

ACCELERATION OF THE COMPUTATIONAL PROCESS OF TOPOLOGICAL OPTIMIZATION BASED ON THE CONVOLUTIONAL NEURAL NETWORK U-NET

Abstract. At every point in history, one of the actual problems in the engineering and construction industry was reducing the cost of materials or weight for any solution. Lighter products (swords, armor, etc.) meant efficiency for the user, and also reduction in cost meant cheaper production. Throughout engineering history, topology optimization was developed to achieve desired results and was scientifically introduced around the 18th century. Right now, there are two of the most famous topology optimization algorithms that are used: Simplified Isotropic Material with Penalization (SIMP) or Bi-directional evolutionary structural optimization (BESO). As SIMP is commonly used by researchers and has implementations in ready tools such as SolidWorks, this paper will mainly focus on this methodology. The goal of this paper is to try to improve overall efficiency, as, for even a simple 2D model, the SIMP optimization algorithm can take over 10 minutes to process. Two ways were proposed to make the algorithm faster while keeping its accuracy: partly and fully using deep learning models. For the first model, the SIMP algorithm was used for earlier iterations, and then the job was passed to deep learning models. As a second proposition, it was tried to use the same algorithm only using deep learning models. In both cases, overall time execution was successfully reduced, while preserving the accuracy of the algorithm.

Key words: topological optimization, finite element method, acceleration of calculations, SIMP, convolutional neural network**.**

Введение. Топологическая оптимизация (ТО) – это метод оптимизации распределения материала в пределах рассматриваемого пространства для заданного набора нагрузок и граничных условий. Целью ТО является определение оптимального распределения материала в проектной области. Данный метод основан на повторяющихся шагах анализа и обновления дизайна (Sigmund et al., 2013).

Топологическая оптимизация обычно достигается с помощью численных расчетов, где область проектирования дискретизируется конечными элементами. Другими словами, метод конечных элементов (МКЭ) является основным числовым инструментом. Топологическая оптимизация на основе МКЭ классифицируется как топологии изотропно-твердый / пустой (ИТП), анизотропно-твердый / пустой (АТП) и изотропно-твердый / пустой / пористый (ИТПП). Среди них наиболее важным классом является ИТП, где элементы МКЭ считаются либо заполненными выбранным изотропным материалом, либо не содержащими какого-либо материала. На самом деле применение твердых конструкций на основе изотропных материалов широко распространено во всех отраслях обрабатывающей промышленности и строительства. Точно так же 3D-печать структур с использованием одного материала является наиболее надежной, поскольку более прочная связь между печатными слоями легче достичь по сравнению с подходом с несколькими материалами.

Задачу ТО можно описать как поиск распределения материала, которое минимизирует целевую функцию F при ограничении объема. Распределение материала описывается переменной плотности $\rho(x)$, которая может принимать значение 0 (пустота) или 1 (твердый материал) в любой точке проектной плоскости Ω. Задача оптимизации может быть записана в математической форме как (Kozhnik et al., 2017):

$$\begin{aligned} &\text{мин: } F = F(u(\rho), \rho) = \int_{\Omega} f(u(\rho), \rho) dV \\ &\text{x} \\ &\text{подвергается: } G_0(\rho) = \int_{\Omega} \rho(x) dV - V_0 \leq 0 \\ &: G_i(u(\rho), \rho) \leq 0, \ i = 1, \dots, M \\ &: \rho(x) = 0 \text{ or } 1, \forall x \in \Omega \end{aligned} \quad (1)$$

где u соответствует уравнению состояния. Для простоты дальнейших обозначений мы здесь предполагаем, что целевая функция может быть вычислена как интеграл по локальной функции $f(u(\rho), \rho)$. Также, в связи с тем, что в реальных условиях имеются несколько ограничений, в общую формулировку включены M дополнительных ограничений.

Основная задача оптимизации топологии (1) может быть решена двумя способами: либо как задача оптимизации формы, либо как подход плотности (узловые или поэлементные плотности). Эти два подхода можно также назвать лагранжевым (граница следует за сеткой) и эйлеровым (фиксированная сетка) соответственно.

До настоящего времени было разработано несколько методов для решения задач ТО: плотностный подход (density approach), подход с набором уровней (level-set approach), подход фазового поля (phase field approach) и дискретные подходы. Имеется несколько работ, в которых сравнивают преимущества и недостатки каждого из этих методов с точки зрения их вычислительной эффективности (Munk et al., 2015; Rozvany, 2009; Deaton et al., 2014; Sigmund et al., 2013). Среди них метод плотностный подход, и, в частности, метод SIMP (твердый изотропный материал с пениализацией) обычно используется исследователями / инженерами и включен в коммерческие программные обеспечения, такие как SolidWorks / COMSOL / ANSYS и так далее. Также метод SIMP получил широкое применение в аддитивных технологиях (технологии 3D-печати) **(**Kozhnik et al., 2017**)**. Таким образом, обычная задача ТО на основе подхода SIMP, где цель состоит в том, чтобы минимизировать степень соответствия (то есть максимизировать жесткость конструкции), может быть записана как (Sigmund et al., 2001):

$$\begin{aligned} &\text{мин: } C(x) = U^T K U = \sum_{e=1}^{N} (x_e)^p u_e^T k_0 u_e \\ &\quad x \\ &\text{подвергается: } \frac{V(x)}{V_0} = f \\ &: KU = F \\ &: 0 < x_{min} \le x \le 1 \end{aligned} \qquad (2)$$

где C - целевая функция; U и F - векторы глобального смещения и силы соответственно; K - матрица глобальной жесткости; u_e и k_e - вектор смещения элемента и матрица жесткости соответственно; x - вектор проектных переменных; x_{min} - вектор минимальных относительных плотностей (ненулевой, чтобы избежать сингулярности); N - количество элементов МКЭ, p - мощность штрафа,V(x) и - объем материала и объем расчетной области соответственно, а f - предписанная доля объема.

Задача оптимизации может быть решена с использованием различных итерационных методов, таких как метод критериев оптимальности

(OC), метод последовательного линейного программирования (SLP) или метод перемещения асимптот (MMA), (Svanberg, 1987). Любой из этих методов требует значительного вычислительного времени. В качестве примера можно рассмотреть ТО симметричной половины Messerschmitt-Bölkow-Blohm (MBB) как показано на рисунке 1. Результаты, полученные на основе метода MMA, показывают, что общий контур/форма возможной структуры достигается после 13 итераций, в то время как для получения окончательного точно настроенного двоичного результата требуются дополнительные 67 итераций.

Рисунок 1 – Пример луча MBB: a) область проектирования, граничные условия; и b) промежуточные и окончательные результаты ТО на основе метода SIMP для элементов МКЭ размером 120 × 40

Более того, в последнее десятилетие техника ТО была значительно усовершенствована (Sigmund et al., 2013), и ее применение для решения сложных задач заметно расширилось (Lundgaard et al., 2018). Такие задачи, как улучшение теплопередачи (Wadbro et al., 2009; Dbouk, 2017), взаимодействие жидкости и конструкции (Neofytou et al., 2021), термоупругое поведение конструкции (Gao et al, 2016) и улучшение геометрии электротермомеханических приводов (Ramírez-Gil et al., 2016; Ramírez-Gil et al., 2021) были численно дискретизированы в трехмерных областях и оптимизированы с использованием ТО для повышения производительности. Изучение задач с несколькими взаимодействующими физическими свойствами в трехмерных измерениях приводит к заметному повышению вычислительной сложности, и, следовательно, увеличению времени вычислений задач ТО.

Существует несколько методов для решения данной проблемы. Одними из них являются параллельные вычисления разработанные, как и для центральных процессоров (ЦПУ) (Mahdavi et al., 2006; Vemaganti et al., 2005; Borrvall et al., 2001), так и для графических процессоров (ГПУ) (Suresh et al., 2010; Schmidt et al., 2011; Zegard et al., 2013). Так же, с появлением технологий искусственного интеллекта

(ИИ) и платформ ИИ с открытым исходным кодом, применение методов машинного обучения (МО) в передовых производственных процессах значительно расширилось (Paraskevoudis et al., 2020; Johnson et al., 2020). Такая тенденция заметна и в области ТО, где с помощью обученных нейронных сетей прогнозируются структурная топология и такие свойства, как прочность, модуль упругости, поля деформаций и напряжений (Wang et al., 2021; Sosnovik et al., 2019; Guo et al., 2021). Данная работа посвящена ускорению ТО методом МО с использованием архитектуры U-net.

Архитектура U-net. Основная идея. U-net — это особый тип сверточной нейронной сети, который был представлен факультетом компьютерных наук Фрайбургского университета, и предназначен для решения задач связанных с сегментацией изображений. Он состоит из двух основных стадий: свертка и развертка. На каждом шаге первой стадии (свертки) модель генерирует N уменьшенных (сжатых или свертанных) отфильтрованных версий изображения, причем при каждом шаге количество фильтров будет увеличиваться вдвое. Последующая стадия (развертка) представляет собой обратную версию свертки. На каждом шаге данной стадии изображения будут увеличиваться в размере и фильтроваться, причем начальное количество фильтров соответствует конечному количеству фильтров 1-й стадии, также количество фильтров уменьшается вдвое при каждом шаге. Ключевым действием в этой модели является сохранение и передача промежуточных изображений со стадии свертки на развертку, причем результат первого шага свертки будет использоваться в последнем шаге развертки, второй – в предпоследнем и т.д. Конечным решением данной задачи является изображение размером, соответствующим с первоначальным.

Промежуточные этапы, слои: Как было указано выше, U-net является особым типом сверточной нейронной сети, которая состоит из следующих слоев: свертки (convolution), уменьшения (pooling), отсева (dropout) и увеличения (up-sampling).

Сверточный слой представляет собой набор карт (набор матриц) с обучаемыми фильтрами (в разных источниках его называют по-разному: сканирующее ядро или синаптическое ядро), которые имеют небольшое рецептивное поле, но проходят на всю глубину входного объема. В рамках данного эксперимента на каждом шаге свертки количество фильтров с активационной функцией ReLU (Rectified Linear Unit) будет увеличиваться вдвое с начальным количеством

равным шестнадцати. Далее, на каждом шаге развертки количество фильтров будет уменьшаться вдвое. Последний набор из шестнадцати отфильтрованных изображений будет отфильтрован в последний раз с функцией активации сигмоид.

Слой уменьшения — это выборочный процесс дискретизации, целью которого является уменьшение разрешения входной матрицы. В рамках этого эксперимента для выделения области были использованы горизонтальные и вертикальные шаги равные двум (N = 2, M = 2), также максимальное значение каждой области было присвоено полученной матрице.

$$f'(x,y) = [\, f(N * x + i, M * y + j) \,]$$

Слой отсева — метод регуляризации, используемый в искусственных нейронных сетях для предотвращения переобучения. По сути, он обнуляет случайные значения во входном наборе данных и увеличивает эффективность алгоритмов машинного обучения.

$$w'_j = \{w_j \ , c\ P(C) \quad , \quad \text{иначе } 0$$

$P(c) -$ вероятность сохранения матрицы

$w_j -$ первоначальная матрица до отсева

$w'_j -$ матрица после отсева

Слой увеличения – это процесс увеличения ширины и длины входной матрицы в N и M раз соответственно. В экспериментах данная матрица заполнялась по следующей формуле, где N=2, M=2:

$$f'(x,y) = f\left(\text{int}\left(\frac{x}{N}\right), \text{int}\left(\frac{y}{M}\right)\right),$$

где $f'(x,y)$ - полученная матрица, $f(x,y)$ - первоначальная матрица

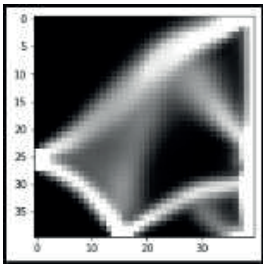
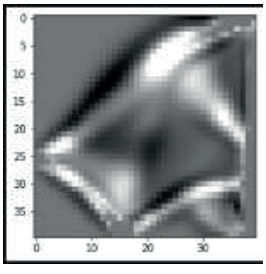
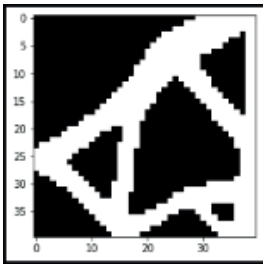
Материалы и методы исследования. Первая стадия: Свертка. Архитектура первой стадии соответствует обычной архитектуре сверточной нейронной сети. Она состоит из нескольких чередующихся шагов, которые условно обозначают глубину модели. Входными данными первого шага являются два изображения: N-ная итерация SIMP-а и разница между N-ным и предыдущим изображением. Каждый шаг начинается с применения двух повторных сверток с ядрами 3x3 с увеличивающимся количеством фильтров и активационной функцией ReLU, с промежуточным слоем отсева в 10%. Далее следует слой уменьшения с шагом 2x2. В данном эксперименте начальное количество

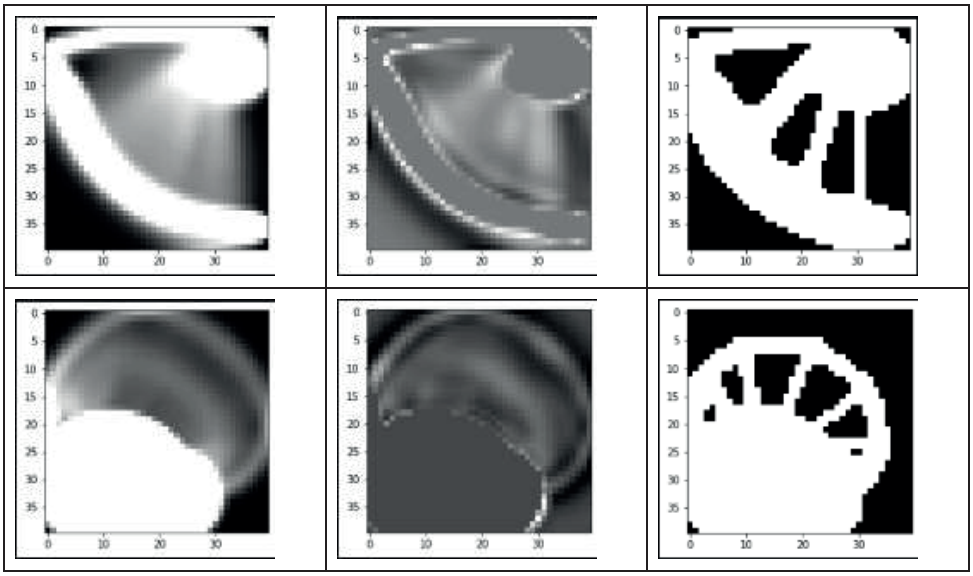
фильтров было взято за 16 и для каждого шага оно увеличивалось вдвое. Также конечный результат каждого шага сохраняется для передачи в следующую стадию.

Вторая стадия: Развертка. Стадия развертки (вторая стадия) является обратной к первой стадии и отличается тем, что в качестве входных данных для каждого шага, помимо полученных данных используются данные, ранее сохранённые в стадии свёртки. При каждом шаге происходят две повторные свертки 3x3 с умещающимся количеством фильтров и активационной функцией ReLU, с промежуточным слоем отсева в 10%. Далее следует слой увеличения с коэффициентом равным двум по вертикали и горизонтали. После 4-й, финальной итераций, полученный результат проходит через последний слой свертки с единственным фильтром и активационной функцией сигмоид. После округления полученного результата мы получаем бинарное изображение с первоначальным разрешением, где 0 соответствует пустоте, а 1 материалу.

Данные. Для реализации вышеописанной модели необходимы изображения итераций SIMP модели вместе с ожидаемым результатом. Были использованы синтетические данные, созданные И. Сосновиком и И. Оселедетом (Sosnovik et al., 2022) с использованием автоматического вычислителя 2D и 3D топологий SIMP Topy (Hunter et al., 2017), который находится в открытом доступе. В итоге было сгенерировано и использовано 10 000 мнимых задач, также для каждой задачи было сгенерировано 100 итераций SIMP наряду с ожидаемым результатом. В таблице 1 показаны 3 примера входных данных для N-ной итерации, градиент (разница N-ной итераций с предыдущей), и ожидаемый результат.

Таблица 1 – Входные данные.

N-ная итерация SIMP	Градиент	Ожидаемый результат
		



Метод оценки. Одним из распространенных применений машинного обучения является выполнение бинарной идентичности, которая просматривает входные данные и предсказывает, к какому из двух возможных классов они принадлежат. Практическое использование включает анализ настроений, обнаружение спама и обнаружение мошенничества с кредитными картами. Такие модели обучаются с помощью наборов данных, помеченных 1 и 0, представляющих два класса, используют популярные алгоритмы обучения, такие как логистическая регрессия и наивный байесовский алгоритм, и часто строятся с помощью библиотек глубокого обучения.

Бинарная идентичность. Данный способ оценки проверяет идентичность каждого пикселя изображения. Условно он будет обозначаться β и, учитывая, что каждый пиксель изображения имеет множество значений состоящих из двух цифр (ноль для обозначения пустоты и один для обозначения материала), будет вычисляться согласно формуле:

$$\beta = \frac{\omega_{00} + \omega_{11}}{n_0 + n_1}$$

Коэффициент Жаккара либо же отношение пересечения к объединению будет использоваться для вычисления уровня перекрытия предугаданного изображения с истинным.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Благодаря тому, что имеется всего два интересующих класса мы можем модифицировать оригинальную формулу и привести ее к следующему виду:

$$K = 0.5 * \left[\frac{\omega_{00}}{n_0 + \omega_{10}} + \frac{\omega_{11}}{n_1 + \omega_{01}}\right]$$

Во всех вышеперечисленных формулах, ω_{tp} это количество экземпляров класса t предугаданных как класс p, а n_x это суммарное количество экземпляров класса x в исходном изображении.

Результаты и обсуждения. В качестве примера рассмотрим ТО балки Мессершмитта-Белькова-Блома (МВВ). С помощью методов глубокого обучения уменьшается объемная доля балки МВВ.

На начальном этапе исследования были проведены два эксперимента с определения действия глубины на модели. Для более полной характеристики рассматриваемого вопроса были изучены две различные глубины разного размера: в первом эксперименте использовалась глубина – 3, а во втором максимальная – 4. Максимальная глубина ограничена разрешением изображения и на одну единицу меньше минимального логарифма разрешения по основанию два. Для данного случая (40x40) эта величина равна четырем.

Также каждый эксперимент имеет четыре раздела, где исследуется эффективность выбора начальных данных. Для этого наряду с равномерным распределением было использовано распределение Пуассона. Следует отметить, что распределение Пуассона было взято с тремя различными коэффициентами 5, 10, 30 и для каждого коэффициента были проведены эксперименты. В таблице 2 даны основные начальные параметры всех моделей. Как видно из таблицы, были проведены 8 эксперимента, 4 эксперимента с глубиной равной 3, и ещё 4 эксперимента с глубиной 4.

Таблица 2 – Начальные параметры модели.

№	Распределение	Глубина	Оптимизатор	Количество фильтров
1	Равномерное [1-100]	3	Адам	16, 32, 64, 64, 32, 16
2	Пуассон (5)	3	Адам	16, 32, 64, 64, 32, 16
3	Пуассон (10)	3	Адам	16, 32, 64, 64, 32, 16

№	Распределение	Глубина	Оптимизатор	Количество фильтров
4	Пуассон (30)	3	Адам	16, 32, 64, 64, 32, 16
5	Равномерное [1-100]	4	Адам	16, 32, 64, 128, 128, 64, 32, 16
6	Пуассон (5)	4	Адам	16, 32, 64, 128, 128, 64, 32, 16
7	Пуассон (10)	4	Адам	16, 32, 64, 128, 128, 64, 32, 16
8	Пуассон (30)	4	Адам	16, 32, 64, 128, 128, 64, 32, 16

Результаты начальных параметров каждого из экспериментов, а также, какие данные были использованы, и как модель была построена – представлены в таблице 3. Вследствие чего мы получили 8 различных моделей из 8 различных экспериментов. Каждая модель была оценена с помощью различных итераций SIMP. Таким образом, в данном эксперименте были выбраны итерации SIMP с 5-го по 80-го, с интервалом в 5 итераций. Для получения удовлетворительной средней точности в 94% было достаточно 5/10 итераций. Подводя промежуточные итоги, хотелось бы отметить, что каждая из этих итераций была проверена с помощью представленной модели.

Таблица 3 – Результаты эксперимента с глубиной в три слоя. Р- распределение, РР- равномерное распределение, П(х)- Распределение Пуассона с коэффициентом х.

№	Р	5	10	15	20	30	40	50	60	70	80
1	РР	93.59	95.44	96.23	96.68	97.20	97.59	97.77	97.97	98.08	98.22
2	П(5)	94.03	95.66	96.20	96.62	96.99	97.35	97.52	97.55	97.65	97.76
3	П(10)	94.06	95.48	96.09	96.47	96.82	97.23	97.42	97.48	97.53	97.64
4	П(30)	93.88	95.96	96.85	97.19	97.61	97.83	98.00	98.19	98.28	98.42

На основе вышесказанного результаты первых 4-х экспериментов указаны в таблице 3, а последующих четырёх в таблице 4 ниже. Исходя из данных таблиц, можно прийти к выводу, что точность представленной модели высокая, и что она может быть использована для ускорения SIMP. Таже можно отметить, что минимальная точность модели равная 94% наблюдалась при использовании пятой итераций. А максимальная точность в 99% была достигнута при использовании 80-й итераций. Вместе с тем следует подчеркнуть, что в данном случае изменение глубины модели минимально повлияло на результаты. Это можно объяснить тем, что изображения имеет малый размер, а увеличение глубины с 3-х до 4-х не делает большой разницы, однако увеличивает вычислительную сложность и время выполнения модели.

Таблица 4 – Результаты эксперимента с глубиной в четыре слоя. Р- распределение, РР- равномерное распределение, П(х)- Распределение Пуассона с коэффициентом х.

№	Р	5	10	15	20	30	40	50	60	70	80
5	РР	93.78	95.38	96.33	96.70	97.16	97.47	97.73	97.89	97.87	98.00
6	П(5)	94.08	95.59	95.94	96.38	96.92	97.10	97.34	97.46	97.59	97.55
7	П(10)	93.98	95.43	96.03	96.46	96.95	97.15	97.36	97.46	97.56	97.66
8	П(30)	93.53	95.74	96.51	96.74	97.42	97.76	97.97	98.14	98.19	98.23

На рисунке 2 показаны ожидаемые результаты (верхний ряд) и результаты вычислений (нижний ряд) представленной модели для трех различных случаев. Можно заметить, что точность во всех показанных случаев превышает 98%. Также можно отметить, что неточности в результатах в основном связаны с тонкими соединениями.

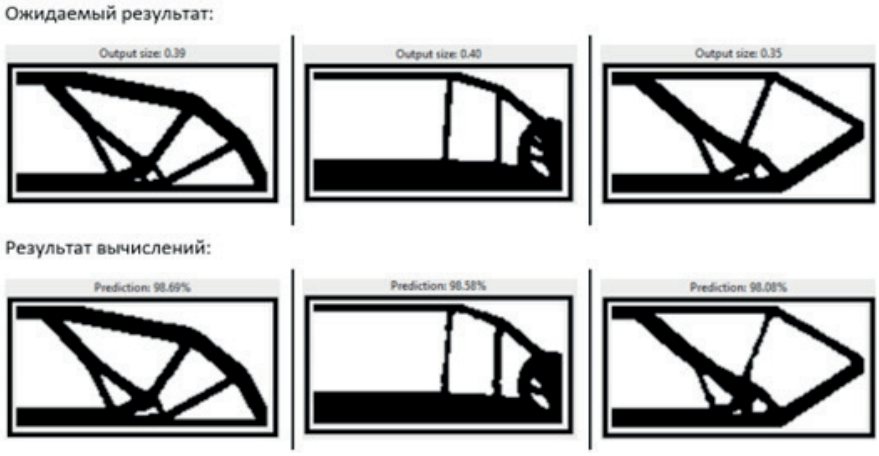


Рисунок 2 – Визуализация результатов.

Такие методы обладают высокой точностью и могут использоваться вместо традиционных методов топологической оптимизации, которые требуют много времени для получения окончательного дизайна структуры. Подход, основанный на данных, с передовыми алгоритмами, является многообещающим методом для будущих сценариев «Промышленности 4.0», таких как цифровые двойники или интеллектуальное производство, где необходимы крупномасштабные и долгосрочные симуляции.

Заключение. Данная работа посвящена разработке метода ускорения задачи ТО для балки МББ. Была представлена модель на основе архитектуры U-Net для ускорения вычисления рассматриваемой задачи. Результаты с использованием бинарной идентичности, где проверяется

идентичность каждого пикселя, показали, что для получения точности выше 95% требуется данные с 10-ой итерации SIMP. А при использовании 80 итераций, точность возрастает до 98%. Кроме того, было замечено, что изменение глубины с 3-го до 4-го уровня не улучшает точность эксперимента, хотя увеличивает вычислительную сложность и время выполнения задачи. Это может быть связано с маленьким размером рассматриваемых изображений.

***Подтверждение**. Это исследование было профинансировано Комитетом науки Министерства образования и науки Республики Казахстан по теме «Разработка метода топологической оптимизации на основе Глубокого Обучения и GPU-ускоренных вычислении для создания аэродинамических структур» (2020-2022) (Номер гранта AP08856141).*

Information about authors:

Nogaibayeva Makpal – senior lecturer, al-Farabi Kazakh National University, Almaty, Kazakhstan, *mnogaibayeva@gmail.com*, ORCID ID: http://orcid.org/0000-0003-1205-2564;

Akhmetov Bakytzhan– PhD, Satbayev University, Almaty, Kazakhstan, *eng.akhmetov@gmail.com*, ORCID ID: http://orcid.org/0000-0003-3323-0059;

Rasulzade Jalal – master's student, ADA University, Baku, Azerbaijan, *jalal.rasulov@gmail.com*;

Maksum Yelaman – PhD student, Satbayev University, Almaty, Kazakhstan, *maksum.yelaman@gmail.com*, ORCID ID: http://orcid.org/0000-0001-6573-1689;

Rustamov Samir – PhD, professor, ADA University, Baku, Azerbaijan, *samir.rustamov@gmail.com*, ORCID ID: http://orcid.org/0000-0002-3247-5882.

REFERENCES:

Kozhnik A.M., Guzh T.S., Ilyichev V.A. Modern trends in the optimization of metal structures // Youth. sci. forum: tech. and matem. science. – 2017. – February – number 2(42) – pp. 51-57 URL: *http://www.nauchforum.ru/archive/MNF_tech/2(42).pdf* (in Russ.).

A. Mahdavi, R. Balaji, M. Frecker, and E. M. Mockensturm, Topology optimization of 2D continua for minimum compliance using parallel computing, Structural and Multidisciplinary Optimization, 32 (2), 2006, pp. 121–132. *https://doi.org/10.1007/s00158-006-0006-1* (in Eng.).

A. Neofytou, F. Yu, L. Zhang, and H.A. Kim, Level Set Topology Optimization for Fluid-Structure Interactions, (January), 2021, pp. 1–17 *https://doi.org/10.2514/6.2021-1686* (in Eng.).

C. Lundgaard, J. Alexandersen, M. Zhou, C. Schousboe and A. Ole, Revisiting density-based topology optimization for fluid-structure-interaction problems, 2018, pp. 969–995. *https://doi.org/10.1007/s00158-018-1940-4* (in Eng.).

D.J. Munk, G.A. Vio and G.P. Steven, Topology and shape optimization methods using evolutionary algorithms: a review, Structural and Multidisciplinary Optimization, 52 (3), 2015. *https://doi.org/10.1007/s00158-015-1261-9* (in Eng.).

D. Wang, C. Xiang, Y. Pan, A. Chen, X. Zhou, and Y. Zhang, A deep convolutional neural network for topology optimization with perceptible generalization ability, Engineering Optimization, 2021. *https://doi.org/10.1080/0305215X.2021.1902998* (in Eng.).

E.P. Paraskevoudis, K. Karayannis P. and Koumoulos, Real-Time 3D Printing Remote Defect Detection (Stringing) with Computer Vision and Artificial Intelligence, Processes, 8(11), 2020. *https://doi.org/10.3390/pr8111464* (in Eng.).

E. Wadbro and M. Berggren, Megapixel topology optimization on a graphics processing unit, SIAM Review, 51 (4), 2009. *https://doi.org/10.1137/070699822* (in Eng.).

F.J. Ramírez-Gil, E.C.N. Silva and W. Montealegre-Rubio, Topology optimization design of 3D electrothermomechanical actuators by using GPU as a co-processor, Computer Methods in Applied Mechanics and Engineering, 302, 2016. *https://doi.org/10.1016/j.cma.2015.12.021* (in Eng.).

F.J. Ramírez-Gil, C.M. Pérez-Madrid, E.C.N. Silva, and W. Montealegre-Rubio, Parallel computing for the topology optimization method: Performance metrics and energy consumption analysis in multiphysics problems, Sustainable Computing: Informatics and Systems, 30, 2021. *https://doi.org/10.1016/j.suscom.2020.100481* (in Eng.).

G.I.N. Rozvany, A critical review of established methods of structural topology optimization, Structural and Multidisciplinary Optimization, 37 (3), 2009. *https://doi.org/10.1007/s00158-007-0217-0* (in Eng.).

Huang X, Xie Y.M. (2007) Convergent and mesh-independent solutions for the bidirectional evolutionary structural optimization method. Finite Elements in Analysis and Design 43(14):1039–1049 *https://doi.org/10.1016/j.finel.2007.06.006* (in Eng.).

I. Sosnovik and I. Oseledets, Neural networks for topology optimization, Russian Journal of Numerical Analysis and Mathematical Modelling, 34 (4), 2019. *https://doi.org/10.1515/rnam-2019-0018* (in Eng.).

I. Sosnovik and I. Oseledets, Neural networks for topology optimization. Retrieved 17 April 2022, from https://arxiv.org/abs/1709.09578 (in Eng.).

J.D. Deaton and R.V. Grandhi, A survey of structural and multidisciplinary continuum topology optimization: Post 2000, Structural and Multidisciplinary Optimization, 49 (1). 2014. *https://doi.org/10.1007/s00158-013-0956-z* (in Eng.).

K. Guo, Z. Yang, C.-H. Yu, and M.J. Buehler, Artificial intelligence and machine learning in design of mechanical materials, Materials Horizons, 2021. *https://doi.org/10.1039/D0MH01451F* (in Eng.).

K. Suresh, A 199-line Matlab code for Pareto-optimal tracing in topology optimization, Structural and Multidisciplinary Optimization, 42 (5), 2010. *https://doi.org/10.1007/s00158-010-0534-6* (in Eng.).

K. Svanberg, The method of moving asymptotes—a new method for structural

optimization, International Journal for Numerical Methods in Engineering, 24 (2), 1987. *https://doi.org/10.1002/nme.1620240207* (in Eng.).

K. Vemaganti and W.E. Lawrence, Parallel methods for optimality criteria-based topology optimization, Computer Methods in Applied Mechanics and Engineering, 194 (34–35), 2005, pp. 3637–3667. *https://doi.org/10.1016/j.cma.2004.08.008* (in Eng.).

M.P. B Bendsøe, E. Lund, N. Olhoff, O. Sigmund, Topology optimization-broadening the areas of application, Control and Cybernetics 34 (2005) 7–35. journal ISSN: 0324-8569.

M.P. Bendsøe, Optimal shape design as a material distribution problem, Structural and multidisciplinary optimization 1 (4) (1989) 193-202 *https://doi.org/10.1007/BF01650949* (in Eng.).

N.P. Van Dijk, K. Maute, M. Langelaar, and F. Van Keulen, Level-set methods for structural topology optimization: A review, Structural and Multidisciplinary Optimization, 48 (3). 2013. *https://doi.org/10.1007/s00158-013-0912-y* (in Eng.).

N.S. Johnson et al., Invited review: Machine learning for materials developments in metals additive manufacturing, Additive Manufacturing, 36, 2020. *https://doi.org/10.1016/j.addma.2020.101641* (in Eng.).

O. Sigmund and K. Maute, Topology optimization approaches: A comparative review, Structural and Multidisciplinary Optimization, 48 (6), 2013, pp. 1031–1055. *https://doi.org/10.1007/s00158-013-0978-6* (in Eng.).

O. Sigmund, A 99 line topology optimization code written in matlab, Structural and Multidisciplinary Optimization, 21 (2), 2001. *https://doi.org/10.1007/s001580050176* (in Eng.).

S. Schmidt and V. Schulz, A 2589 line topology optimization code written for the graphics card, Computing and Visualization in Science, 14 (6), 2011. *https://doi.org/10.1007/s00791-012-0180-1* (in Eng.).

T. Borrvall and J. Petersson, Large-scale topology optimization in 3D using parallel computing, Computer Methods in Applied Mechanics and Engineering, 190 (46–47), 2001, pp. 6201–6229. *https://doi.org/10.1016/S0045-7825(01)00216-X* (in Eng.).

T. Dbouk, A review about the engineering design of optimal heat transfer systems using topology optimization, Applied Thermal Engineering, 112, 2017, pp. 841–854. *https://doi.org/10.1016/j.applthermaleng.2016.10.134* (in Eng.).

T. Gao, P. Xu, and W. Zhang, Topology optimization of thermo-elastic structures with multiple materials under mass constraint, Computers and Structures, 173, 2016, pp. 150–160. *https://doi.org/10.1016/j.compstruc.2016.06.002* (in Eng.).

T. Zegard and G.H. Paulino, Toward GPU accelerated topology optimization on unstructured meshes, Structural and Multidisciplinary Optimization, 48 (3), 2013. *https://doi.org/10.1007/s00158-013-0920-y* (in Eng.).

W. Hunter, et al., Topy – topology optimization with python, https://github.com/williamhunter/topy (2017).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 214-227
https://doi.org/10.32014/2022.2518-1726.148
UDC 551.3,528.8

G. Turebaeva*, A. Syzdykov, A. Tenchurina, J. Doshakov

Technical University named after Abylkas Saginov,
Kazakhstan, Karaganda.
E-mail: *gulnara_83.06.12@mail.ru*,

NUMERICAL METHODS FOR SOLVING DIFFERENTIAL EQUATIONS USING APPLICATION PROGRAMS

Abstract. The modern stage of society's development is characterized by the strong influence of computer technologies on it, which penetrate into all spheres of human activity. The use of computers activates the process of studying the discipline by students, facilitates and accelerates the assimilation of new material and control, which ultimately improves the quality of education and deepens students' knowledge. At the same time, both standard programs and those developed at departments are used when studying the most important topics of the theoretical course and the material of practical and laboratory classes. Also, in conditions of versatility, the availability of textbooks saturated with theory, and a shortage of classroom time, new approaches to conducting classes are needed to achieve high quality knowledge and skills.

The article discusses the possibilities of using modern computer technologies, in particular, the Mathcad application program for visual representation of physical processes. This article shows methods for solving ordinary differential equations in the Mathcad package based on numerical methods. As an example of a nonlinear process, the Cauchy problem for a second order ordinary differential equation is solved using the Mathcad odesolve function, which is a complication of the linear oscillator equation, and a graph is obtained. It also talks about the advantages of using the Mathcad application program for solving problems in physics, which allows you to not only make the necessary calculations, but also to arrange your

work using graphs, drawings, tables and mathematical formulas. Based on the results of these modeling works, the user gets the system model ready and can only set the initial conditions and control all the parameters of the model during the numerical experiment. In this regard, this program provides an opportunity to expand the teaching activities of the teacher and increase the independence and activity of students.

Key words: physical processes; MathCad; modeling; physical models; Runge-Kutta method; complex systems; solutions of ordinary differential equations; learning process; examples of problem solving.

Г.Б. Туребаева*, А.К. Сыздықов, А.Р. Тенчурина, Ж.Б. Дошакова

Абылқас Сағынов атындағы Қарағанды техникалық университеті, Қазақстан, Қарағанды.
E-mail: *gulnara_83.06.12@mail.ru*,

ҚОЛДАНБАЛЫ БАҒДАРЛАМАЛАРДЫ ҚОЛДАНА ОТЫРЫП ДИФФЕРЕНЦИАЛДЫҚ ТЕҢДЕУЛЕРДІ ШЕШУДІҢ САНДЫҚ ӘДІСТЕРІ

Аннотация. Қоғам дамуының қазіргі кезеңі оған адам қызметінің барлық салаларына енетін компьютерлік технологиялардың қатты әсерімен сипатталады. Компьютерлерді қолдану студенттердің пәнді оқу процесін белсендіреді, жаңа материалды игеруді және бақылауды жеңілдетеді және жылдамдатады, нәтижесінде оқу сапасын арттырады және студенттердің білімін тереңдетеді. Бұл жағдайда стандартты бағдарламалар да, теориялық курстың маңызды тақырыптарын, практикалық және зертханалық сабақтардың материалдарын оқу кезінде кафедраларда жасалған бағдарламалар да қолданылады. Сондай-ақ, көп салалы, теорияға толы оқулықтардың болуы, аудиториялық уақыттың жетіспеушілігі жағдайында білім мен дағдылардың жоғары сапасына қол жеткізуге мүмкіндік беретін сабақтарды өткізудің жаңа тәсілдері қажет

Мақалада заманауи компьютерлік технологияларды, атап айтқанда физикалық процестерді көрнекі түрде көрсету үшін Mathcad қолданбалы бағдарламасын қолдану мүмкіндіктері қарастырылады. Бұл жұмыста сандық әдістерге негізделген Mathcad пакетіндегі қарапайым дифференциалдық теңдеулерді шешу әдістері көрсетілген.

Сызықтық емес процестің мысалы ретінде екінші ретті қарапайым дифференциалдық теңдеу үшін Коши есебі сызықтық осциллятор теңдеуінің күрделенуі болып табылатын Mathcad пакетінің odesolve функциясын қолдана отырып шешілді және график алынды. Сондай-ақ, физика есептерін шешуде Mathcad қолданбалы бағдарламасын қолданудың артықшылықтары туралы айтылады. Бұл өз кезегінде қажетті есептеулерді жүргізуге ғана емес, сонымен қатар графиктер, суреттер, кестелер және математикалық формулалар көмегімен жұмысты ұйымдастыруға мүмкіндік береді. Осы модельдік жұмыстардың нәтижелері бойынша пайдаланушы дайын жүйенің моделін алады және сандық эксперимент кезінде бастапқы шарттарды еркін орнатуға және модельдің барлық параметрлерін басқаруға мүмкіндік алады. Осыған байланысты бұл бағдарлама оқытушының оқу қызметін кеңейтуге және студенттердің дербестігі мен белсенділігін арттыруға мүмкіндік береді.

Түйін сөздер: физикалық процестер, Mathcad, модельдеу, физикалық модельдер, Рунге-Кутта әдісі, күрделі жүйелер, қарапайым дифференциалдық теңдеулерді шешу, зерттеу процесі, проблемаларды шешу мысалдары.

Г.Б. Туребаева*, А.К. Сыздыков, А.Р. Тенчурина, Ж.Б. Дошаков

Карагандинский технический университет имени Абылкаса Сагинова,
Казахстан, Караганда.
E-mail: *gulnara_83.06.12@mail.ru*,

ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ИСПОЛЬЗОВАНИЕМ ПРИКЛАДНЫХ ПРОГРАММ

Аннотация. Современный этап становления общества характеризуется мощным влиянием на него компьютерных технологий, которые проникают во все сферы человеческой деятельности. Использование компьютеров активизирует процесс изучения дисциплины студентами, облегчает и ускоряет усвоение нового материала и контроль, что в результате повышает качество обучения и углубляет познания студентов. При этом применяются как стандартные программы, так и разрабатываемые на кафедрах при изучении особенно важных тем теоретического курса и материала практических и лабораторных занятий. Также в условиях многопрофильности, наличия учебников насыщенных теори-

ей, нехватки аудиторного времени нужны новые подходы к проведению занятий, позволяющие добиваться высокого качества знаний и умений.

В статье рассматриваются возможности применения современных компьютерных технологий, в частности, прикладной программы Mathcad для наглядного представления физических процессов. В данной работе показаны методы решения обыкновенных дифференциальных уравнений в пакете Mathcad на основе численных методов. Как пример нелинейного процесса была решена задача Коши для обыкновенного дифференциального уравнения второго порядка с помощью функции odesolve пакета Mathcad, являющегося усложнением уравнения линейного осциллятора и получен график. Также говорится о преимуществах применения прикладной программы Mathcad при решении задач по физике, что позволяет не только провести необходимые расчеты, но и оформить свою работу с помощью графиков, рисунков, таблиц и математических формул. По результатам данных модельных работ, пользователь получает модель системы готовой и имеет возможность лишь произвольно задавать начальные условия и управлять всеми параметрами модели в ходе численного эксперимента. В связи с этим данная программа дает возможность расширить обучающую деятельность преподавателя и повысить самостоятельность и активность студентов.

Ключевые слова: физические процессы, Mathcad, моделирование, физические модели, Метод Рунге-Кутта, сложные системы, решения обыкновенных дифференциальных уравнений, процесс изучения, примеры решения проблем.

Introduction. In modern conditions of intensive development of information technology, there is a need to create a new system (methods, forms, etc.) of the educational environment. Currently, the urgent issue is the use of program-pedagogical and telecommunication facilities in the educational process of a higher educational institution, and, in particular, in teaching physics

Since the introduction of information technology in the educational process, the role of their use has increased. Physics as an educational discipline lends itself perfectly to the process of computerization. Information technology in the process of teaching physics can be used to study theoretical material, training, as a means of modeling and visualization, as well as in solving physical problems.

One of the main disciplines in technical universities is “Physics”. Solving

problems in physics classes contributes to the formation and consolidation of acquired knowledge and skills in practice in order to use them in professional activities. It is applied physical problems that reflect the technical content and essence of the future professional activities of a university graduate. The solution to this type of problem allows students to get acquainted with the various principles of the operation of technical devices, physical research methods (Bursian et al, 2015).

However, in the educational process, students encounter many difficulties and the use of computer technology in solving problems greatly facilitates and solves these difficulties. And also computer technologies in the process of solving problems develop students\' interest in the subject of physics and computer technology.

Computer programs help students solve applied physical problems. Many tasks contain huge calculations, in the solution of which, by making a small mistake, you can get the wrong answer. Computer programs make it possible to prevent such errors and make it possible to come to the correct answer quickly enough. The use of such programs can reduce the time for calculations, and increase the time spent on analysis and conclusion.

To date, computer programs that help solve applied physical problems are many. Many of them allow the calculation of tasks; some of them make charts, graphs and diagrams for which students would spend a lot of time; others allow virtual experiments and experiments, that is, many physical phenomena are clearly demonstrated to students.

Modern computer programs and telecommunication technologies provide students with access to non-traditional sources of information - electronic hypertext textbooks, application software packages, distance learning systems, etc., this is designed to increase the efficiency of development of cognitive independence and provide new opportunities for creative personal growth

The development of information and telecommunication technologies is so fast that the existing pedagogical research does not have time to analyze new methods, means and forms of teaching physics [Dyakonov et al, 2019].

The use of computers in the study of physics provides great opportunities. For example, computers are increasingly being used in laboratory work. More and more, the so-called virtual laboratory work is being introduced into the educational process in universities. Many programs such as Matcad, Matlab and others are used to process the results, which leads to a reduction in the time for this calculation of results and more time is left for better laboratory work. The use of computers in teaching physics allows you to change the

teaching methodology, and this leads to the facilitation of the work of the teacher. A personal computer is not replaced by traditional teaching aids, but supplementing them and together with them form a system of teaching aids focused on the use of new information technologies, the use of which creates the conditions for teaching physics in the educational information environment.

Materials and basic methods. Computer labs satisfy almost all the requirements of a physical experiment, except that they are not familiar with specific devices. Further, they can be used in distance learning. Computer labs can be used when a real physical experiment is not possible at all. For example, in the implementation of thought experiments that play an important role in the development of physics. They can also be used when the material is very complex and to study it requires increased visibility. It is computer labs that have such increased visibility. During computer experiments, the studied physical processes are visualized, and due to the application of computer modeling methods, graphic symbols on the screen depicting physical objects move in accordance with the laws of physics. However, these computer experiments were created precisely as interactive laboratory work, i.e. students in the course of their implementation can independently change the values of physical parameters and take measurements of "virtual" physical quantities, and then by calculating or plotting to determine other physical quantities. These "virtual" laboratory work can be performed along with the usual ones, they are not alternative, but mutually complement each other. In most of these laboratory works, phenomena are studied, the study of which is difficult using ordinary field works (Guld, Tobochnik et al, 2017).

First of all, it is extremely convenient to use computer models in a demo version when explaining new material or in solving problems. Problem solving is a necessary element of teaching physics and the formation of a creative personality. Using a computer in the process of solving physical problems allows not only to better absorb physics, but also demonstrates the importance of a computer as an instrument of creativity and as an effective assistant in the study of the laws of nature.

Numerical methods for solving ordinary differential equations (odes). As you know, any physical phenomena are described by differential equations, so the solution of ordinary differential equations for modeling physical processes is of great practical importance. The solution of ordinary differential equations based on numerical methods is widely used in the practice of scientific and technical calculations, as well as in solving various problems of physics, mechanics and other natural Sciences.

The main form of setting the initial conditions, which is used for modeling real processes using ordinary differential equations, is associated with determining the values of all lower derivatives at the starting point of the variable change interval. So, in physics, ordinary differential equations usually describe the change in the studied characteristic over time, and the initial conditions are determined at the moment $t = 0$ (Gorbatenko, 2019).

Thus, given ordinary differential equations and systems of differential equations are called Cauchy tasks, that is, if additional conditions are given for a single value of an independent variable, then such a task is called a Cauchy task.

Methods for numerical solution of ordinary differential equations in the form of the Cauchy task are developed in great detail. The most popular of them are deservedly the Runge-Kutta algorithm, which is successfully used for solving the vast majority of differential equations.

To solve an ordinary differential equation, you need to know the values of the dependent variable and the derivatives for some values of the independent variable. If the conditions are set for two or more values of an independent variable, the problem is called a boundary value problem. Such tasks require finding a function (or several functions) of a single variable, if, first, a differential equation (or system of equations) containing the derivative of the function is defined, and, second, the necessary number of additional conditions specifying the value of the function at some starting point (Kondratiev et al, 2015).

Solving Cauchy tasks for ordinary differential equations is a technology that has been developed in detail for a long time. With "good" ordinary differential equations, no computational problems usually arise at all (most often they are solved using the Euler and Runge – Kutta algorithm), and for a special type of ordinary differential equations, called hard ones, special methods must be used. All these features are embedded in Mathcad, and the user is allowed to choose a specific solution algorithm.

The Cauchy task can be formulated as follows: let be an ordinary differential equation:

$$\frac{dy}{dx} = f(x, y) \qquad (1)$$

$y(x)$ initial condition $y(x_0) = y_0$. You need to find a function $y(x)$ that satisfies both the specified equation and the initial condition.

The numerical solution of the Cauchy task consists in constructing a table of approximate values of $y_1, y_2,...,y_n$ for solving the equation $y(x)$ at points $x_1, x_2,..., x_n$. Most often $x_i=x_0+ih, i=0,1,...n,$ where h is the increment step of

the variable x, n is the number of intervals of the solution with a step h.

Here we consider two groups of numerical methods for solving the Cauchy task: one-step and multi-step.

One-step methods are methods where finding the next point on the $y(x)=f(x)$ curve requires information about only one previous step. The simplest one step method is the Euler method:

$$y_{i+1}=y_i+f(x_i,y_i)h \quad (2)$$
$$i=0,1,...,n-1.$$

The Euler method has low accuracy (on the order of h).

To achieve higher accuracy (order h^4), the fourth-order Runge-Kutta method is used:

$$y_{i+1}=y_i+\frac{k_0+2k_1+2k_2+k_3}{6}, \text{ где, } \begin{array}{l} k_0=h\cdot f(x_i,y_i), \\ k_1=h\cdot f(x_i+\frac{h}{2},y_i+\frac{k_0}{2}) \end{array},$$
$$k_2=h\cdot f(x_i+\frac{h}{2},y_i+\frac{k_1}{2}),$$
$$k_3=h\cdot f(x_i+h,y_i+k_2)$$

Based on the above, now let's try to solve physical problems using Matcad based on numerical methods.

Results. The results of numerical modeling. The solution of linear differential equations and their systems in Mathcad is presented in two forms: as a computational block and as inlinefunctions. The first form is preferable from the point of view of visual representation of the solution and technical simplicity, while the second opens up much wider possibilities (Levitskij et al, 2016).

The differential equation of the first order can by definition contain, in addition to the function $y(t)$ itself, only its first derivative $y'(t)$. In the vast majority of cases, the differential equation can be written in standard form (Cauchy form):

$$y'(t)=f(y(t,)t) \quad (3)$$

and only with this form can the Mathcad computing processor work. Correct from a mathematical point of view, the formulation of the corresponding Cauchy task for first-order ordinary differential equations must contain one initial condition in addition to the equation itself -function value $y(t_0)$ the value of the function x at some point t_0 .You need to explicitly define the function $y(t)$ on the interval from t_0 to t_x. By the nature of the performance the Cauchy tasks are also called tasks with initial conditions (Matros, Polev, Melnikova et al, 2017).

To numerically integrate a single ordinary differential equation, the Mathcad user has the choice of either using the given/odesoive computing block, or using built-in functions, such as the rkfixed function, as in previous versions of Mathcad. The first way is preferable for reasons of visual representation of the problem and results, and the second gives the user more leverage over the parameters of the numerical method.

Computing block Given/Odesolve

A computational block for solving a single ordinary differential equation that implements the Runge-Kutta numerical method consists of three parts:

- Given-keyword;

- the ordinary differential equation and the initial condition written using logical operators must be typed on the Boolean toolbar (Boolean operators) and the initial condition must be in the form of $y(t0)=b$;

- odesolve x (t,t_1) - inline function for solving an ordinary differential equation with respect to a variable t on an interval (t_0, t_1), with $t_0<t_1$.

It is acceptable, and even often preferable, to set the Odesolve function $(t, t_1, step)$ with three parameters, where step is an internal parameter of the numerical method that determines the number of steps in which the Runge - Kutta method will calculate the solution of the differential equation. The larger the step, the more accurate the result will be, but the more time will be spent searching for it. Keep in mind that selecting this parameter can significantly (several times) speed up calculations without significantly impairing their accuracy.

An example of solving the Cauchy task for a first order ordinary differential equation $y'=y-y^2$ using a computational block is given in listing 1.

Listing 1. Solution of the Cauchy problem for an ordinary differential equation of the first order

Given

$y(0)=0.1$ $y'(t)=y(t)-y(t)^2$ $y:=Odesolve\ (t,10)$

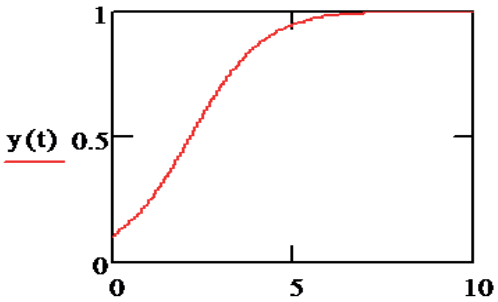


Fig. 1. Solution of the Cauchy task

Keep in mind that you should insert Boolean operators using the Boolean operators toolbar. When entering from the keyboard, remember that the logical equal sign corresponds to the keyboard shortcut <Ctrl>+<=>.The derivative symbol can be entered using the Calculus panel. Mathcad requires that the end point of integration of an ordinary differential equation lies to the right of the initial one: $(t_0<t_1)$ (in listing 1. $t_0=0$, $t_1=10$), otherwise an error message will be returned.

As you can see, the result of using the Given/odesoive block is a function *y(t)* defined on the interval (t_0,t_1). You should use the usual Mathcad tools to plot it or get the function value at some point in the specified interval, for example: *y(3)=0.691*. The user has the option to choose between two modifications and the Runge-Kutta numerical method. To change the method, right-click on the odesolve function area to open the context menu and select one of two options: Fixed (Fixed step) or Adaptive (Adaptive). By default, the first one is used, i.e. the Runge-Kutta method with a fixed step (Verzhbitsky at al, 2001).

Now let's try using the method of solving equations using the odesolve function to simulate various processes, in particular the model of a nonlinear harmonic oscillator. It is based on the solution of the Cauchy problem for an ordinary differential equation of the second order, which is a complication of the equation $(2\omega y'' + \beta y' + y + \gamma y^2 = 0)$ of a linear oscillator, where ω is the cyclic frequency of oscillations, γ is the attenuation coefficient. The harmonic oscillator model describes, in particular, the pendulum oscillations: *y(t)* describes changes in the angle of its deviation from the vertical, *y'(t)*-the angular speed of the pendulum, *y''(t)* -acceleration, and the initial conditions, respectively, the initial deviation of the pendulum *y(0)=1.0* and the initial speed *y'(0)=0*. It is important to note that the model is linear, that is, the unknown function (and its derivatives) are included in the equation in the first degree (Samarsky , Mikhailov et al, 2001).

The method for solving the equation for the nonlinear case is given in Listing 2, and the result is shown in Fig.2. Here the symbol of the derivative is allowed to enter the tools panel, Calculus (Computing). Once again, we emphasize that the result of applying the Given/odesolve block is the function *y(t)* defined on the interval (t_0,t_1). You should use the usual Mathcad tools to plot it or get the function value at some point in the specified interval, for example: *y(10)=0.048*.

Listing 2. Modeling of a nonlinear oscillator
w: = 0.5 β:=0.2 γ : = 0.95

Given

$$w^2 \cdot \frac{d^2}{dt^2} y(t) + \beta \cdot \left(\frac{d}{dt} y(t) \right) + y(t) + \gamma \cdot y(t)^2 = 0$$

y(0)=0 y'(0)=3
y:=Odesolve (t,20)
y(10)=0.048

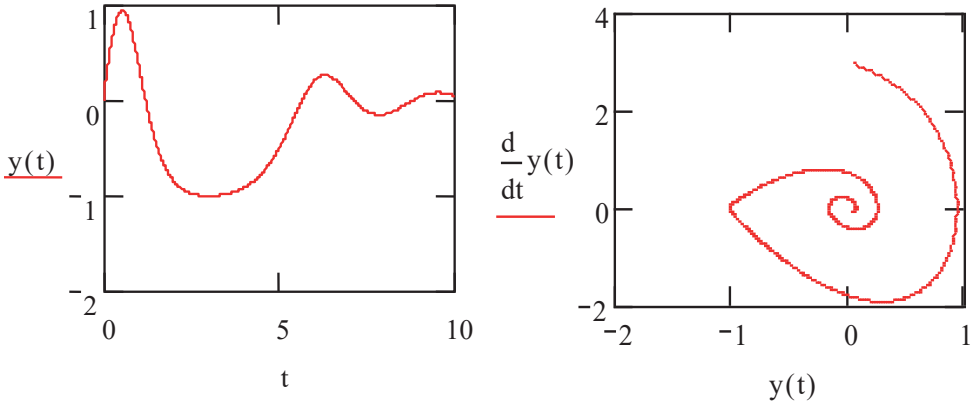


Fig 2. The model of nonlinear harmonic oscillator

Discussion. As Listing 2 shows, in addition to the equation itself, it was necessary to define two initial conditions (the third and fourth lines of the listing) – initial values of y(t) and y'(t) for t=0. Generally speaking, ordinary differential equations have a unique solution if, in addition to the equation, the initial or boundary conditions are specified in a certain way (Turin, Markov, Poyarkov, 2018).

The resulting nonlinear model is widely used in the study of sections of physics: mechanics, molecular, electricity and magnetism, vibrations and waves. Nonlinear systems with dynamic chaos are used in systems of hidden information transmission, as well as in communication systems that use dynamic chaos as a source of vibrations that carry information.

From the above, it follows that the use of modern application packages in the educational process allows you to significantly change the methodology of studying some issues of the physics course related to conducting cumbersome, repetitive computational procedures, solving systems of differential equations, plotting graphs and surfaces, with a visual representation of the results of solving the problem using application packages. If before the behavior of a physical system was analyzed exclusively analytically, now

it is possible to use numerical methods of computer simulation, which have certain advantages (Tarasova, 2014).

Computer modeling, conducting a computational experiment is one of the modern methods of studying physical phenomena. It has its own characteristics, advantages and disadvantages compared to other methods of studying physical systems. It is quite obvious that students of higher educational institutions should have ideas about computer models, numerical methods for studying various objects of cognition, and be fairly free to navigate in modern software products. It is modern application packages that make it possible to solve a complex system of equations in a few seconds, construct a graph of the studied dependence, and simulate a difficult reproducible experiment.

The advantages of modern packages are expressed in providing the ability to enter mathematical formulas or functions for numerical calculation by them, setting various values of the quantities used, plotting graphs for a visual representation of the simulation results, generating random variables (modeling random processes), performing logical operations, which allows you to implement various numerical methods. Using Mathcad, the student does not waste time coding the computational algorithm and programming auxiliary blocks i.e. saves the student from the mass of routine computing work. Also, its advantages are that Mathcad makes studying physics easier, the Mathcad program itself is easy to learn, and does not require reading thick books, conducting abstracts and memorizing complex rules for studying and applying. Mathcad is simple in that a solution to a problem of interest can be obtained in a short period of time. In this regard, the Mathcad package is very effective in the educational process, makes it possible to teach a number of educational disciplines (computer modeling of physical processes, physics, mathematical modeling, numerical methods, etc.) at a higher level (Robert et al, 1994).

The results of the work can be used when reading courses of computer modeling of physical processes, when conducting practical, lecture, laboratory classes. The complex complements traditional forms of teaching (lectures, seminars, physical laboratory) and can be used in computer classes of all universities, technical, pedagogical institutes and other higher educational institutions as a modern addition to physics courses.

Conclusion. Thus, solving the problem of the General course of physics using Mathcad allows you to form an idea of the possibilities of using this mathematical package for further research. Using the Mathcad program not only reduces the time for laboratory work, but also activates the process of

studying the discipline by students, facilitates and accelerates the assimilation of new material and control, which ultimately improves the quality and deepens student' knowledge

The integration of physics, mathematics and computer technology, and the creation of a whole set of exercises and tasks, and tasks that are specific, and not abstract from practice, will allow you to achieve a deeper understanding of the physical foundations and a more focused and meaningful development of the mathematical apparatus.

This application of mathematical modeling can result in huge cost savings and a significant reduction in research time. Mathematical modeling for the control and evaluation of design solutions and experimental methods created not only significantly improves the quality of design solutions, but also dramatically reduces the cost of creating experimental installations and conducting scientific research using them.

Information about the authors

Gulnara Beisengazievna Turebaeva – Karaganda Technical University named after Abylkas Saginov, Master of Physics, Senior Lecturer, gulnara_83.06.12@mail.ru, http://orcid/0000-0002-8800-174X;

Syzdykov Alpys Kasarbekovich – Karaganda Technical University named after Abylkas Saginov, senior lecturer, alpis_62@mail.ru http://orcid / 0000-0002-2278-2838;

Tenchurina Alfiya Rishatovna – Karaganda Technical University named after Abylkas Saginov, Associate Professor, Candidate of Chemical Sciences, altenchurina@mail.ru, http://orcid / 0000-0002-4861-8411;

Doshakova Zhanar Baizakova – Gymnasium No. 39, Master of Physics, *m29kt@mail.ru* http://orcid/ 0000-0002-1388-9377.

REFERENCE

Bursian E.V.`(2015). Tasks on physics for a computer M.: Inlightening (in Russ).

Dyakonov V.P. Handbook of MathCAD PLUS 7.0 PRO: Textbook. Moscow: SK Press, 2019. – 270c. (in Russ).

Guld K. & Tobochnik Y. (2017). Computer modeling in physics. M.: World (in Russ).

Gorbatenko A.I. Organization of the educational process based on the application of innovative teaching methods in an agricultural University // Physics and modern technologies in agriculture. Materials of the X International youth conference of young scientists, students and schoolchildren. Orel: EBS Orelgau, 2019. - Pp. 353-358. 612 (in Russ).

Kondratiev A.S. (2015). Physics. Tasks on a computer. [Physics. Tasks on a computer]. Publishing house MSTU (in Russ).

Levitskij A.A. (2016). Matlab 3.05, MathCad 2.5. Krasnayarsk: publishing house Phismatlit (in Russ).

Matros D.Sh., Polev D.M., Melnikova N.N. Quality management of education based on new information technologies: Textbook.Moscow: Pedagogical Society of Russia, 2017. - 95 p.

Methodological developments using mathematical packages. Electron. Dan. Mode of access: *http://www.exponenta.ru*. (in Russ).

Robert I.V. Modern information technologies in education: didactic problems, prospects of use: Textbook. M.: School-Press, 1994. - 205 p (in Eng).

Samarsky A.A., Mikhailov A.P. Mathematical modeling: Ideas. Methods. Examples. GL. ed. Fiz. Mat. lit., - Moscow: Nauka 2001. - 320 p. (in Russ).

Turin V.O., Markov O.I., Poyarkov V.N. Aberration of light and matter wave. / / Modern problems of physical and mathematical Sciences: Materials of the IV all-Russian scientific and practical conference with international participation. November 22-25, 2018, Orel: I.S. Turgenev OSU, - Pp. 199-207 (in Russ).

Tarasova M.A. Interdisciplinary integration of academic disciplines – an effective technology for the formation of professional and active components of competencies / Tarasova M.A., Grishina S.Yu. / / Scientific notes of the Oryol state University. Scientific journal. Series “Natural, technical and medical Sciences”. - 2014-No. 5 (61), OSU publishing house, Pp. 409-412.(in Russ).

Verzhbitsky V.M. Numerical methods. Mathematical analysis and ordinary differential equations, Moscow: Higher school of Economics, 2001, 383 p. 12. E.A. Volkov. Numerical method. Saint Petersburg: LAN, 2004. - 248 p. (in Russ).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 228-246
https://doi.org/10.32014/2022.2518-1726.149
УДК 004.716

К.С. Чежимбаева*, А.Н. Хайруллина

НАО «Алматинский университет энергетики и связи им. Г. Даукеева»,
Казахстан, Алматы.
E-mail: k.chezhimbayeva@aues.kz

ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ ПРИЕМОПЕРЕДАТЧИКА LORA

Аннотация. Сети LPWAN – относительно новая модель связи, где традиционные технологии сотовой связи и беспроводные связи малой дальности комбинируются и дополняются обработкой различных запросов в приложениях. Технология обеспечивает широкополосное подключение при низком энергопотреблении и подходят для устройств с низкой скоростью передачи данных при низкой стоимости, что является безусловным преимуществом. Прогнозируется, что рынок сетей LPWA также будет значительным, что в итоге приведет к подключению к Интернету около 30 миллиардов устройств.

LoRa / LoRaWAN в настоящее время является самой популярной технологией глобальной сети с низким энергопотреблением (LPWAN), позволяющей использовать приложения Интернета вещей (IoT). LoRa/LoRaWAN — это технология сотовой связи, поддерживающая широкий диапазон параметров связи для множества узлов. Прежде чем приступить к развертыванию Интернета вещей поверх сети LoRaWAN, рекомендуется провести исследования на основе моделирования, чтобы оптимизировать дизайн-сети LoRaWAN для рассматриваемого Интернета вещей. LoRaSim в настоящее время является самым популярным симулятором для LoRa/LoRaWAN.

Ключевыми элементами интеллектуальных электросетей являются умные счетчики, которые будут составлять основу систем мониторинга и учета энергопотребления. В частности, особый интерес представляет именно мониторинг, в котором обмен большим количеством данных

осуществляется с высокой частотой; поэтому в данной статье будет смоделирована работа интеллектуальной электросети на основе LoRa, а именно проведена оценка эффективности сети для кейса мониторинга. Таким образом, целью статьи является получение реальной картины различных аспектов применения технологии LoRa в системах мониторинга интеллектуальных сетей [7].

Ключевые слова: Интернет вещей, беспроводные сенсорные сети, модуляция, интеллектуальные системы учета ресурсов, надежность канала передачи данных, медленная волна, коэффициент замедления, энергетические и фазовые характеристики.

К.С. Чежимбаева*, А.Н. Хайруллина

Ғ. Даукеев атындағы Алматы энергетика және байланыс университеті,
Қазақстан, Алматы.
E-mail: k.chezhimbayeva@aues.kz

LORA ҚАБЫЛДАҒЫШ/ЖІБЕРГІШІНІҢ ӨНІМДІЛІГІН БАҒАЛАУ

Аннотация. LPWAN желілері салыстырмалы түрде жаңа байланыс моделі болып табылады, онда дәстүрлі ұялы байланыс технологиялары мен қысқа қашықтықтағы сымсыз байланыс қосымшалары әртүрлі сұраныстарды өңдеумен біріктіріліп, толықтырылады. Технология төмен қуат тұтынуда кең жолақты қосылуды қамтамасыз етеді және құндылығы төмен деректерді жіберу жылдамдығы төмен құрылғылар үшін қолайлы, әрі сөзсіз артықшылық болып есептеледі. LPWA желілері нарығыда маңызды болады деп болжануда, нәтижесінде 30 миллиардқа жуық құрылғы интернетке қосылады.

LoRa / LoRaWAN қазіргі уақытта интернет заттары (IoT) қосымшаларын пайдалануға мүмкіндік беретін ең танымал төмен қуатты ғаламдық желі (LPWAN) технологиясы. LoRa / LoRaWAN – бұл көптеген түйіндерге арналған байланыс параметрлерінің кең спектрін қолдайтын ұялы байланыс технологиясы. LoRaWAN желісінің үстіне заттар интернетін орналастыруды бастамас бұрын, қарастырылып отырған заттар интернеті үшін LoRaWAN желісінің дизайнын оңтайландыру үшін модельдеу негізінде зерттеулер жүргізген жөн. LoRaSim қазіргі уақытта LoRa /LoRaWAN үшін ең танымал тренажер болып табылады.

Ақылды есептеуіштер интеллектуалды электр желілерінің негізгі

элементтері, олар тұтынуды есепке алу және мониторингілеу жүйелерінің негізі. Атап айтқанда, мониторинг ерекше қызығушылық тудырады, онда көптеген мәліметтермен алмасу жоғары жиілікте жүзеге асырылады; сондықтан осы мақалада LoRa негізінде интеллектуалды электр желісінің жұмысы модельденеді, атап айтқанда, мониторинг жағдайы үшін желінің тиімділігі бағаланады. Осылайша, мақаланың мақсаты интеллектуалды желілерді бақылау жүйелерінде LoRa технологиясын қолданудың әртүрлі аспектілерінің нақты көрінісін алу болып табылады.

Түйін сөздер: заттар интернеті, сымсыз сенсорлық желілер, модуляция, интеллектуалды ресурстарды есепке алу жүйелері, деректерді жіберу арнасының сенімділігі, баяу толқын, баяулау коэффициенті, энергетикалық және фазалық сипаттамалары.

K.S. Chezimbayeva*, A.N. Khairullina

Non-ptofit JSC «Almaty University of Poer Engineering and Telecommunications named after G. Daukeyev», Kazakhstan, Almaty.
E-mail: k.chezhimbayeva@aues.kz

EVALUATION OF LORA TRANSCEIVER PERFORMANCE

Abstract. LPWAN networks are a relatively new communication model, where traditional cellular communication technologies and short-range wireless communications are combined and supplemented by processing various requests in applications. The technology provides broadband connectivity with low power consumption and is suitable for devices with low data transfer rates at low cost, which is an absolute advantage. It is predicted that the LPWA network market will also be significant, which will eventually lead to an Internet connection of about 30 billion devices.

LoRa / LoRaWAN is currently the most popular low-power wide area network (LPWAN) technology that enables Internet of Things (IoT) applications. LoRa/LoRaWAN is a cellular communication technology that supports a wide range of communication parameters for multiple nodes. Before proceeding with the deployment of the Internet of Things over the LoRaWAN network, it is recommended to conduct simulation-based research to optimize the design of the LoRaWAN network for the Internet of Things in question. LoRaSim is currently the most popular simulator for LoRa/LoRaWAN.

The key elements of intelligent power grids are smart meters, which will form the basis of energy consumption monitoring and accounting systems. In particular, monitoring is of particular interest, in which a large amount of data is exchanged with a high frequency; therefore, this article will simulate the operation of an intelligent power grid based on LoRa, namely, an assessment of the network efficiency for the monitoring case. Thus, the purpose of the article is to get a real picture of various aspects of the application of LoRa technology in intelligent network monitoring systems [7]

Key words: Internet of Things, wireless sensor networks, modulation, intelligent resource accounting systems, data channel reliability, slow wave, deceleration coefficient, energy and phase characteristics.

Введение. LoRa (Long Range Radio) – беспроводная технология, которая предназначена для сетей M2M и Интернета вещей. Эта технология позволяет общедоступным или многопользовательским сетям подключать несколько приложений, которые работают в одной сети. С помощью датчиков LoRa и автоматизированных приложений технология позволит развивать умный город [11]. LoRa Alliance - это открытый, некоммерческий Международный Альянс фирм и партнеров по обмену партнерской степенью, который разделяет миссию по институционализации организации LPWAN, которая постоянно распространяется по всему миру для настройки инноваций в области Интернета вещей и межмашинных взаимодействий, «умных» городских сообществ, и современных приложений [3].

Материалы и методы исследования. Моделирование будет осуществляться в LoRaSim. LoRaSim – это дискретно-событийный симулятор для сетей LoRa, который ориентирован на исследование масштабируемости сети и коллизий. LoRaSim написан на языке программирования Python 2.7. Симулятор включает в себя четыре скрипта для различных экспериментов: loraDir.py, loraDirMulBs.py, directionalLoraIntf.py и oneDirectionalLoraIntf.py. Скрипт loraDir.py имитирует работу одной базовой станции, loraDirMulBs.py – более чем одной базовой станции (до 24-х), loraDirMulBs.py имитирует работу ОУ с направленными антеннами и нескольких сетей, oneDirectionalLoraIntf.py имитирует работу базовых станций с направленными антеннами и нескольких сетей. Все четыре скрипта требуют, чтобы были установлены библиотеки matplotlib, SimPy и NumPy [4].

В LoRaSim можно задавать ряд параметров, таких как количество оконечных устройств, количество базовых станций (или LoRa-шлюзов), расстояние между БС, количество посторонних сетей, наличие

направленных антенн. Симулятор также позволяет проводить проверку на количество коллизий. Более того, LoRaSim может оценить потребление энергии всей сети. Однако следует иметь в виду, что симулятор рассчитывает только энергию, потребляемую радиостанцией для передачи пакетов. Он не учитывает время простоя или потребление энергии оконечными устройствами или самими радиостанциями [5].

У каждого скрипта свои входные параметры:

- loraDir.py: <NODES> <AVGSEND> <EXPERIMENT> <SIMTIME> [COLLISION];
- loraDirMulBS.py: <NODES> <AVGSEND> <EXPERIMENT> <SIMTIME> <BASESTATIONS> [COLLISION];
- directionalLoraIntf.py: <NODES> <AVGSEND> <EXPERIMENT> <SIMTIME> <BASESTATIONS> <COLLISION> <DIRECTIONALITY> <NETWORKS> <BASEDIST>;

<BASESTATIONS> <COLLISION> <DIRECTIONALITY> <NETWORKS> <BASEDIST>.

Обозначения параметров:

- NODES – количество ОУ на одну базовую станцию (шлюз);
- AVGSEND – средний интервал отправки пакетов к БС в миллисекундах;
- EXPERIMENT – параметр, определяющий основные конфигурации сети (от 0 до 5); LoRaSim предлагает шесть различных конфигураций, отличающихся друг от друга коэффициентом расширения спектра, шириной полосы пропускания, скоростью кодирования и т. д.;
- SIMTIME – длительность симуляции в миллисекундах;
- BASESTATIONS – количество базовых станций (1, 2, 3, 4, 6, 8 или 24);
- COLLISION – "1" означает полную проверку на коллизии, "0" – упрощенную (по умолчанию); при упрощенной проверке два пакета накладываются, если они прибывают в одно и то же время, на той же частоте и с тем же коэффициентом расширения спектра; полная проверка на коллизии подразумевает "эффект захвата", при котором один из двух накладывающихся пакетов все еще может достигнуть места назначения в зависимости от момента времени и разницы в мощности приема;
- DIRECTIONALITY – "1" означает, что ОУ имеют направленные антенны;
- NETWORKS – количество сетей LoRa;
- BASEDIST – расстояние между двумя базовыми станциями.

В результате каждой симуляции программа создает файл под названием "expX.dat", в котором "X" – это параметр эксперимента (от

0 до 5). Файл состоит из столбцов, разделенных пробелом: количество ОУ, коллизии, энергопотребление и т. д.

Данные из файла можно с легкостью визуализировать, например, при помощи gnuplot. В данной работе будут использоваться скрипты loraDir.py, коды даны в приложении А. Рассмотрим пять основных регулируемых параметров, определяющих работу передатчика LoRa:

1. Коэффициент расширения спектра (SF). В технологии LoRa коэффициент расширения спектра может быть определен, как соотношение между скоростью передачи элементов сигнала и символьной скоростью, которое описывается значением от 7 до 12. SF тесно связан с такими параметрами, как скорость кодирования и ширина полосы пропускания. Регулируя значение SF, можно добиться различных скорости передачи данных и времени пребывания в эфире. Большое значение коэффициента расширения означает низкую скорость передачи, но большие дальность и чувствительность приемника. К тому же, большие значения SF оказывают сильное влияние на потребление энергии оконечными устройствами.

2. Ширина полосы пропускания (BW). Согласно LoRa Alliance, ширина полосы пропускания может принимать значения 125 КГц, 200 КГц и 500 КГц. Большее значение ширины полосы пропускания обеспечивает большую скорость передачи и меньшее время эфира. Однако это уменьшает чувствительность и, как следствие, возрастает влияние на устойчивость к интерференции.

3. Скорость кодирования (CR). Скорость кодирования определяется как количество битов упреждающей коррекции ошибок (Forward Error Correction, или FEC), которые прибавляются к передаваемому пакету с целью гарантировать возможность восстановления поврежденных данных в результате интерференции. Она может принимать значения 4/5, 4/6, 4/7 или 4/8. Скорость кодирования оказывает значительное влияние на время передачи: чем больше значение CR, тем больше информации закладывается в полезную нагрузку сообщения и, следовательно, больше время передачи. Однако это также улучшает защищенность от пакетов ошибок.

4. Несущая частота (CF). Несущая частота – это центральная частота, которая может быть установлена между 137 МГц и 1020 МГц с шагом 61 Гц. Более высокая частота означает меньшее время передачи и распространения.

5. Мощность передатчика (TP). В зависимости от того, как устроен приемник LoRa, мощность передатчика может принимать значение от -4

дБм до 20 дБм с шагом 1 дБ. Особенности регулирования использования радиочастотных ресурсов и ограниченность аппаратного обеспечения наложили свой след на разрешенные мощности передатчика. Таким образом, значения мощности передатчика ограничены в пределах от 2 дБм до 20 дБм.

В таблице 1 приведены параметры конфигурации, которые будут использоваться в моделировании. В LoRaSim конфигурациям №1, 2 и 3 соответствуют параметры <EXPERIMENT> 0, 2 и 4, соответственно. Конфигурация №1 – это сеть с наименьшей скоростью передачи. Конфигурация №2 – сеть с наибольшей скоростью передачи. Конфигурация №3 использует параметры, рекомендуемые спецификацией LoRaWAN [8].

Таблица 1 - Параметры конфигурации

Параметр	Конфиг. №1	Конфиг. №2	Конфиг. №3
Количество ОУ, N	-	-	-
Размер полезной нагрузки, B	20 байт	20 байт	20 байт
Средний период отправки пакетов, λ	300 000 мс (5 мин)	300 000 мс (5 мин)	300 000 мс (5 мин)
Радиус ячейки, R	-	-	-
Мощность передатчика, TR	14 дБм	14 дБм	14 дБм
Несущая частота, CF	868 МГц	868 МГц	868 МГц
Коэффициент расширения спектра, SF	12	7	12
Скорость кодирования, CR	4/8	4/5	4/5
Ширина полосы пропускания, BW	125 КГц	500 КГц	125 КГц

Во всех экспериментах размер пакета по умолчанию будет составлять 20 байт, что является вполне достаточным для поддержки таких реальных способов применения, как интеллектуальный учет и, самое главное, мониторинг. Количество оконечных устройств N – это положительное целое число, которое будет меняться в ходе экспериментов. Все эти устройства разбросаны по территории в радиусе R вокруг базовой станции.

Согласно спецификации различных смарт-счетчиков электроэнергии, отправку пакетов можно организовать с периодом от 5 минут до одного месяца, в зависимости от выполняемых задач.

Мы предполагаем, что оконечные устройства отправляют пакеты к базовой станции каждые 900 000 миллисекунд, или 15 минут, т.к., согласно разным источникам, это самый оптимальный интервал, с которым умные счетчики должны отправлять ряд показателей энергии

в системах мониторинга. Следует также отметить, что этот временной интервал может меняться в зависимости от кейсов использования LoRa, например, автоматическое снятие показаний электропотребления в жилых домах можно проводить каждый час или даже один раз в день.

Мощность передатчика примем равным 14 дБм. Остальные параметры следующие: скорость кодирования 4/8 или 4/5, ширина полосы пропускания 125 КГц или 500 КГц, коэффициент расширения спектра SF 12 или 7, несущая частота 868 МГц. В данных экспериментах мы будем использовать следующие обозначения названных параметров: TP (мощность передатчика - transmission power), CF (несущая частота - carrier frequency), SF (коэффициент расширения спектра - spreading factor), BW (ширина полосы пропускания - bandwidth), CR (скорость кодирования - coding rate), λ (временной интервал отправки пакетов), B (размер полезной нагрузки сообщения).

Для оценки работы всей сети мы будем использовать скорость извлечения данных (Data Extraction Rate, или DER) и потребление энергии.

DER вычисляется по формуле (1) [2]:

$$DER = \frac{N_{пер} - N_{кол}}{N_{пер}}, \quad (1)$$

где $N_{пер}$ – количество переданных пакетов, $N_{кол}$ – количество коллизий.

Проще говоря, DER – это отношение количества успешно принятых пакетов к количеству переданных пакетов за определенный промежуток времени. В идеальных условиях DER = 1, т. е. все переданные пакеты корректно доставляются к базовой станции.

Экспериментальное исследование различных конфигураций сети LoRaWAN при наличии одной базовой станции.

В этом эксперименте мы изучаем работу простейшего случая сети LoRa (рисунок 1), в котором N оконечных устройств отправляют пакеты единственной базовой станции (или LoRa-шлюзу). Для данного эксперимента мы воспользуемся тремя конфигурациями, приведенными в таблице 1. Длительность симуляции составляет один день (8 640 000 мс). Радиус ячейки по умолчанию: R = 99 м.

Рисунок 1 – Сеть для эксперимента

Последняя строка на рисунке 2 – образец того, как осуществляется внесение исходных данных в программу: наименование скрипта (loraDir.py), количество ОУ (500), период отправки пакетов (900 000 мс, или каждые 15 мин), конфигурация №2 (в LoRaSim она обозначена числом 2), длительность симуляции (8 640 000 мс, или 24 ч) и проверка на коллизии (1).

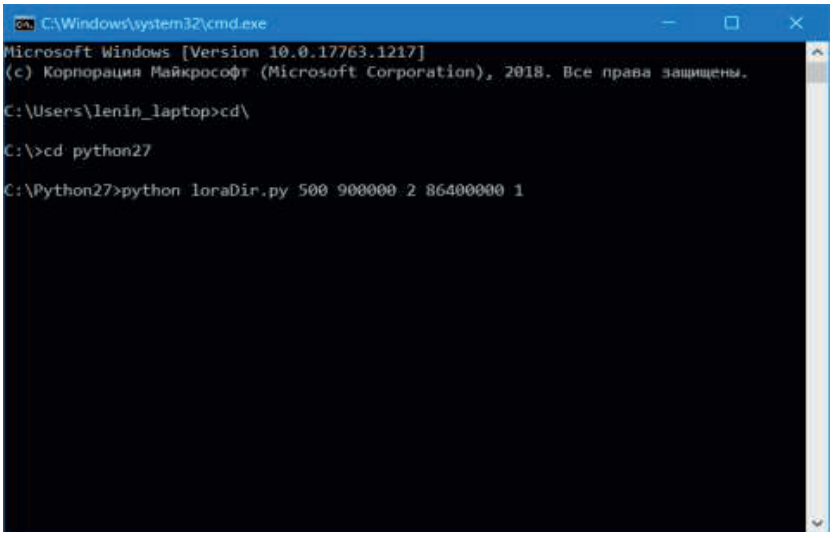


Рисунок 2 – Ввод данных

Входные данные для конфигурации №1:
python loraDir.py 1 900000 0 86400000 1
python loraDir.py 100 900000 0 86400000 1
python loraDir.py 200 900000 0 86400000 1

python loraDir.py 300 900000 0 86400000 1
python loraDir.py 400 900000 0 86400000 1
python loraDir.py 500 900000 0 86400000 1
python loraDir.py 600 900000 0 86400000 1
python loraDir.py 700 900000 0 86400000 1
python loraDir.py 800 900000 0 86400000 1
python loraDir.py 900 900000 0 86400000 1
python loraDir.py 1000 900000 0 86400000 1

Входные данные для конфигурации №2:
python loraDir.py 1 900000 2 86400000 1
python loraDir.py 100 900000 2 86400000 1
python loraDir.py 200 900000 2 86400000 1
python loraDir.py 300 900000 2 86400000 1
python loraDir.py 400 900000 2 86400000 1
python loraDir.py 500 900000 2 86400000 1
python loraDir.py 600 900000 2 86400000 1
python loraDir.py 700 900000 2 86400000 1
python loraDir.py 800 900000 2 86400000 1
python loraDir.py 900 900000 2 86400000 1
python loraDir.py 1000 900000 2 86400000 1

Входные данные для конфигурации №3:
python loraDir.py 1 900000 4 86400000 1
python loraDir.py 100 900000 4 86400000 1
python loraDir.py 200 900000 4 86400000 1
python loraDir.py 300 900000 4 86400000 1
python loraDir.py 400 900000 4 86400000 1
python loraDir.py 500 900000 4 86400000 1
python loraDir.py 600 900000 4 86400000 1
python loraDir.py 700 900000 4 86400000 1
python loraDir.py 800 900000 4 86400000 1
python loraDir.py 900 900000 4 86400000 1
python loraDir.py 1000 900000 4 86400000 1

На рисунках 3, 4, 5 отображены результаты для различных конфигураций сети LoRa в случае мониторинга. Файлы с результатами содержат информацию о количестве ОУ, коллизий, отправок, DER и затраченной на передачу энергии.

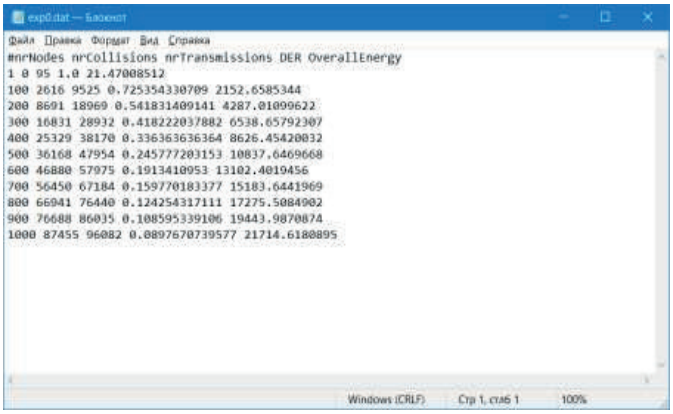


Рисунок 3 – Результат эксперимента для сети с конфигурацией №1

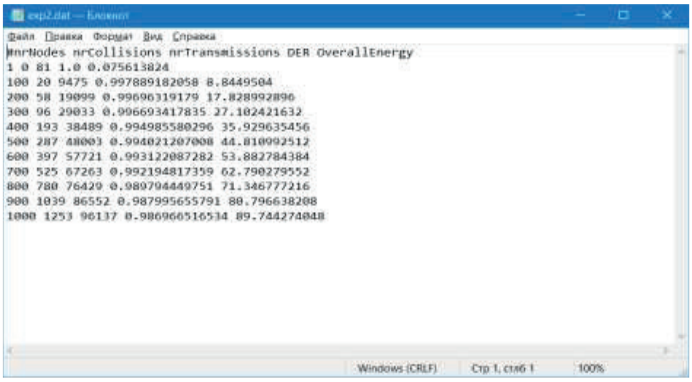


Рисунок 4 – Результат эксперимента для сети с конфигурацией №2

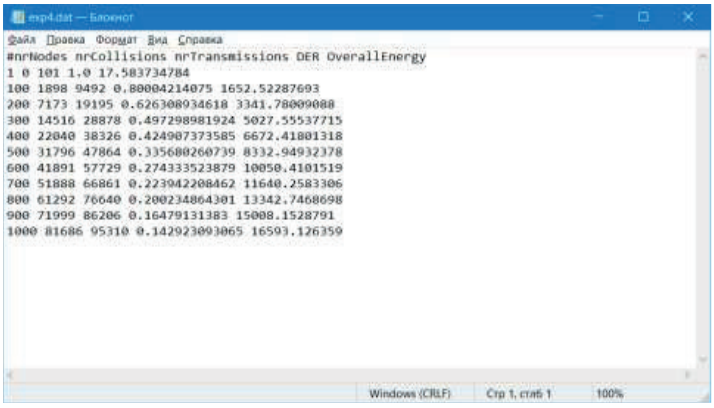


Рисунок 5 – Результат эксперимента для сети с конфигурацией №3

На рисунках 6, 7 приведено визуальное обобщение эксперимента для всех трех конфигураций сети, построенное в gnuplot [5], а именно

зависимость DER и энергопотребления от количества ОУ. Каждому маркеру соответствует результат симуляции. Отношение количества корректно принятых сообщений к количеству отправленных сообщений для каждой конфигурации различно.

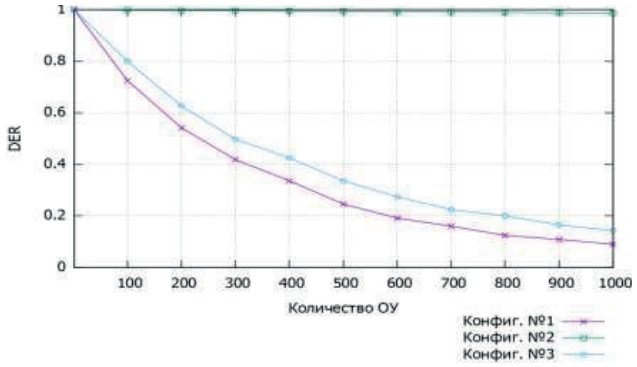


Рисунок 6 – Зависимость DER от количества ОУ

Из рисунка 6 видно, что соотношение между принятыми и отправленными сообщениями DER уменьшается по экспоненте с увеличением количества ОУ при низких скоростях передачи (конфигурации №1 и №3). Сеть с конфигурацией №2 имеет хорошие показатели работы: DER больше 0,9 при различных значениях количества оконечных устройств N.

С точки зрения энергоэффективности, сеть с конфигурацией №2 - лучший вариант (рисунок 7). Такая сеть потребляет малое количество энергии вне зависимости от числа оконечных устройств, в то время как в конфигурациях №1 и №3 наблюдается рост энергопотребления по мере увеличения N.

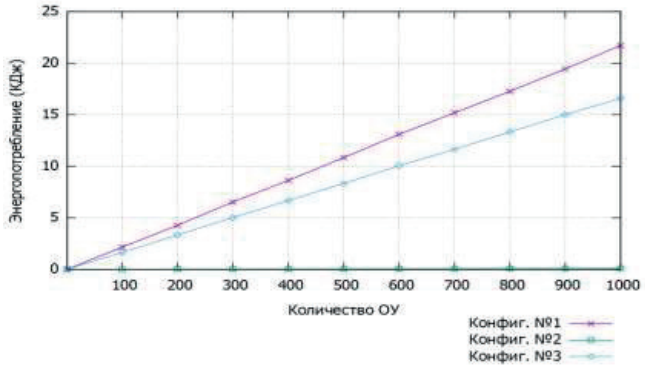


Рисунок 7 – Зависимость энергопотребления от количества ОУ

Таким образом, можно заключить, что в случае мониторинга, где энергопотребление сети и соотношение DER являются критическими факторами, конфигурация №2 - идеальное решение. Однако следует иметь в виду, что в сетях LPWAN топология может оказать как положительное, так и отрицательное воздействие на потребление энергии каждого оконечного устройства. Использование звездной топологии помогает повысить энергетическую эффективность. Вдобавок, передача вычислительных функций серверам может в дальнейшем снизить энергопотребление.

Проведение экспериментального исследования надежности канала передачи данных, основанного на технологии LoRaWAN. Одним из основных требований при проектировании каналов передачи данных в сетях Интернета вещей, является требование к надежности системы. Личные данные пользователей или конфиденциальная производственная информация должны быть защищены надлежащим образом. В свою очередь, немаловажным фактором достижения высокой надежности канала связи является обеспечение помехозащищенности канала [9].

В данной главе попытаемся экспериментальным путем исследовать помехозащищенность канала связи, основанного на данной технологии беспроводной передачи данных LoRa. В качестве основных показателей эффективности исследуемого канала связи будем использовать такие характеристики как RSSI и SNR.

RSSI или Received Signal Strength Indicator (индикатор уровня принимаемого сигнала) – один из критериев качества связи. Отображает уровень мощности принимаемого сигнала и позволяет дать оценку качества этого сигнала. Данный показатель измеряется в дБ(dB) и может принимать значения от -120 до 0 дБ. Чем ближе к нулю значение RSSI, тем выше уровень сигнала [10].

SNR или Signal-to-Noise-Ratio (отношение сигнал/шум) – отношение полезного сигнала к мощности шума. SNR измеряется в дБ и может быть вычислено по следующей формуле [1]:

$$\mathrm{SNR(dB)} = 10\log_{10}\left(\frac{\mathrm{P_{signal}}}{\mathrm{P_{noise}}}\right) = 20\log_{10}\left(\frac{\mathrm{A_{signal}}}{\mathrm{A_{noise}}}\right), \qquad (2)$$

где P – средняя мощность, а A – среднеквадратичное значение амплитуды.

Для проведения эксперимента потребуются два приемопередатчика Ra-02 (модули беспроводной передачи данных на основе трансивера

SX1278) [7] и два контроллера Arduino Uno [10]. Для создания экспериментального канала связи необходимо создать два соединения: для передающей и принимающей стороны. Сначала необходимо припаять к модулю Ra-02 соединительные провода для дальнейшего соединения с контроллером. В качестве антенны будем использовать металлическую проволоку. Соединив соответствующие контакты, получим следующую конструкцию. Конструкция одинакова и для приемника, и для передатчика.

В качестве интегрированной среды разработки (Integrated Development Environment) будет использоваться Arduino IDE. Для работы с модулем LoRa будем использовать готовую библиотеку LoRa (библиотека свободно распространяется разработчиком через Интернет).

Эксперимент № 1. Измерение в условиях прямой видимости с различными расстояниями между передатчиком и приемником: 0 м, 50 м, 100 м, 150 м, 200 м, 250 м. Полученные значения RSSI и SNR отображены на гистограммах (рисунок 8).

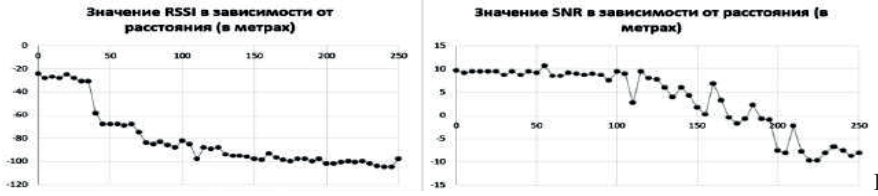


Рисунок 8 – Гистограммы значений RSSI и SNR для эксперимента № 1

Также нужно отметить, что после преодоления расстояния в 200 м наблюдаются искажения и потеря пакетов данных. Здесь уровень SNR опускается ниже -2 дБ.

Исходя из результатов проведенного эксперимента, можно сделать следующие выводы: канал связи на основе технологии LoRa способен эффективно передавать пакеты данных на расстоянии до 150 м; помехозащищенность канала связи достигает критического уровня при значениях SNR < -2 дБ.

Эксперимент № 2. Измерение при разнесении приемника и передатчика в разные помещения на расстоянии 5 м. (разделенные бетонной стеной).

Полученные значения RSSI и SNR отображены на гистограммах (рисунок 9). Значения для пакетов данных, полученных при созданных условиях, выделены прямоугольником.

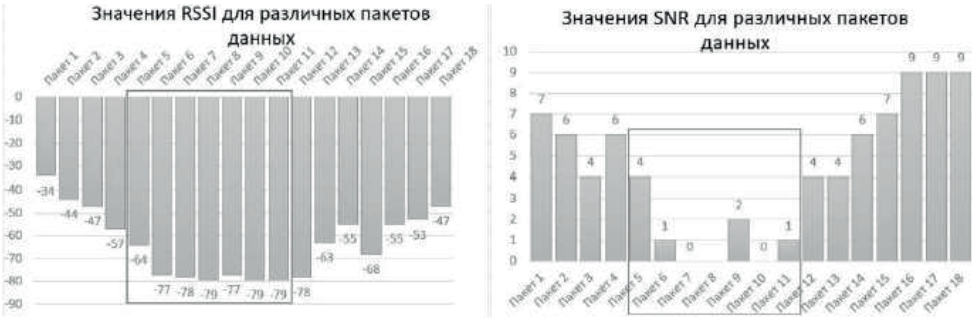


Рисунок 9 – Гистограммы значений RSSI и SNR для эксперимента № 2

Анализируя полученные результаты, можно сделать вывод о том, что физические препятствия (в нашем случае – бетонная стена) сильно влияют на канал связи. Зарегистрированные значения RSSI и SNR намного меньше, чем на тех же расстояниях в условиях прямой видимости (в сравнении с результатами, полученными в ходе эксперимента № 1). В данных условиях мощность помехи становится примерно равной мощности сигнала, соответственно, качественная передача пакетов данных становится невозможной.

Эксперимент № 3. Измерение при создании электромагнитного излучения частотой 2,45 ГГц (СВЧ-печь).

Полученные значения SNR в условиях создания помехи отображены на гистограмме (рисунок 10). Значения для пакетов данных, полученных при одновременном включении СВЧ-печи, выделены прямоугольником.

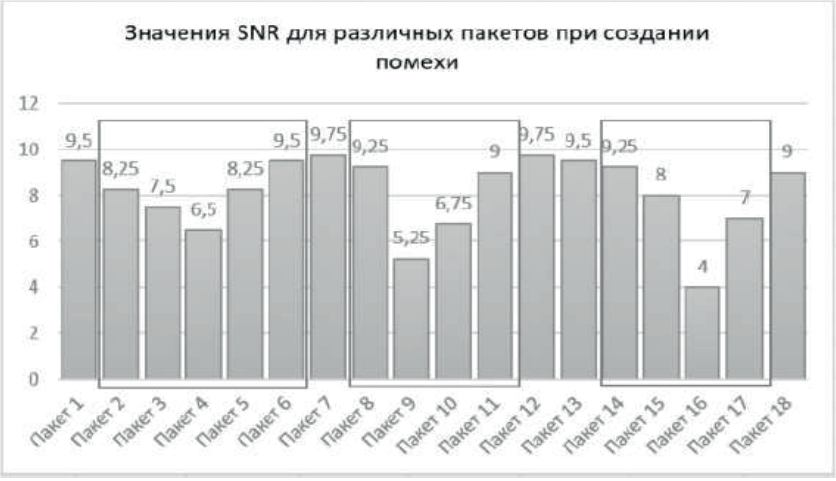


Рисунок 10 – Гистограмма значений SNR для эксперимента № 3

Результаты и обсуждение. Анализируя полученные результаты, можно сделать следующий вывод: расположенный поблизости источник высокочастотного сигнала (в данном случае СВЧ-печь) достаточно сильно влияет на помехозащищенность канала связи на основе технологии LoRa. Из гистограммы (рисунок 10) видно, что уровень SNR изменяется волнообразно и минимальный уровень зарегистрирован в момент создания помехи.

Таким образом, подводя итоги проведенных экспериментов, можно сказать, что использование канала связи на основе технологии LoRa возможно для создания сети Интернета вещей умных домов и городов, транспорта, промышленного интернета вещей и так далее. Результаты, полученные в ходе эксперимента, можно значительно улучшить использованием более мощной антенны, а также созданием сети с архитектурой, включающей промежуточные шлюзы (базовые станции). Тем не менее, использование сети Интернета вещей на основе технологии LoRa на объектах критической информационной инфраструктуры ставится под сомнение и требует гораздо более глубокого изучения и является предпосылкой к дальнейшей научной работе.

Заключение. В экспериментальной части были проведены несколько экспериментов, демонстрирующих особенности технологии LoRa, была проведена оценка эффективности работы сети LoRa с тремя различными конфигурациями. Была проведена оценка отношения количества корректно принятых пакетов к количеству переданных пакетов (DER) и энергопотребления сетей в качестве показателей работы при разных конфигурациях, чтобы определить сильные и слабые стороны технологии LoRa в конкретных случаях применения.

Используя три различные конфигурации №1, №2 и №3, было установлено, что при малых скоростях передачи данных, а именно при большом коэффициенте расширения спектра, равном 12, LoRa очень чувствительна к количеству оконечных устройств и их большое число может привести к резкому уменьшению параметра DER. Более того, в двух сценариях сети LoRa из трех, а именно при коэффициенте расширения спектра SF, равном 12, линейно возрастает энергопотребление на передачу при увеличении количества оконечных устройств, а то время, как при SF = 7 энергопотребление постоянно и минимально.

Надежность системы – одно из основных требований при проектировании каналов передачи данных в сетях Интернета вещей.

В ходе написания данной статьи были проведены три эксперимента в различных условиях для оценки помехозащищенности канала связи, основанного на технологии беспроводной передачи данных LoRa (Long Range), как одного из основных показателей надежности. В качестве основных показателей помехозащищенности был рассчитан Received Signal Strength Indicator (индикатор уровня принимаемого сигнала) и Signal-to-Noise-Ratio (отношение сигнал/шум). Для проведения были использованы два приемопередатчика Ra-02 и два контроллера Arduino Uno. Создав приемник и передатчик на основе вышеуказанных элементов, были проведены три эксперимента с условиями: изменения расстояния между приемником и передатчиком от 0 до 250 метров; преграды между приемником и передатчиком (бетонная стена); создания электромагнитных помех (сверхвысокочастотная печь).

Результаты проведенных экспериментов позволяют утверждать, что использование канала связи на основе технологии LoRa возможно для создания сетей Интернета, но использование на объектах критической информационной инфраструктуры ставится под сомнение и требует гораздо более глубокого изучения и является предпосылкой к дальнейшей научной работе.

Information about authors:

K.S. Chezimbayeva – PhD in Technical Sciences, professor, Non-profit JSC «Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeyev». E-mail: k.chezhimbayeva@aues.kz, https://orcid.org/0000-0002-1661-2226;

A.N. Khairullina – master's student of 2nd group, Non-profit JSC «Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeyev». *https://orcid.org/0000-0003-2890-1962.*

ЛИТЕРАТУРА

Bankov Dmitry, Khorov Evgeny, Lyakhov Andrey. LoRaWAN Modeling and MCS Allocation to Satisfy Heterogeneous QoS Requirements. Sensors. 2019. Vol. 19, no. 19. Pp. 1–23.

Бабаев А.А., Банков Д.В., Хоров Е.М. Анализ эффективности метода доступа к каналу в сетях LoRaWAN. Сборник трудов 40-й междисциплинарной школы-конференции ИППИ РАН «Информационные технологии и системы 2016». 2016. С. 688–694.

Дао Ч.Н. Модели концентрации трафика M2M и оценка его влияния на QOS в сетях 5G. Ч.Н. Дао, А.И. Парамонов. Электросвязь. – 2018. – № 4. – С. 47–54.

Кучерявый А.Е. Сети связи общего пользования. А.Е. Кучерявый, А.И. Парамонов, Е.А. Кучерявый. Тенденции развития и методы расчёта. ФГУП ЦНИИС, 2008.

Кучерявый А.Е. Тактильный интернет. А.Е. Кучерявый, А.И. Выборнова. Актуальные проблемы инфотелекоммуникаций в науке и образовании сборник научных статей V международной научно-технической и научно-методической конференции. – 2016. – С. 6 –11.

Кучерявый А.Е. Самоорганизующиеся сети. А.Е. Кучерявый, А.В. Прокопьев, Е.А. Кучерявый. – СПб: Любавич, 2011. – 312 с.

Мутханна А.С. Сравнительный анализ протоколов маршрутизации rpl и aodv. А.С. Мутханна, А.В. Прокопьев, А.Е. Кучерявый. Актуальные проблемы инфотелекоммуникаций в науке и образовании. II-я международная научно-техническая и научно-методическая конференция: сб. научных статей. Под ред. Доценко С.М. – СПб.: СПбГУТ, 2013. – С. 16 –171.

МСЭ-Т, Y-2060, Глобальная информационная инфраструктура, аспекты протокола интернет и сети последующих поколений. – Сектор стандартизации электросвязи МСЭ, 06/2012.

Махмуд О.А. Обзор методов передачи данных в устройствах IoT. О.А. Махмуд, Р.В. Киричек. 72-я Всероссийская научно-техническая конференция, посвященная Дню радио — СПб.: СПбГЭУ «ЛЭТИ» им. В. И. Ульянова (Ленина). – 2017. – С. 174 –175.

Хуссейн О.А. Анализ влияния технологий D2D на функционирование беспроводных сетей связи. О.А. Хуссейн, А.И. Парамонов. Информационные технологии и телекоммуникации. – 2018. – Т. 6. № 2. – С. 79 – 86.

К.С. Чежимбаева, М.Ж. Батырова Ақылды үйді модельдеу үшін деректер желісіне (IOT) жасанды интеллект ісерін зерттеу. NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN PHYSICS AND INFORMATION TECHNOLOGY SERIES ISSN 1991-346X Volume 1, Number 341 (2022), 107–116 https://doi.org/10.32014/2022.2518-1726.122.

http://www.rfwireless-world.com/Terminology/LoRa-technology-basics.html.

http://info.link-labs.com/lpwan.

REFERENCES

Bankov Dmitry, Khorov Evgeny, Lyakhov Andrey. LoRaWAN Modeling and MCS Allocation to Satisfy Heterogeneous QoS Requirements // Sensors. 2019. Vol. 19, no. 19. Pp. 1–23.

Babayev A.A., Bankov D.V., Khorov E.M. Analysis of the effectiveness of the channel access method in LoRaWAN networks. Proceedings of the 40th Interdisciplinary school-conference of IPPI RAS "Information technologies and Systems 2016". 2016. Pp. 688-694 (in Russ).

Dao, Ch.N. M2M traffic concentration models and estimation of the impact on QOS in 5G networks. Ch.N. Dao, A.I. Paramanov. Telecommunications. – 2018. – № 4. – Pp. 45-54 (in Russ).

Kucheryavyi A.E. Public communication networks. A.E. Kucheryavyi, A.I. Paramanov, E.A. Kucheryavyi. Development trends and calculation methods. FSUE TSNIIS, 2008 (in Russ).

Kucheryavyi A.E. Tactile Internet. A.E. Kucheryavyi, A.I. Vybornova. Actual problems of information and communication science and an educated collection of scientific articles

of the V Interstate Scientific-technical and scientific- methodical conference. – 2016. – Pp.6-11 (in Russ).

Kucheryavyi A.E. Self-organizing networks. A.E. Kucheryavyi, A.V. Prokopyev, E.A. Kucheryavyi. — St. Petersburg: Lyubavich, 2011. — 312 p (in Russ).

Khussein O.A. Analysis of the impact of B2B technologies on the functioning of wireless communication networks. O.A. Khussein, A.I. Paramanov. Information technology and telecommunications. – 2018. – Vol. 6. No. 2. – Pp. 79-86 (in Russ).

Mutkhanna A.S. Comparative analysis of rpl and aodv routing protocols. A.S. Mutkhanna, A.V. Prokopyev, A.E. Kucheryavyi. Actual problems of information and communication science and education. II-th international scientific-technical and scientific-methodical conference: collection of scientific articles. Edited by Docenko S.M. – St. Petersburg: SPbGUT, 2013. - Pp. 16 – 171 (in Russ).

Makhmud O.A. Overview of data transmission methods in IoT devices. O.A. Makhmud, R.V. Kirichek // 72nd All-Russian Scientific and Technical Conference dedicated to the Radio Day – STDs. SPbGEU "LETI" named after V. I. Ulyanov (Lenin). – 2017. – Pp. 174 –175 (in Russ).

ITU-T, Y-2060, Global information infrastructure, aspects of the Internet protocol and next-generation networks. –ITU Telecommunication Standardization Sector, 06/2012 (in Russ).

K.S. Chezhimbayeva, M.Zh. Batyrova Study of Artificial Intelligence in Internet of Data (IOT) for Smart Home Modeling. NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN PHYSICS AND INFORMATION TECHNOLOGY SERIES ISSN 1991-346X Volume 1, Number 341 (2022),107–116 https://doi.org/10.32014/2022.2518-1726.122.

http://www.rfwireless-world.com/Terminology/LoRa-technology-basics.html.

http://info.link-labs.com/lpwan.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 247-259
https://doi.org/10.32014/2022.2518-1726.150
UDC 14.35.07

A.G. Shaushenova[1*], A.A. Nurpeisova[1], Z.S. Mutalova[2], D.B. Dosalyanov[3], M.B.Ongarbaeva[4]

[1]S. Seifullin Kazakh Agro Technical University, Kazakhstan, Astana;
[2]Zhangir Khan Agrarian Technical University, Kazakhstan, Uralsk; [3]Narxoz University, Kazakhstan, Almaty;
[4]International Taraz Innovation Institute, Kazakhstan, Taraz.
E-mail: *Shaushenovsa_78@mail.ru*

FEATURES OF FOREIGN SYSTEMS OF VIDEO MONITORING AND IDENTIFICATION OF STUDENTS IN DISTANCE LEARNING

Abstract. COVID-19 changed the mode of life of all mankind in the spring of 2020. The epidemic also affected the education sector of Kazakhstan. Schools, colleges and universities were forced to provide distance education. At the beginning of the summer, the university chose the proctoring system for online exams. Educational institutions faced the question of choosing an automated online proctoring system that can objectively evaluate the results of students ' academic achievements.

The article is devoted to the analysis of foreign online proctoring systems used to control the knowledge of students in the conditions of distance learning. The main comparative information about organizations that supply products for online proctoring is given and the features of foreign video monitoring systems and identification of students in distance learning are described. Various functions of blocking programs offered by online proc-toring providers and automatic ways of identifying a person are described. A description of the technical support, namely web cameras, is given, indi-cating the characteristics of the resolution and cost, as well as some of their advantages and disadvantages.

Proctoring systems in distance learning demonstrate the relevance of its

application in terms of the effectiveness of such indicators as the reliability of identity verification, the reduction of time and material costs. The article has formulated promising directions for the development of the online proctoring system in Kazakhstan.

Key words: automated proctoring system, identification, online proctoring, biometric identification, video monitoring.

А.Г. Шаушенова[1*], А.А. Нурпейсова[1], Ж.С. Муталова[2], Д.Б. Досалянов[3], М.Б. Онгарбаева[4]

[1]С. Сейфуллин атындағы Қазақ агротехникалық университеті, Қазақстан, Астана;
[2]Жәңгірхан атындағы Батыс Қазақстан агротехникалық университеті, Қазақстан, Орал;
[3]Нархоз университеті, Қазақстан, Алматы;
[4]Халықаралық Тараз инновациялық институты Қазақстан, Тараз.
E-mail: *Shaushenovsa_78@mail.ru*

ҚАШЫҚТЫҚТАН ОҚЫТУДА БІЛІМ АЛУШЫНЫ ИДЕНТИФИКАЦИЯЛАУ ЖӘНЕ БЕЙНЕМОНИТОРИНГТЕУ ШЕТЕЛДІК ЖҮЙЕЛЕРІНІҢ ЕРЕКШЕЛІКТЕРІ

Аннотация. 2020 жылдың көктемінде COVID-19 бүкіл адамзаттың өмірін өзгертті. Індет Қазақстанның білім саласына да әсерін тигізді. Мектептер, колледждер мен университеттер қашықтықтан білім беруге мәжбүр болды. Жаздың басында университеттер онлайн емтихандарды тапсыру үшін прокторинг жүйесін таңдауға мәжбүр болды. Білім беру мекемелерінің алдында білім алушылардың оқу жетістіктерінің нәтижелерін объективті бағалай алатын онлайн-прокторингтің автоматтандырылған жүйесін таңдау мәселесі тұрды.

Мақала қашықтықтан оқыту жағдайында білім алушылардың білімін бақылау үшін қолданылатын шетелдік онлайн-прокторинг жүйелерін талдауға бағытталған. Онлайн-прокторингке арналған өнімдерді жеткізетін ұйымдар туралы негізгі салыстырмалы ақпарат берілген және шетелдік қашықтықтан оқытудағы білім алушыны идентификациялау және бейнемониторингтеу жүйелерінің ерекшеліктері сипатталған. Онлайн прокторинг провайдерлері ұсынатын бұғаттау бағдарламаларының әртүрлі функциялары және тұлғаны анықтаудың автоматты

әдістері сипатталған. Веб-камералардың рұқсаттылығы мен құны бойынша, сондай-ақ олардың кейбір артықшылықтары мен кемшіліктері секілді техникалық сипаттамасы беріледі.

Қашықтықтан оқытудағы прокторинг жүйелері жеке басын тексерудің сенімділігі, уақыт пен материалдық шығындардың азаюы сияқты көрсеткіштердің тиімділігі тұрғысынан оны қолданудың өзектілігін көрсетеді. Мақалада Қазақстандағы онлайн-прокторинг жүйесін дамытудың перспективалық бағыттары тұжырымдалды.

Түйін сөздер: автоматтандырылған прокторинг жүйесі, идентификация, онлайн-прокторинг, биометриялық идентификация, бейнемониторинг.

А.Г. Шаушенова[1*], А.А. Нурпейсова[1], Ж.С. Муталова[2], Д.Б. Досалянов[3], М.Б. Онгарбаева[4]

[1]Казахский агротехнический университет имени С. Сейфуллина, Казахстан, Астана;
[2]Западно-Казахстанский аграрно-технический университет имени Жангир хана, Казахстан, Уральск;
[3]Университет Нархоз, Казахстан, Алматы;
[4]Международный Таразский инновационный институт, Казахстан, Тараз.
E-mail: *Shaushenovsa_78@mail.ru*

ОСОБЕННОСТИ ЗАРУБЕЖНЫХ СИСТЕМ ВИДЕОМОНИТОРИНГА И ИДЕНТИФИКАЦИИ ОБУЧАЮЩЕГОСЯ В ДИСТАНЦИОННОМ ОБУЧЕНИИ

Аннотация. COVID-19 изменила режим жизни всего человечества весной 2020 года. Эпидемия коснулась и сферы образования Казахстана. Школы, колледжи и университеты были вынуждены предоставлять дистанционное образование. В начале лета вуз выбрал систему прокторинга для сдачи онлайн экзаменов. Перед образовательными учреждениями встал вопрос выбора автоматизированной системы онлайн-прокторинга, которая сможет объективно оценить результаты учебных достижений обучающихся.

Статья посвящена анализу зарубежных систем онлайн-прокторинга применяемых для контроля знаний обучающихся в условиях дистан-

ционного обучения. Дана основная сравнительная информация об организациях, поставляющих продукты для онлайн-прокторинга и описаны особенности зарубежных систем видеомониторинга и идентификации обучающегося в дистанционном обучении. Описаны различные функции программ блокировки, предлагаемых поставщиками онлайн-прокторинга и автоматические способы идентификации человека. Дается описание технического обеспечения, а именно веб-камер с указанием характеристик по разрешению и стоимости, а также посредством некоторых их преимуществ и недостатков. Рассмотрены разновидности процедуры прокторинга, классификация технологий на основе отечественных продуктов Oqylyq, Aero, Oes, а также преимущества и недостатки системы прокторинга. На основе анализа прокторинговых систем была подтверждена обоснованность ее применения с точки зрения эффективности с точки зрения таких показателей, как надежность проверки личности, сокращение временных и материальных затрат. Системы прокторинга в дистанционном обучении свидетельствуют об актуальности его применения с точки зрения эффективности таких показателей, как надежность проверки личности, сокращение временных и материальных затрат. В статье были сформулированы перспективные направления развития системы онлайн-прокторинг в Казахстане.

Ключевые слова: автоматизированная прокторинговая система, идентификация, онлайн-прокторинг, биометрическая идентификация, видеомониторинг.

Введение. Дистанционные образовательные технологии могут быть использованы при реализации всех форм обучения. Ключевым вопросом, замедляющим этот процесс, является низкая степень доверия к результатам обучения студента, в частности аутентификация личности при аттестации, а также соответствие условий проведения аттестации требованиям высшей школы. Данная проблема во многом решается прокторингом – специальной процедурой наблюдения и контроля за дистанционным испытанием. Технически в процессе прокторинга осуществляется визуальный контроль за студентом, программный контроль технического средства студента, аудиоконтроль окружения студента и фиксация его действий (Humbert et al., 2022).

Онлайн-прокторинг, иногда называемый удаленным прокторингом, обычно относится к прокторам, наблюдающим за экзаменом через Интернет с помощью веб-камеры. Она включает в себя также процессы, происходящие на расстоянии, для идентификации экзаменуемого как

лица, которое должно сдавать экзамен. В дополнение к этому определению онлайн-прокторинг включает в себя любые автоматизированные процессы, которые помогают обеспечить безопасность события администрирования тестирования (Phillips et al., 2005).

Термин "онлайн-прокторинг" является более описательным и предпочтительным по сравнению с удаленным прокторингом. В нем подчеркивается критическое использование Интернета и автоматизированных процессов для создания безопасного решения при мониторинге тестируемых. С другой стороны, дистанционный прокторинг – это термин, который может относиться к любому прокторингу, происходящему в ситуации, удаленной от стандартного места тестирования (например, центра тестирования или Вуза) (Seaman et al., 2018).

Онлайн-прокторинг подтверждает личность экзаменуемого, отслеживает его действия через веб-камеру и «видит», что происходит на мониторе компьютера (Camara et al., 2022). Такая технология позволяет с высокой вероятностью подтвердить личность экзаменуемого, объективно оценить его знания, исключить шпаргалки на экзамене (Kentnor, 2015).

Материалы и методы исследования. К сожалению, при бурном развитии информационных технологий не существует универсального метода, подходящего для решения всех задач распознавания, идентификации и диагностики (Griego et al., 2022). Поэтому, несмотря на богатый арсенал средств для решения задач идентификации и множество успешно решенных практических вопросов, интерес к данной теме не ослабеет. Проводится обзор отечественной продукции с анализом зарубежных систем прокторинга. Это объясняется многообразием новых производств, сложностью конкретных задач, необходимостью создания все более совершенных моделей, правильно характеризующих эти конкретные задачи. В статье методика исследования определяется новизной решения проблемы и соответствующими результатами. В научной статье рассмотрены особенности систем прокторинга Kryterion, ProctorU, Tegrity, Respondus, B Virtual, Software Secure, ProctorCam и loyalist Exam Services. В данных системах проведены сравнения по особенностям уровня звука, данных в реальном времени, блокировки, идентификации, веб-камеры, настройки программы. Проведены сравнения по требованиям к Веб-камере, видам блокировок.

Технологические альтернативы, такие как онлайн-прокторинг, становятся все более эффективными и привлекают к себе все больше внимания (Lee et al., 2020). Технологические помощники, такие как бло-

кировка компьютера/системы, контроль нажатия клавиш, возможность остановить/запустить тест и многие другие вспомогательные процессы прокторинга, были относительно легко интегрированы в процесс прокторинга.

Результат и обсуждение. Онлайн-прокторинг впервые был представлен и поддержан в США компанией Kryterion Inc. в 2006 году. Несколько других организаций последовали примеру Kryterion, это следующие программные обеспечения как ProctorU, Tegrity, Respondus, B Virtual, Software Secure, ProctorCam и Loyalist Exam Services (Miller, 2013).

Многие в индустрии тестирования, наконец, признают слабые стороны безопасности традиционного прокторинга. Например, трудно не заметить сообщения о мошенничестве в образовательных государственных программах оценки. Местные прокторы могут знать тестируемых студентов и, следовательно, иметь заинтересованность в результатах тестов, что делает тесты уязвимыми.

Местные прокторы, как правило, считаются “добровольцами”, то есть им не платят (или плохо платят), они относительно не мотивированы и плохо обучены. В индустрии тестирования с высокими ставками мало моделей, где внимание уделяется качественному прокторингу.

У некоторых поставщиков есть более одного продукта для онлайн-прокторинга. Обычно они различаются по степени предлагаемой безопасности. Например, Kryterion Online Proctoring, или OLP, обеспечивает большую безопасность, чем его аналог ProctorU. Программное обеспечение Secure имеет для тестов с высокими ставками Remote Proctor Pro, но теперь предлагает Remote Proctor для программ, требующих меньшей безопасности или желающих просто платить меньше. Другие организации предлагают единую услугу, хотя могут быть доступны варианты или настройки.

Другие соответствующие продукты/услуги:

Software Secure предлагает своим клиентам с высокими ставками программ аппаратное устройство, называемое Remote Proctor, которое включает в себя 360-градусную камеру и считыватель отпечатков пальцев.

Продукт KryterionInc. оборачивает свое решение ProctorU вокруг систем управления обучением (LMS), таких как Blackboard, обеспечивая пользователям LMS дополнительную безопасность во время сдачи экзаменов студентами.

Прокторам Kryterion не разрешается просматривать содержимое экранов рабочих станций экзаменуемых. Внутренние веб-камеры

ноутбука не могут просматривать экран, но, по крайней мере, один поставщик (ProctorU) записывает и хранит содержимое экранов. SoftwareSecure описывает своих прокторов как профессиональных прокторов, которые просматривают запись тестового сеанса после завершения теста (Таблица 1).

Таблица 1 – Особенности прокторинговых систем

Особенности прокторинга	Kryterion Inc.	Software Secure	ProctorU	B Virtual	Tegrity	ProctorCam	Loyalist Exam Services	Respondus
Онлайн проктор во время экзамена	+	-	+	+	-	+	+	+
Непрерывный Интернет	+	+	+	+	+	+	+	+
Шифрование для передачи данных	+	+	+	+	+	+	+	+
Проктор менеджмент	+	+	+	+	-	+	+	-
Взаимодействие с тестируемым	+	-	+	+	-	+	+	-
Запрет проктору просмотр экрана	+	-	-	-	+	-	+	+
Более поздний видеообзор прокторинга	-	+	-	-	+	-	-	+
Автоматический прокторинг	+	-	-	-	-	-	-	-
Уровни звука	+	-	-	-	-	-	-	-
Данные в реальном времени	+	+	+	+	+	+	+	+
Блокировка	+	+	-	-	+	-	+	+
Идентификация	+	+	+	+	+	+	+	+
Веб-камера	+	+	+	+	+	+	+	+
Журналы/записи	+	+	-	-	-	+	+	-
Хранение видео	+	+	+	+	+	+	+	+
Инцидент с отметкой времени	+	+	+	+	+	+	+	+
Журналы инцидентов	+	+	+	+	+	+	+	+
Настройка программы	+	-	-	-	-	-	+	-
Уровни решений по обеспечению безопасности	+	-	-	-	-	-	+	-
Разрешенные/указанные вспомогательные средства	+	+	+	-	-	-	+	-
Исследование эффективности	+	-	-	-	-	-	+	-

Некоторые системы онлайн-прокторинга прилагают усилия, чтобы обеспечить программу “блокировки”, но существуют большие различия в том, что это означает, и в различных задействованных компонентах. Это может относиться просто к блокировке браузера, не позволяя тестируемому получить доступ к другим URL-адресам. Или это может означать контроль над компьютером испытуемого, управление

операционной системой, обнаружение использования периферийных устройств или различных компьютерных портов. Это также может повлечь за собой использование более активных мер безопасности, таких как обнаружение нежелательных нажатий клавиш или вызовов функций (например, ctrl-alt-tab или prntscrn на компьютерах с Windows). В этой таблице предпринята попытка перечислить различные функции программ блокировки, предлагаемых поставщиками онлайн-прокторинга. В некоторых системах программы блокировки могут предоставляться третьими лицами, некоторые онлайн-системы прокторинга предлагают сторонние возможности блокировки, в то время как другие поставщики могут использовать свои собственные возможности блокировки. Некоторые онлайн-системы прокторинга не требуют или не используют блокировочный браузер. Для ProctorU прокторы имеют вид экрана рабочей станции испытуемого (что для некоторых может само по себе представлять значительную угрозу безопасности) и могут определить, пытается ли человек скопировать экран или запустить приложение или совершить какое-либо другое запрещенное действие. Для других (B Virtual и ProctorCam) неясно, как проктор может знать о типично заблокированных функциях и/или управляет ими (Таблица 2).

Таблица 2 – Особенности блокировки

Особенности блокировки	Kryterion Inc.	Software Secure	ProctorU	B Virtual	Tegrity	ProctorCam	Loyalist Exam Services	Respondus
Windows и Mac	++	++	-	-	++	-	++	++
Браузер	+	+	-	-	+	-	+	+
Запрет кнопки управления браузером	+	+	-	-	+	-	+	+
Запрет на навигацию	+	+	-	-	+	-	+	+
Предотвращение одновременных тестов	+	-	-	-	-	-	+	-
Контроль тестового выхода	+	-	-	-	+	-	+	+
Операционная Система/Компьютер	+	+	-	-	+	-	+	+
Запрет щелчка правой кнопкой мыши	+	+	-	-	+	-	+	+
Скрытие панели задач и рабочий стол	+	-	-	-	-	-	-	-
Предотвратить копирование/вставку	+	+	-	-	+	-	-	+
Запрет запуска приложений	+	+	-	-	+	-	+	+

Существует множество способов идентификации человека, в Таблице 3 перечислены те способы, которые предлагаются различными онлайн-системами прокторинга.

Идентификация в традиционных моделях тестирования является обязанностью проктора или администратора тестирования, часто одного и того же лица. В последнее время, благодаря технологическому тестированию, эта ответственность может перейти на автоматизированные процессы. Идентификация может быть хорошо обработана автоматически системой тестирования без участия человека-проктора.

Таблица 3 – Особенности идентификации

Особенности блокировки	Kryterion Inc.	Software Secure	ProctorU	B Virtual	Tegrity	ProctorCam	Loyalist Exam Services	Respondus
Идентификация	+	+	+	+	+	+	+	+
Имя Пользователя/Пароль Логин	+	+	+	+	+	+	+	+
Данные удостоверения личности/паспорт	+	+	+	+	+	+	+	+
Сравнение фотографий	+	+	+	+	+	+	+	+
Аналитика нажатий клавиш	+	-	-	-	-	-	-	-
Сложные Вопросы	+	-	+	-	-	-	-	-
Распознавание лиц	+	-	-	-	-	-	-	-
Биометрическая идентификация	-	-	-	-	-	-	-	-
Распознавания голоса	-	+	-	-	-	-	-	-
Распознавания по отпечаткам пальцев	-	-	-	-	-	-	-	-
Распознавания радужной оболочки	-	-	-	-	-	-	-	-

Все системы онлайн-прокторинга полагаются на веб-камеру со встроенным микрофоном (это могут быть отдельные функции ноутбука или планшета или автономная беспроводная или проводная камера/микрофон). Веб-камера с микрофоном в основном используется для наблюдения, общения и записи поведения, экзаменуемого во время экзамена, но может также использоваться в процессе аутентификации. Для последнего он может быть использован для облегчения программного обеспечения распознавания лиц, для захвата/сравнения фотографии испытуемого, для захвата произнесенной фразы для распознавания голоса или для фотографирования удостоверения личности. 45^0 – это угол обзора для рекомендуемой камеры Kryterion. Стандартные веб-камеры варьируются от 58° (основные) до 80° (широкоугольные). Удаленный проктор Software Secure имеет поле зрения 360° с программным обеспечением (Таблица 4).

Таблица 4 – Особенности веб-камеры

Особенности веб-камеры	Kryterion Inc.	Software Secure	ProctorU	B Virtual	Tegrity	ProctorCam	Loyalist Exam Services	Respondus
Внешняя или внутренняя камера	+ +	+ +	- +	- +	- +	- +	- +	- +
Угол обзора камеры	45^0	45	45^0	45^0	45^0	45^0	45^0	45^0
Угол обзора внешней камеры	110^0	360^0	Нет ответа	Нет ответа	Нет ответа	Нет ответа	Нет ответа	Нет ответа
Широкоформатный	+	+	+	-	-	-	+	-

Различные типы веб - камер используются сегодня в системах онлайн-прокторинга. Они различаются с точки зрения их поля зрения и того, интегрированы ли они в компьютерное оборудование, соответствующие различия для критического компонента процедур безопасности. Веб-камеры сравниваются по разрешению и стоимости, а также посредством некоторых преимуществ и недостатков (Таблица 5).

Таблица 5 – Сравнение моделей веб-камер

Модели веб-камер	Внутренняя веб-камера	70 градусов	110 градусов	360 градусов
Разрешение	Высокая	Высокая	Средняя	Низкая/ Средняя
Стоимость	0	$	$	$$$
Преимущества	Простая поддержка	Хорошее разрешение	Хорошее разрешение	Полный вид комнаты
Недостатки	Ограниченный обзор	Не вся комната видна	Не вся комната видна	Низкое разрешение; запутанное изображение

К тому же проктор не обязательно должен быть преподавателем. Его основная обязанность – следить, чтобы тестируемые не нарушали правила сдачи теста.

Рассмотрим автоматизированные системы проведения онлайн-экзаменов в учебных заведениях страны. В основном казахстанские вузы используют российские автоматизированные системы ProctorEdu, Examus и отечественные Oqylyq, Oes, Aero и др.

Казахстанская система прокторинга «Oes» также позволяет следить за ходом прохождения экзаменов. Система с помощью искусственного

интеллекта автоматически верифицирует студента и наблюдает за нарушениями с начала и до конца экзамена, затем выдает информацию в виде подробного отчета. Также система ведет запись вебкамеры, микрофона и содержимого экрана пользователя. Все видеозаписи хранятся на сервере, и можно просмотреть в любое время. С прокторинговой системой сотрудничают 30 учебных учреждений РК (Kazakh-American University, Turan University, КРУ имени А. Байтурсынова, КазГАСА, Южно-Казахстанский государственный педагогический университет, Академия Болашак, Евразийский гуманитарный институт, Атырауский инженерно гуманитарный институт, Актауский гуманитарно-технический университет и т.д.).

Казахстанская система прокторинга «Aero» проводит онлайн-экзамены с удобным мониторингом, быстрой аналитикой и подробными отчетами. В системе прокторинга «Aero» одновременно сдвали 10000 студентов. Проводит онлайн-экзамены одновременно с двух устройств на человека. С прокторинговой системой сотрудничают более 10 учебных учреждений РК (КазНПУ, Холдинг Зерде, Университет Назарбаева, Университет Ахмеда Ясави, КарГУ, Медицинский университет Астана и т.д.). С прокторинговой системой сотрудничают более 10 учебных учреждений РК (КазНПУ, Холдинг Зерде, Университет Назарбаева, Университет Ахмеда Ясави, КарГУ, Медицинский университет Астана и т.д.).

Казахстанская система прокторинга «Oqylyq», прокторинговая система включает в себя дополнительные модули автоматизированного прокторинга и антиплагиата. Это позволяет проводить весь цикл по принципу «одного окна» без перехода в сторонние системы, что дает удобство всем пользователям системы. С прокторинговой системой сотрудничают более 10 учебных учреждений РК (Казахский национальный университет им. аль-Фараби, Республиканская физико-математическая школа, Казахский Национальный педагогический университет имени Абая, Казахский Национальный женский педагогический университет и т.д.).

Несмотря на уже существующие современные разработки и технологии в этой области, проблема предупреждения фальсификаций итогов тестирования остается не до конца разрешенной и материальные ресурсы не всех вузов могут позволить себе приобрести или арендовать данный продукт.

Отечественные прокторинг-платформы разработаны без достаточного исследования научно-методологических основ организации

контроля во время экзаменов в условиях дистанционного обучения; не разработаны психолого-педагогические рекомендации организации контроля во время экзаменов в условиях дистанционного обучения. Кроме того, при совершенствовании отечественных платформ-прокторингов стоит учитывать следующие моменты:

- запрет проктору просмотр экрана;
- автоматический прокторинг;
- запрет кнопки управления браузером;
- предотвращение одновременных тестов;
- запрет щелчка правой кнопкой мыши;
- скрытие панели задач и рабочий стол;
- предотвратить копирование / вставку.

Вывод. Контроль знаний студентов в условиях дистанционного обучения является особенно актуальным. Для объективной оценки знаний обучающихся необходимо использовать прокторинг-системы.

В условиях постоянного совершенствования технологий организации образования должны оперативно осваивать современные технологические новшества, особенно технологии онлайн-прокторинга, направленные на обеспечение высокого качества подготовки обучающихся, в том числе в процессе дистанционного обучения.

Прокторинг позволяет повысить надежность и достоверность результатов обучения студентов. Онлайн-прокторинг контролирует процесс соблюдения студентами правил при сдаче онлайн-экзаменов (для самостоятельного выполнения заданий и не использования внешних материалов и дополнительных ресурсов). С развитием цифровых технологий в образовании прокторинг становится все более востребованным, в связи с чем необходимо продолжить изучение возможностей оптимизации данного процесса. Показаны особенности систем: уровень звука, данные в реальном времени, блокировка, идентификация, веб-камера, настройка программы. В научной статье анализ систем прокторинга Kryterion, ProctorU, Tegrity, Respondus, B Virtual, Software Secure, ProctorCam и Loyalist Exam Services находит применение в совершенствовании отечественных систем.

Признание. *Данная научная статья подготовлена в рамках проекта №AP09259657 «Исследование и разработка автоматизированной системы прокторинга для контроля знаний студентов в условиях дистанционного обучения» по программе 217 «Развитие науки», подпрограмме 102 «Грантовое финансирование научных исследований».*

Information about the authors:

Shaushenova Anargul – Candidate of Technical Sciences, Head of the Department of Information Systems, S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan, Shaushenova_78@mail.ru, https://orcid.org/0000-0002-3164-3688;

Nurpeisova Ardak – Doctor PhD, Senior teacher of Department of Information System, S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan, Has an H-index: 2 (Scopus) (Scopus ID: 57220128907, https://www.scopus.com/authid/detail.uri?authorId=57220128907), nurpeisova.ardak81@gmail.com, https://orcid.org/ 0000-0002-1245-8313;

Dossalyanov Damir – Doctor PhD, Public and local management, Narxoz University, Almaty, Kazakhstan, ms_018@mail.ru, https://orcid.org/0000-0001-9796-4049;

Ongarbayeva Maral – International Taraz Innovation Institute Head of the Department of Information and Communication Technologies, Taraz, Kazakhstan, ongarbaevam@mail.ru, https://orcid.org/0000–0003–0698–666X.

REFERENCES

Camara W.J., Mattern K. Inflection Point: The Role of Testing in Admissions Decisions in a Postpandemic Environment. 2022. Educational Measurement: Issues and Practice. 41(1), c. 10-15.

Grieco M., Elmore U., Vignali A., Caristo M.E., Persiani R. Surgical Training for Transanal Total Mesorectal Excision in a Live Animal Model: A Preliminary Experience. 2022. Journal of Laparoendoscopic and Advanced Surgical Techniques. 32(8), c. 866-870.

Humbert M., Lambin X., Villard E. The role of prior warnings when cheating is easy and punishment is credible. 2022. Information Economics and Policy. 58,100959.

Kentnor H. Distance Education and the Evolution of Online Learning in the United States. Curriculum and Teaching Dialogue, 2015, vol. 17, no. 1–2, pp. 28–29.

Lee K., Fanguy M. Online exam proctoring technologies: Educational innovation or deterioration? British Journal of Educational Technology. 2020.53(3), c. 475-490.

Miller G. Et al. Leading the e-learning transformation of higher education: Meeting the challenges of technology and distance education. Stylus Publishing, LLC, 2013.

Phillips P.J., Flynn P.J., Scruggs T., Bowyer K.W., Chang J., Hoffman K., Worek, W. Overview of the face recognition grand challenge // Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference. – IEEE, 2005. – Vol. 1. – p.947-954.

Purohit H., Ajmera P.K. Multi-modal biometric fusion based continuous user authentication for E-proctoring using hybrid LCNN-Salp swarm optimization. 2022. Cluster Computing. 25(2), c. 827-846.

Seaman J.E. Grade Increase Tracking Distance Education in the United States / J.E. Seaman, I.E. Allen, J. Seaman. – Oakland; Babson Survey Research Group, 2018. – 49 p.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 3, Number 343 (2022), 260-274
https://doi.org/10.32014/2022.2518-1726.151
УДК 004.4, 004.8, 004.6, 004.89:004.4

К. Якунин[1,2,*], Р.И. Мухамедиев[1,2], М. Елис[1,2], Я. Кучин[1,2], А. Сымагулов[1,2], Н. Юничева[1], Е. Мухамедиева[1]

[1]Институт информационных и вычислительных технологий МОН РК, Казахстан, Алматы;

[2]Satbayev University, Казахстан, Алматы;
E-mail: ykuchin@mail.ru

АНАЛИЗ ТЕМАТИЧЕСКИХ КЛАСТЕРОВ ПУБЛИКАЦИЙ СМИ РЕСПУБЛИКИ КАЗАХСТАН ПО ТЕМЕ ПАНДЕМИИ COVID-19

Аннотация. В настоящей работе сформирован корпус документов по данным русскоязычных СМИ Казахстана с помощью автоматического скрапинга. Корпус состоит из 761831 документа, которые относятся к ведущим новостным изданиям страны. Одним из основных инструментов, применяемых для анализа крупных корпусов текстов, является тематическое моделирование. Наиболее часто для формирования тематической модели исследователи используют так называемое латентное размещение Дирихле (LDA).

Мы использовали ARTM – расширение LDA, отличие которого заключается в применении конфигурируемых регуляризаторов, которые позволяют тонко настроить желаемый результат модели: в том числе уменьшить/увеличить склонность модели к включению слова и/или документа сразу в несколько топиков, изменить склонность модели к большему/меньшему количеству ненулевых весов в итоговой матрице. Анализ результатов показывает, как меняется отношение общества к проблемам COVID-19 в 2021-2022 годах. Во-первых, результаты отражают устойчивую тенденцию снижения интереса электронных СМИ к теме пандемии, хотя и в неравной степени для разных тематических групп. Во-вторых, выявилась тенденция к переносу

фокуса внимания на более прагматичные вопросы, такие как вопросы удалённого обучения, удалённой работы, влияния карантинных ограничений на экономику.

Ключевые слова: машинное обучение (machine learning), геопространственный искусственный интеллект (geoAI), здравоохранение (health care), мультикритериальный анализ решений (multi criteria decision analysis – MCDA), объяснимое машинное обучение (explainable machine learning), обработка естественного языка (natural language processing).

К. Якунин[1,2*], Р.И. Мухамедиев[1,2], М. Елис[1,2], Я. Кучин[1,2], Н. Юничева[1], А. Сымагулов[1,2], Е. Мухамедиева[1]

[1]Ақпараттық және есептеуіш технологиялар институты, Қазақстан, Алматы;
[2]Сатбаев университеті, Қазақстан, Алматы.
E-mail: ykuchin@mail.ru

КОВИД-19 ПАНДЕМИЯСЫ ТАҚЫРЫБЫ БОЙЫНША ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БАҚ БАСЫЛЫМДАРЫНЫҢ ТАҚЫРЫПТЫҚ КЛАСТЕРЛЕРІН ТАЛДАУ

Аннотация. Осы жұмыста Қазақстандағы орыстілді БАҚ деректері бойынша автоматты қырғышты қолдану арқылы құжаттар корпусы қалыптастырылды. Корпус еліміздің жетекші ақпарат агенттіктеріне тиесілі 761 831 құжаттан тұрады. Үлкен мәтіндік корпусты талдау үшін қолданылатын негізгі құралдардың бірі тақырыптық модельдеу болып табылады. Көбінесе зерттеушілер тақырыптық модельді қалыптастыру үшін жасырын Дирихле орналастыру (LDA) деп аталады.

Бұл жұмыста ARTM қолданылды - LDA кеңейтімі, оның айырмашылығы үлгінің қалаған нәтижесін дәл келтіруге мүмкіндік беретін конфигурацияланатын регуляризаторларды пайдалануда жатыр: соның ішінде модельдің сөзді және/немесе қосуға бейімділігін азайту / арттыру. Соңғы матрицадағы нөлдік емес салмақтар санының көп/кеміне модель үрдісін өзгерту арқылы бірден бірнеше тақырыпта құжат. Нәтижелерді талдау 2021-2022 жылдары қоғамның COVID-19 проблемаларына деген көзқарасының қалай өзгеретінін көрсетеді. Біріншіден, нәтижелер әртүрлі тақырыптық топтар үшін бірдей емес

дәрежеде болса да, электронды БАҚ-тың пандемия тақырыбына қызығушылығының тұрақты төмендеу үрдісін көрсетеді. Екіншіден, назарды қашықтықтан оқыту, қашықтан жұмыс істеу және карантиндік шектеулердің экономикаға әсері сияқты прагматикалық мәселелерге аудару үрдісі байқалды.

Түйін сөздер: Машиналық оқыту (machine learning), геокеңістіктік жасанды зият (geoAI), Денсаулық сақтау (health care), мультикритериалдық шешімдерді талдау (multi criteria decision analysis-MCDA), түсінікті машиналық оқыту (explainable machine learning), табиғи тілді өңдеу (natural language processing).

K. Yakunin[1,2*], R.I. Mukhamediev[1,2], M. Elis[1,2], Ya. Kuchin[1,2], N. Yunicheva[1], A. Symagulov[1,2], E. Mukhamedieva[1]

[1]Institute of Information and Computational Technologies, Kazakhstan, Almaty;
[2]Satbayev University, Kazakhstan, Almaty.
E-mail: ykuchin@mail.ru

ANALYSIS OF THEMATIC CLUSTERS OF KAZAKHSTAN MEDIA PUBLICATIONS ON THE TOPIC OF THE COVID-19 PANDEMIC

Abstract. In this paper, a corpus of documents based on the Russian-language media of Kazakhstan was formed using automatic scraping. The corpus consists of 761831 documents, which belong to the leading news publications of the country. One of the main tools used to analyze large corpora of texts is thematic modeling. Most often researchers use the so-called latent Dirichlet placement (LDA) to form a thematic model.

We used ARTM, an extension of LDA, the difference of which lies in the use of configurable regularizers that allow us to fine-tune the desired model result: including reducing/increasing the propensity of the model to include a word and/or document in several topics at once, changing the model propensity to have more/less non-zero weights in the resulting matrix. Analysis of the results shows how public attitudes toward COVID-19 issues change between 2021 and 2022. First, the results reflect a steady downward trend in electronic media interest in the pandemic topic, albeit to an unequal degree for different thematic groups. Second, there is a tendency to shift the

focus to more pragmatic issues, such as distance learning, telecommuting, and the impact of quarantine restrictions on the economy.

Key words: Machine Learning, geospatial artificial intelligence (geoAI), health care, multi criteria decision analysis – MCDA, explainable machine learning, natural language processing.

Введение. Период пандемии ясно показал, что системы здравоохранения почти во всем мире сталкиваются с многочисленными проблемами, связанными с повышенным спросом на медицинские услуги, высокими ожиданиями населения и увеличивающимися расходами (Атун Р., 2015:2). При этом обнаруживается, что в системе здравоохранения важны не только экономическая, но и социальная, и медицинская эффективность, поскольку, как отмечается в (Орлов и др., 2010:6), «медицинские мероприятия лечебно-профилактического характера могут быть экономически невыгодными, но медико-социальный эффект требует их проведения». И это особенно справедливо в условиях пандемии. Однако пандемия COVID-19 является хорошим примером того, как слухи и неполнота знаний влияют на общество. Массмедиа для людей становится основным источником информации, и они ощущают неопределенность, когда в окружающей среде возникают угрозы (Болл-Рокич и др., 1976:18). В соответствии (Казахстан и COVID-19: «Делюкс Типография», 2021) пандемия спровоцировала всплеск слухов и дезинформации, которые препятствовали рациональному поведению населения и в некоторой степени способствовали ускорению распространения вируса. Во время пандемии COVID-19 сообщения СМИ существенно повлияли на эмоции людей и их психологическую устойчивость (Гири и др., 2021). В этот период более 51% заголовков новостей в англоязычных СМИ были негативными и только около 30% – позитивными (Аслам и др., 2020:8). Подаваемая в таком ключе информация может вызывать отрицательные эмоции у большого числа людей и может представлять угрозу для психики человека (Хамидин и др., 2020).

Материалы. В результате сопутствующего стресса страдает иммунитет и человек становится более восприимчивым к инфекционным заболеваниям. Значительную часть информации население получает с помощью электронных СМИ, поскольку большая часть населения планеты является пользователями Интернет, например, в Казахстане – практически все взрослое население (14,73 млн.) (Кемп, 2020). Поэтому оценка объективности и качества подачи материалов электрон-

ных СМИ в период пандемии позволяет понять, как СМИ реагируют на текущую ситуацию и на необходимые мероприятия в системе здравоохранения. Эта оценка может отражать качество подачи материалов, «эмоциональную перегретость» информации и может использоваться для коррекции публикуемых материалов с целью увеличения эмоционально-психологической устойчивости читателей в период серьезных социальных потрясений. Такую оценку можно провести с использованием одного из подразделов искусственного интеллекта (AI) - методов обработки естественного языка (natural language processing – NLP) (Мухамедиев и др., 2021:26) (рисунок 1):

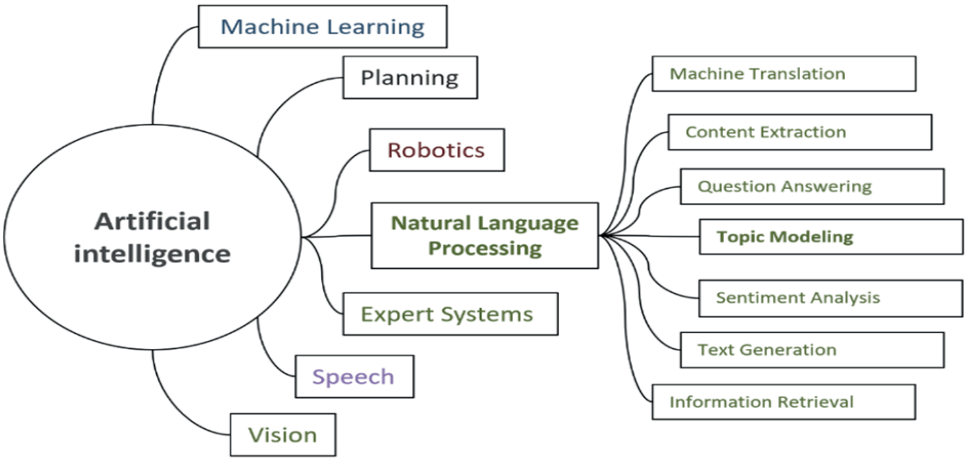


Рисунок 1. Области ИИ и НЛП

Одним из методов, продуктивно применяемых в области NLP, является тематический анализ или тематическое моделирование. Тематическое моделирование – метод, основанный на статистических характеристиках коллекций документов, который применяется в задачах автоматического реферирования, извлечения информации, информационного поиска и классификации (Машечкин и др., 2013:12). Смысл данного подхода заключается в интуитивном понимании того, что документы в коллекции образуют группы, в которых частота встречаемости слов или сочетаний слов различается.

Основой современных тематических моделей является статистическая модель естественного языка. Вероятностные тематические модели

описывают документы дискретным распределением на множестве тем, а темы – дискретным распределением на множестве терминов (Воронцов и др., 2012::13). Другими словами, тематическая модель определяет, к каким темам относится каждый документ и какие слова образуют каждую тему. Кластеры терминов и фраз, формируемые в процессе тематического моделирования, в частности, позволяют решать задачи синонимии и полисемии терминов (Пархоменко и др., 2017:139).

Недавние исследования предлагают использовать для тематического моделирования векторное представление слов (word embeddings), что позволяет учесть контекст, в котором используется тот или иной термин (Диенг и др., 2020:14).

Для построения тематической модели корпуса документов применяют: Вероятностный латентно-семантический анализ (PLSA), весьма популярное, латентное размещение Дирихле (LDA) (Блей и др., 2003:29), ARTM (Additive regularization of topic models) (Воронцов и др., 2015:11).

Иными словами, методы NLP можно использовать и в контуре обратной связи здравоохранения для оценки того, как СМИ и общество реагируют на предпринимаемые меры в чрезвычайных обстоятельствах.

При этом мы исходим из следующих предположений. Во-первых, СМИ влияет на общественное мнение, а, следовательно, те вопросы, которые освещаются в СМИ, начинают сильнее волновать общество. Во-вторых, СМИ, как и всякий бизнес, пытается предоставлять продукт (публикации) в соответствии с общественным спросом, и освещают те вопросы, которые волнует общество в большей степени. Поэтому в данной работе проведен анализ тематических кластеров публикаций СМИ Республики Казахстан по теме пандемии COVID-19 на основе автоматического скрапинга и использовании конфигурируемых регуляризаторов.

Метод анализа СМИ. Для решения поставленной задачи исследован основной аспект, связанный с представлением динамики публикационной активности СМИ в числовой форме: автоматически формируемые темы. Метод включает следующие основные шаги, продемонстрированные на рисунке 2:

1) Формирование корпуса документов с помощью автоматических систем сбора данных.

2) Создание иерархической тематической модели с применением методов, описанных в (Мухамедиев и др., 2020).

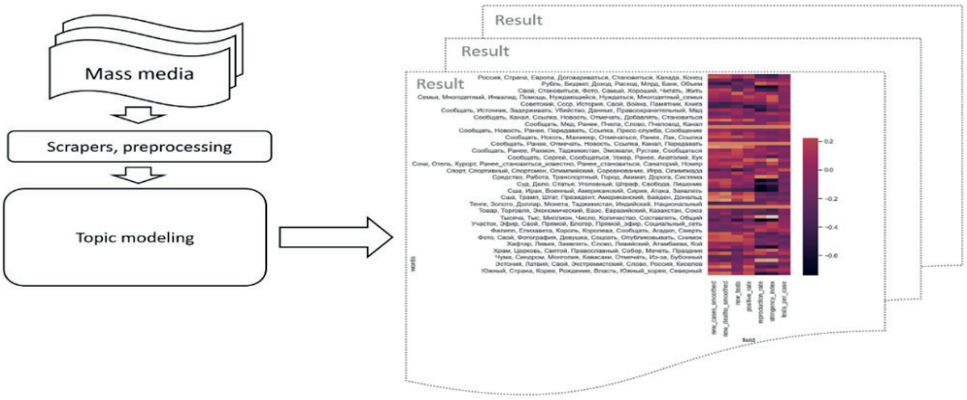


Рисунок 2. Метод оценки средств массовой информации

Рассмотрим перечисленные шаги подробнее.

- Корпус документов сформирован по данным русскоязычных СМИ Казахстана с помощью автоматического скрапинга (Якунин и др., 2021:31). Корпус состоит из 761831 документов относящихся к ведущим новостным изданиям страны.

- Создание иерархической тематической модели. Одним из основных инструментов применяемых для анализа крупных корпусов текстов является тематическое моделирование. Наиболее часто для формирования тематической модели исследователи используют так называемое латентное размещение Дирихле (LDA) (Джелодар, 2018:42).

В настоящей работе мы использовали ARTM – расширение LDA, отличие которого заключается в применении конфигурируемых регуляризаторов, которые позволяют тонко настроить желаемый результат модели: в том числе уменьшить/увеличить склонность модели к включению слова и/или документа сразу в несколько топиков, изменить склонность модели к большему/меньшему количеству ненулевых весов в итоговой матрице и т.п.

Иными словами, используется подход аддитивной регуляризации (ARTM), в котором максимизация логарифма правдоподобия, восстанавливающая исходное распределение слов W по документам D, добавляется к взвешенной сумме регуляризаторов (2), по многим критериям:

$$\sum_{d \in D} \sum_{w \in D} n_{dw} \ln \sum_{t \in T} \varphi_{wt} \theta_{td} + R(\varphi, \theta) \rightarrow \max \qquad (1)$$

$$R(\varphi, \theta) = \sum_{i=1} \tau_i R_i(\varphi, \theta) \quad (2)$$

где n_{dw} количество вхождений слова w в документе d φ_{wt} распределение слова в теме t, θ_{td} распределение темы t по документам d. Это слагаемое, $\sum_{i=1} \tau_i R_i(\varphi, \theta)$, является взвешенной линейной комбинацией регуляризаторов с неотрицательными весами τ_i. ARTM предлагает набор регуляризаторов, реализованных на основе дивергенции Кульбака-Лейблера, в данном случае демонстрирующих энтропийные различия между распределениями исходной матрицы p'(w|d) и модели p'(w|d): Сглаживающий регуляризатор, Уменьшающий регуляризатор, декоррелирующий регуляризатор.

Результаты и обсуждение. На основе библиотеки BigARTM (Воронцов и др., 2015:11), сформирована тематическая модель, состоящая из 200 кластеров документов, из которых экспертами было выбрано 12 тех, которые относятся к медицине. Эти 12 кластеров образовали субкорпус из 119956 документов, на котором было вновь выполнено тематическое моделирование с формированием 150 тематических групп из которых затем вновь отобраны 47 наиболее релевантных с точки зрения матрицы принадлежности (порог выше 0.05). Это позволило сформировать финальную модель на субкорпусе из 100481 документа. В каждый получившийся кластер помещены тексты с принадлежностью более 0.1. Итоговая тематическая модель с наиболее релевантными каждому кластеру словами показана на рисунке 3.

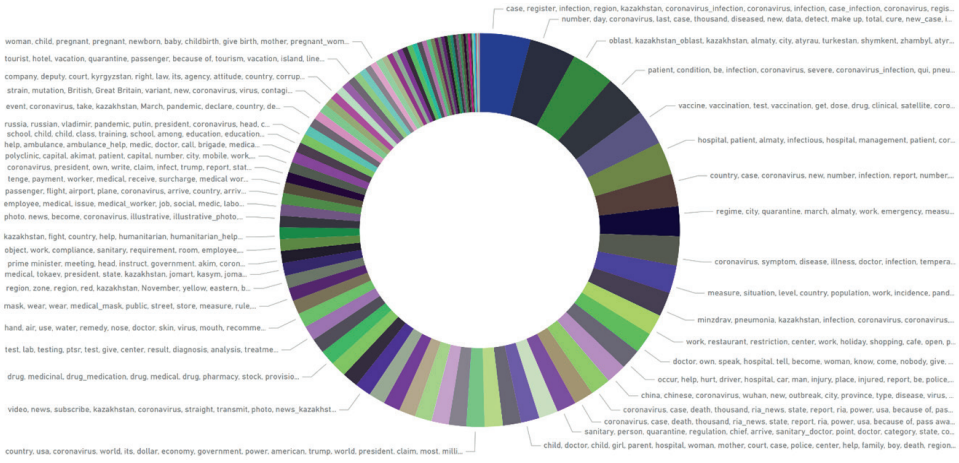


Рисунок 3. Тематические группы текстов, ранжированные по количеству документов

Количество тематических кластеров (топиков) и другие параметры тематической модели задаются вручную, обычно исходя из объективных метрик качества (Perplexity, Sparsity Score (Воронцов и др., 2015:11) и так называемого принципа плеча (Маруто и др., 2018:6). Соответственно, ввиду ограниченности количества топиков, некоторые субъективно важные темы не обязательно будут сформированы автоматически.

По этой причине, предлагается дополнительно использовать набор сформированных вручную поисковых запросов, которые позволяют проверить отдельные гипотезы о связи публикационной активности определенной тематики с эпидемиологической ситуацией.

В (Якунин и др., 2021:23) были предложены запросы, представлены ниже:

- Фальсификация, дезинформация, антивакцинация;
- Безработица, бедность;
- Кризис, экономический упадок;
- Голод, голод, бездомные, нищета;
- Удаленное образование;
- Фриланс, удаленная работа, утечка мозгов;
- Криминал, грабежи, кражи, убийства;
- Кризис, кредитование, долги, микрокредиты;
- Здравоохранение, больницы, проблемы, скандалы в здравоохранении;
- Вакцинация, вакцины COVID-19.

В настоящей работе динамика публикационной активности анализируется по этому же списку запросов. Поиск по этим запросам (на русском языке) выполнялся с помощью ElasticSearch с использованием метода полнотекстового поиска, который возвращал список совпадающих документов с относительными весами релевантности.

Например, проанализирован тематический кластер «Заболеваемость, Школа, Ребенок, Рост, Эпидемиологический». Определено, что в 2020 году тема удалённого обучения для школьников активно поднималась только с началом учебного года (в сентябре).

Аналогичная ситуация с еще большей амплитудой наблюдалась в 2021 году. Можно предположить, что население в целом ожидало ослабления карантинных мер и переход на очное обучение, однако ухудшение эпидемиологической ситуации вызвало рост интереса к данной теме.

Рассмотрен пример анализа динамики по поисковому запросу «Вакцинация, Вакцины, Прививка, COVID». В то же время резкий

рост заболеваемости начала 2022 года, связанный с распространением омикрон-штамма коронавируса, напротив, не вызвал отклика в медиапространстве, а динамика публикационной активности продолжила снижение, в соответствии с общим трендом уменьшения интереса к теме COVID-19.

Далее проанализирована динамика публикационной активности по топику, отражающему доступность и цены лекарственных средств. Анализ показал, что, например, к весне 2021 постепенно снижается публикационная активность по этой теме, что может свидетельствовать как о стабилизации ситуации с поставками лекарственных средств, так о и снижении напряжённости в обществе по поводу поставок лекарственных препаратов во время пандемии.

При рассмотрении тематического кластера «Фейк, ложная информация, дезинформация» определено, что тема фейков в области здравоохранения продолжает оставаться актуальной, и в отличии от темы вакцинации, лекарственных средств и общей эпидемиологической обстановки, публикационная активность не затухает.

Для анализа динамики интереса СМИ к теме пандемии в целом был выбран самый крупный топик, прямо относящийся к COVID-19, и рассмотрена динамика его изменения с начала 2020 года по 23 февраля 2022 года. На рисунке 4 показана эта динамика.

Рисунок 4. График нормированной публикационной активности по топику «Случай, Инфекция, Коронавирусный, Коронавирусный_Инфекция, Коронавирус, Зарегистрировать».

По рисунку 4 видно, что интерес к COVID-19 стабильно падает. Так, если сравнить январь 2021 и январь 2022 года, интерес к данной теме упал примерно в два раза. Значение на оси ординат представляет

собой отношение суммы весов документов в этом топике к сумме всех весов по всем топикам за данный период. То есть, значение может интерпретироваться как доля данной темы в потоке информации. Видно, что в пике интереса в начале 2020 года единственный из топиков имеющий отношение к COVID-19 занимал около 8% всей информации в масс-медиа, а на начало 2022 года этот показатель упал до 1%. Цифра в один процент является значимой, но вполне сравнимой с другими темами. Например, тема искусственного интеллекта, оцененная подобным образом, составляет от 1 до 5 процентов. Если учесть все топики, связанные с COVID-19, то общая доля информации СМИ, касающаяся пандемии достигала в пике 10-15%.

Заключение. В работе на основе сформированного с помощью автоматического скрапинга корпуса документов по данным русскоязычных СМИ Казахстана относящихся к ведущим новостным изданиям страны выполнен тематический анализ публикаций относящихся к COVID-19. Мы использовали ARTM – модификацию латентного размещения Дирихле LDA, отличие которого от оригинала заключается в применении конфигурируемых регуляризаторов, которые позволяют настроить желаемый результат модели. Например, с его помощью можно уменьшить либо увеличить склонность модели к включению слова или документа сразу в несколько топиков; изменения склонности модели к большему либо меньшему количеству ненулевых весов в итоговой матрице и т.п. Проведенный анализ результатов показал изменение отношения общества к проблемам COVID-19 в период последних двух лет: появилась общая устойчивая тенденция снижения интереса электронных СМИ к теме пандемии на фоне постепенного увеличения интереса к другим аспектам общественной жизни.

Задачей ближайших исследований является разработка методики корреляционного анализа между информационными трендами в электронных СМИ Казахстана и показателями эпидемиологической ситуации по COVID-19 по данным Всемирной организации здравоохранения (ВОЗ).

Благодарности. *Данная публикация профинансирована Комитетом по науке Министерства образования и науки Республики Казахстан, грант № AP09259587 "Разработка методов и алгоритмов интеллектуальной ГИС для многокритериального анализа данных здравоохранения".*

Information about authors:

Yakunin Kirill – PhD, senior researcher of the Institute of Information and Computational Technologies of the Ministry of Education and Science of the Republic of Kazakhstan, yakunin.k@mail.ru, *https://orcid.org/0000-0002-7378-9212;*

Mukhamediev Ravil – PhD, professor, K.I. Satpayev KazNITU, professor, ravil.muhamedyev@gmail.com, *https://orcid.org/0000-0002-3727-043X;*

Symagulov Adilkhan – PhD student, software engineer, ISRT KN MES RK; asmogulove00@gmail.com; ORCID ID: *https://orcid.org/0000-0001-9974-3215;*

Kuchin Yan – PhD student, senior researcher of the Institute of Information and Computational Technologies of the National Academy of Sciences of the Republic of Kazakhstan; ykuchin@mail.ru; ORCID ID: *https://orcid.org/0000-0002-5271-9071;*

Yunicheva Nadiya – PhD, leading researcher of the Institute of High Technologies of the National Academy of Sciences of the Republic of Kazakhstan; naduni@mail.ru; ORCID ID: *https://orcid.org/0000-0001-6351-3450;*

Yelis Marina – PhD student, junior researcher of the Institute of Information and Computational Technologies of the National Academy of Sciences of the Republic of Kazakhstan; k.marina92@gmail.com; ORCID ID: *https://orcid.org/0000-0003-4203-800X;*

Mukhamedieva Elena – is a researcher at the Institute of Information and Computational Technologies of the Ministry of Education and Science of the Republic of Kazakhstan, muhamedijeva@gmail.com, ORCID ID: *https://orcid.org/0000-0001-9596-4432.*

ЛИТЕРАТУРА:

Атун Р. (2015) Переход систем здравоохранения на мультиморбидность. Ланцет 386: 721-722.

Орлов Е.М. (2010) Категория эффективности в системе здравоохранения. Фундаментальные исследования 4:70-75. (На русском).

Болл-Рокич С.Дж., Дефлер М. (1976) Модель зависимости эффектов средств массовой информации. Коммуникационные исследования, 3(1): 3–21. https://doi.org/10.1177/009365027600300101.

«ДЕЛЮКС Типография». (2021) КАЗАХСТАН И COVID-19: СМИ, КУЛЬТУРА, ПОЛИТИКА, http://library.fes.de/pdf-files/bueros/kasachstan/18218.pdf.

Гири С.П., Маурья А.К. (2021) Забытая реальность средств массовой информации во время COVID-19: влияние новостей о пандемии на положительные и отрицательные

эмоции и психологическую устойчивость человека. Личность и индивидуальные различия, 180:110962.

Аслам Ф., Аван Т.М., Сайед Дж.Х., Кашиф А., Парвин М. (2020) Чувства и эмоции, вызванные заголовками новостей о вспышке коронавирусной болезни (COVID-19). Коммуникации гуманитарных и социальных наук, 7: 1-9.

Хамидин З., Хатам Дж., Резапур Т. (2020) Как люди эмоционально реагируют на новости о COVID-19: онлайн-опрос. Базовая и клиническая неврология, 11:171.

Кемп С. (2020) Digital 2020: Казахстан. www.datareportal.com. Дата обращения:16.10.2020.

Мухамедиев Р.И., Сымагулов А., Кучин Ю., Якунин К., Елис М. (2021) От классического машинного обучения к глубоким нейронным сетям: упрощенный наукометрический обзор. Прикладные науки, 11(12):5541, https://doi.org/10.3390/app11125541.

Машечкин И.В., Петровский М.И., Царев Д.В. (2013) Методы расчета релевантности текстовых фрагментов на основе тематических моделей в задаче автоматического аннотирования. Вычислительные методы и программирование, 14(1):91-102.

Воронцов К.В., Потапенко А.А. (2012) Регуляризация, надежность и разреженность вероятностных тематических моделей. Компьютерные исследования и моделирование, 4(4):693-706.

Пархоменко П.А., Григорьев А.А., Астраханцев Н.А. (2017) Обзор и экспериментальное сравнение методов кластеризации текста. Труды ИСП РАН, 29(2):161-200. DOI: 10.15514/ИСПР РАН-2017-29(2)-6.

Диенг А.Б., Руис Ф.Дж., Блей Д.М. (2020) Тематическое моделирование в пространстве вложений. Труды Ассоциации компьютерной лингвистики, 8:439-453.

Блей Д., Нг Э.Й., Джордан М. (2003) Скрытое распределение Дирихле. Журнал исследований машинного обучения, 3(1):993-1022.

Воронцов К. и соавт. (2015) Bigartm: библиотека с открытым исходным кодом для регуляризованного мультимодального тематического моделирования больших коллекций. Международная конференция по анализу изображений, социальных сетей и текстов, Springer, Cham. С. 370-381.

Мухамедиев Р.И., Якунин К., Мусабаев Р., Булдыбаев Т., Кучин Ю., Мурзахметов С., Елис М. (2020) Классификация негативной информации на общественно значимые темы в СМИ. Симметрия, 12(12):1945.

Якунин К., Калимолдаев М., Мухамедиев Р.И., Мусабаев Р., Барахнин В., Кучин Ю., Мурзахметов С., Булдыбаев Т., Оспанова Ю., Елис М. (2021) KazNewsDataset: Единый национальный корпус цифровых СМИ, 6:31. https://doi.org/10.3390/data6030031.

Джелодар Х. (2018) Скрытое распределение Дирихле (LDA) и тематическое моделирование: модели, приложения, обзор. Мультимедийные инструменты и приложения. 78(5): 15169–15211. https://doi.org/10.1007/s11042-018-6894-4.

Воронцов К., Фрей О., Апишев М., Ромов П., Дударенко М. (2015) BigARTM: библиотека с открытым исходным кодом для регуляризованного мультимодального тематического моделирования больших коллекций. На Международной конференции по анализу изображений, Soc. Сети и тексты, Springer: Cham, Швейцария. стр. 370-381. https://doi.org/10.1007/978-3-319-26123-2_36.

Маруто Д., Хандака С.Х., Виджая Э. (2018) Определение числа кластеров по

k-среднему с использованием метода локтя и оценка чистоты в заголовках новостей. Международный семинар по применению информационных и коммуникационных технологий 2018 г., стр. 533–538, https://doi.org/10.1109/ISEMANTIC.2018.8549751.

Якунин К., Мухамедиев Р.И., Зайцева Е., Левашенко В., Елис М., Сымагулов А., Кучин Ю., Мухамедиева Е., Аубакиров М., Гопеженко В. (2021) СМИ как зеркало пандемии COVID-19. Вычисления. 9(12):140. https://doi.org/10.3390/computation9120140.

REFERENCES:

Atun R. (2015) Transitioning health systems for multimorbidity. The Lancet 386:721-722.

Orlov E.M. (2010) The category of effectiveness in the health care system. Basic research 4:70-75. (In Russian).

Ball-Rokeach S.J., DeFleur M.L. (1976) A dependency model of mass-media effects. Communication Research, 3(1): 3–21. https://doi.org/10.1177/009365027600300101.

«DELUXE Printery». (2021) KAZAKHSTAN AND COVID-19: MEDIA, CULTURE, POLITICS, http://library.fes.de/pdf-files/bueros/kasachstan/18218.pdf.

Giri S.P., Maurya A.K. (2021) A neglected reality of mass media during COVID-19: Effect of pandemic news on individual's positive and negative emotion and psychological resilience. Personality and Individual Differences, 180:110962.

Aslam F., Awan T.M., Syed J.H., Kashif A., Parveen M. (2020) Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. Humanities and Social Sciences Communications, 7:1-9.

Hamidein Z., Hatam J., Rezapour T. (2020) How people emotionally respond to the news on COVID-19: An online survey. Basic and Clinical Neuroscience, 11:171.

Kemp S. (2020) Digital 2020: Kazakhstan. www.datareportal.com. Date of access:16.10.2020.

Mukhamediev R.I., Symagulov A., Kuchin Y., Yakunin K., Yelis M. (2021) From Classical Machine Learning to Deep Neural Networks: A Simplified Scientometric Review. Applied Sciences, 11(12):5541, https://doi.org/10.3390/app11125541.

Mashechkin I.V., Petrovsky M.I., Tsarev D.V. (2013) Methods for calculating the relevance of text fragments based on thematic models in the problem of automatic annotation. Computational methods and programming, 14(1):91-102.

Vorontsov K.V., Potapenko A.A. (2012) Regularization, robustness and sparseness of probabilistic thematic models. Computer research and modeling, 4(4):693-706.

Parkhomenko P.A., Grigoriev A.A., Astrakhantsev N.A. (2017) Review and experimental comparison of text clustering methods. Proceedings of ISP RAS, 29(2):161-200. DOI: 10.15514/ISPRAS-2017-29(2)-6.

Dieng A.B., Ruiz F.J., Blei D.M. (2020) Topic modeling in embedding spaces. Transactions of the Association for Computational Linguistics, 8:439-453.

Blei D.M., Ng A.Y., Jordan M.I. (2003) Latent dirichlet allocation. Journal of machine Learning research, 3(1):993-1022.

Vorontsov K. et al. (2015) Bigartm: Open source library for regularized multimodal topic modeling of large collections. International Conference on Analysis of Images, Social Networks and Texts, Springer, Cham. P. 370-381.

Mukhamediev R.I., Yakunin K., Mussabayev R., Buldybayev T., Kuchin Y., Murzakhmetov S., Yelis M. (2020) Classification of Negative Information on Socially Significant Topics in Mass Media. Symmetry, 12(12):1945.

Yakunin K., Kalimoldayev M., Mukhamediev R.I., Mussabayev R., Barakhnin V., Kuchin Y., Murzakhmetov S., Buldybayev T., Ospanova U., Yelis M. (2021) KazNewsDataset: Single Country Overall Digital Mass Media Publication Corpus, 6:31. https://doi.org/10.3390/data6030031.

Jelodar H. (2018) Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. Multimedia Tools and Applications. 78(5): 15169–15211. https://doi.org/10.1007/s11042-018-6894-4.

Vorontsov K., Frei O., Apishev M., Romov P., Dudarenko M. (2015) BigARTM: Open Source Library for Regularized Multimodal Topic Modeling of Large Collections. In International Conference on Analysis of Images, Soc. Networks and Texts, Springer: Cham, Switzerland. pp.370-381. https://doi.org/10.1007/978-3-319-26123-2_36.

Marutho D., Handaka S.H., Wijaya E. (2018) The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. "2018 International Seminar on Application for Technology of Information and Communication. pp. 533-538, https://doi.org/10.1109/ISEMANTIC.2018.8549751.

Yakunin K., Mukhamediev R.I., Zaitseva E., Levashenko V., Yelis M., Symagulov A., Kuchin Y., Muhamedijeva E., Aubakirov M., Gopejenko V. (2021) Mass Media as a Mirror of the COVID-19 Pandemic. Computation. 9(12):140. https://doi.org/10.3390/computation9120140.

МАЗМҰНЫ

СОДЕРЖАНИЕ

CONTENTS

Publication Ethics and Publication Malpracticein
the journals of the National Academy of Sciences of the Republic of Kazakhstan

For information on Ethics in publishing and Ethicalguidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier.com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any otherlanguage, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:
www:nauka-nanrk.kz
http://physics-mathematics.kz/index.php/en/archive
ISSN2518-1726 (Online),
ISSN 1991-346X (Print)

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*
Заместитель директор отдела издания научных журналов НАН РК *Р. Жәлиқызы*
Редакторы: *М.С. Ахметова, Д.С. Аленов*
Верстка на компьютере *Г.Д.Жадыранова*

Подписано в печать 15.09.2022.
Формат 60х881/8. Бумага офсетная. Печать – ризограф.
17,5 п.л. Тираж 300. Заказ 3.

Национальная академия наук РК
050010, Алматы, ул. Шевченко, 28, т. 272-13-19