REDS:

Телекоммуникационные устройства и системы

№2

2023

СОДЕРЖАНИЕ

ИМИТАЦИОННАЯ МОДЕЛЬ SP-СЕТИ

Абызов Артемий Алексеевич,
Российский университет транспорта, студент, Москва, Россия
ascanders@gmail.com

Архипов Алексей Михайлович,
Российский университет транспорта, студент, Москва, Россия
rf.arhipov@yandex.ru

Басюк Степан Александрович,
Российский университет транспорта, студент, Москва, Россия
kenshi.s@mail.ru

Михалевич Игорь Феодосьевич,
Российский университет транспорта, доцент, к.т.н., старший научный сотрудник, Москва, Россия
mif-orel@mail.ru

Тихомиров Владимир Алексеевич,
Российский университет транспорта, студент, Москва, Россия
dphamineless@yandex.ru

Аннотация

В статье представлена реализация модели, имитирующей типовые процессы в криптографических системах, основанных на блоках замен и перестановок. Обращено внимание, что понимание функционирования таких систем их корректное применение требует знаний в широкой области математических наук и умения их интерпретировать. Визуализация всех вычислений облегчает понимание сложных теоретических вопросов. Модель реализована в виде кроссплатформенного приложения, имитирующего шифр AES. Описаны решения, применимые для моделирования других криптографических систем.

Ключевые слова: *Блок замен, блок перестановок, информационная безопасность, криптографическая система, среда разработки программы, AES, S-box, P-box*

Введение

Ускоренная пандемией COVID-19 цифровая трансформация экономики и всего информационного общества повысила внимание к вопросам цифровой гигиены, стимулировала совершенствование средств защиты информации и их масштабное распространение [1, 2]. В период удаленной работы обострились многие проблемы обеспечения безопасности информации, связанные, в том числе, с недостаточностью квалифицированных кадров и недостатками в их подготовке. Трудности в подготовке кадров по информационной безопасности возникают, в частности, по причине сложности материала, основанного на высшей математике, особенно в области криптографии. Подача такого материала исключительно в теоретической форме может отпугивать обучаемых и снижать их самооценку, что в конечном счете приводит к снижению интереса к учебной дисциплине с очевидными негативными последствиями.

Устранить эти трудности, не прибегая к упрощению материала, помогает его визуализация, что, в рассматриваемом случае, достигается путем имитационного моделирования.

Выбор моделируемой SP-сети

Спецификации многих криптографических протоколов содержат алгоритмы и методы, основанные на блоках замен (S-box) и перестановок (P-box), из которых создаются SP-сети. Понимание функционирования таких сетей и их корректное применение требует знаний в широкой области, включая теорию матриц, модулярной арифметики, абстрактной алгебры (групп, колец, полей), математической статистики, принятия решений и других, а также умения эти знания интерпретировать в интересах решаемых задач обеспечения информационной безопасности защищаемых систем.

Для имитационного моделирования SP-сети был выбран шифр AES, как наиболее распространенный и применяемый, в том числе, государственными органами Российской Федерации. Это следует, например, из спецификации криптографического протокола:

(TLS_ECDHE_RSA_WITH_AES_128_GSM_SHA256, длина ключей 128 бит, TLS 1.2), -

используемого официальным сайтом МВД России (рис. 1).

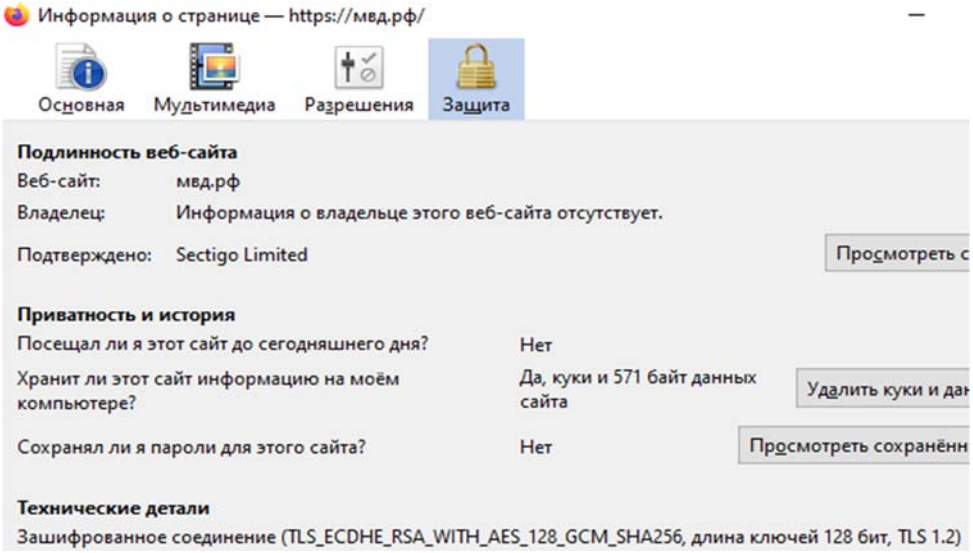

Рис. 1. Технические детали веб-сайта МВД России (браузер Mozilla)

Описание моделируемой SP-сети

Шифр AES [3,4] представляет собой SP-сеть, при работе которой входное сообщение разбивается на блоки размером 128 бит, подвергаемые раундовым преобразованиям. Раунды, за исключением первого и последнего, исполняют одинаковые процедуры, представленные на рисунке 1.

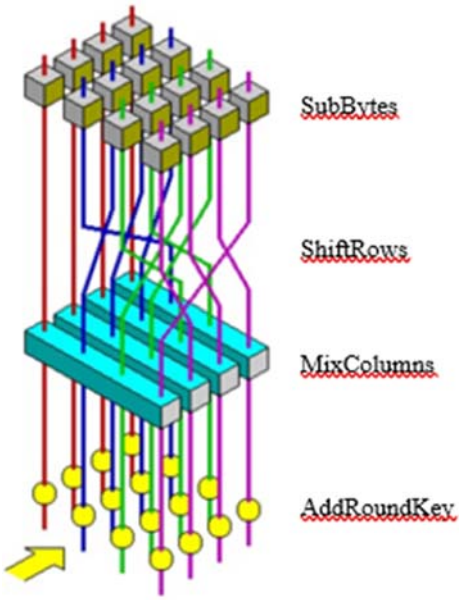

Рис. 2. Раунд AES (Источник: https://yandex.ru/search/?text=AES&clid=2411726&lr=213)

Такими процедурами являются:

SubBytes - замена байтов;

Shift Rows - сдвиг строк;

MixColumns – смешивание столбцов;

AddRoundKey - добавление раундовых ключей, которые вырабатываются путем расширения базового (мастер) ключа процедурой KeyExpansion.

Работа в раундах производится над блоками полного размера. В первом и последнем раундах выполняется только процедура добавления ключей. В зависимости от размера базового ключа меняется число раундов. В остальном работа сети остаётся неизменной.

Основные характеристики шифра AES приведены в таблице 1.

Таблица 1

Характеристики SP-сетей AES

Характеристики	Шифр		
Название шифра	AES-128	AES-192	AES-256
Размер ключа, бит	128	192	256
Число раундов	10	12	14
Размер блока, бит	128		

На стороне отправителя и получателя SP-сеть AES работает одинаково. Полученное на выходе у отправителя зашифрованное сообщение, поступив на вход SP-сети получателя, преобразуется в исходное сообщение отправителя. Подлинность и конфиденциальность сообщения обеспечиваются за счет специальных криптографических свойств блоков замен и их многократного применения совместно с блоками, реализующими процедуры перестановок [5].

Каждый раунд оперирует целым блоком данных размером 128 бит. Раундовые процедуры выполняются над отдельными байтами, которые интерпретируются как элементы поля Галуа $GF(2^8)$, или над словами, образованными четырьмя байтами.

В поле $GF(2^8)$ определены операции сложения и умножения двух элементов, результатом которых является элемент этого же поля.

Сложение выполняется с помощью операции XOR поразрядно. То есть результатом сложения байтов

$P = \{p_7, p_6, p_5, p_4, p_3, p_2, p_1, p_0\}$

и

$Q = \{q_7, q_6, q_5, q_4, q_3, q_2, q_1, q_0\}$

будет

$R = \{r_7, r_6, r_5, r_4, r_3, r_2, r_1, r_0\}$,

где $r_i = p_i$ XOR q_i.

В операции умножения байты представляются в виде полиномов седьмой степени:

$p(\boldsymbol{x}) = p_7 \cdot x^7 + p_6 \cdot x^6 + \ldots + p_2 \cdot x^2 + p_1 \cdot x^1 + p_0 \cdot x^0$

и

$q(\boldsymbol{x}) = q_7 \cdot x^7 + q_6 \cdot x^6 + \ldots + q_2 \cdot x^2 + q_1 \cdot x^1 + q_0 \cdot x^0$**.**

Результат умножения вычисляется по модулю неприводимого полинома

$m(x) = x^8 + x^4 + x^3 + x + 1$.

Таким образом, в результатом умножения байтов P и Q в поле $GF(2^8)$ вычисляется байт, который является остатком от деления на многочлен $m(x)$ произведения $p(x) \cdot q(x)$:

$r(x) = p(x) \cdot q(x) \bmod m(\mathrm{x}) =$

$(p_7 \cdot x^7 + p_6 \cdot x^6 + \ldots + p_2 \cdot x^2 + p_1 \cdot x^1 + p_0) \cdot$

$(q_7 \cdot x^7 + q_6 \cdot x^6 + \ldots + q_2 \cdot x^2 + q_1 \cdot x^1 + q_0) \bmod (x^8 + x^4 + x^3 + x + 1)$.

Для выполнения раундов входные данные разбиваются на блоки по 128 бит (16 байт). При недостаточном объеме для формирования полного последнего блока входные данные дополняются.

Блоки представляются в виде матриц состояний (*state*) размером 4 x 4.

Замена байтов (SubBytes) осуществляется по заранее созданным таблицам (S-box).

Процедура ShiftRows представляет собой циклический сдвиг строк матриц состояний state.

При выполнении процедуры MixColumns происходит умножение каждого столбца матрицы state на фиксированную матрицу. Это является линейным преобразованием столбцов матрицы state.

В ходе процедуры AddRoundKey раундовый ключ с помощью поразрядного XOR добавляется в матрицу state.

Процедура KeyExpansion обеспечивает выработку 44 четырехбайтовых слов (в сети AES-128): 4 слова на основной ключ и по 4 слова ключей на каждый из 10 раундов. В этом случае полная длина расширенного ключа составляет 1408 бит.

Описание имитационной модели

Имитационная модель реализована в виде кроссплатформенного приложения. При ее работе обеспечивается визуализация всех процедур, которая облегчает восприятие описанного выше сложного математического аппарата и дополнительно позволяет поупражняться с вычислениями, в том числе в поле Галуа.

Для реализации имитационной модели с расширенной возможностью визуализации вычислительного процесса была использована среда разработки IDE, язык программирования JAVA и его библиотеки [6].

В процессе разработки были написаны классы для различных целей, а также методы для реализации их функционала. Использованные классы и методы представлены на рисунке 3.

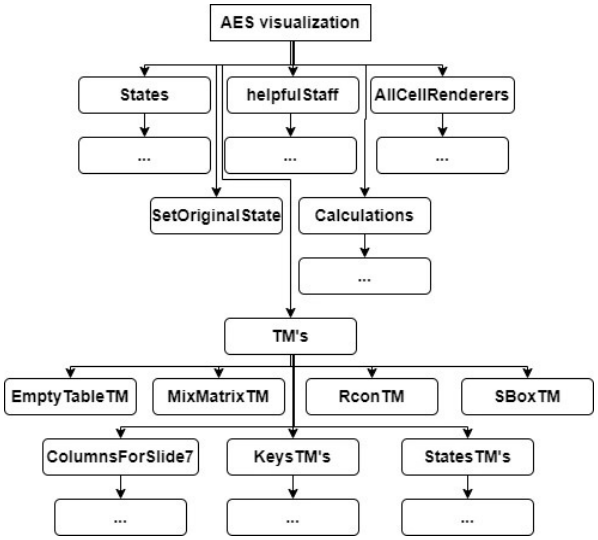

Рис. 3. Классы и методы, примененные для визуализации процедур SP-сети

Основными библиотеками для создания графического интерфейса стали Swing и AWT [7-9]. Графическая библиотека Swing была создана на основе библиотеки AWT для решения таких проблем AWT, как недостаточный выбор графических компонентов и зависимость внешнего вида и поведения AWT графического интерфейса пользователя от конкретной операционной системы. Библиотека AWT использует нативный код (т.е. код, специфичный для конкретной операционной системы) для отображения компонентов. Компоненты Swing поддерживают специфические динамически подключаемые виды и поведения, благодаря которым возможна адаптация к графическому интерфейсу платформы (то есть к компоненту можно динамически подключить другой, специфический для операционной системы, в том числе и созданный программистом вид и поведение). Сочетание этих двух библиотек позволило в полной мере добиться поставленной цели визуализировать все процедуры, выполняемые в SP-сети.

В частности, были написаны следующие классы.

AESEncrypter – класс, отвечающий за шифрование/расшифрование В нем реализованы все раунды SP- сети AES-128

Key – класс, облегчающий вывод значения ключа при тестировании и может генерировать случайный ключ.

KeyExpander – этот класс расширяет исходный 128 битный ключ в раундовые ключи.

Разработка Swing GUI в IDE

GUI Builder в среде IDE разрешает основные проблемы, возникающие при создании графического интерфейса Java путем рационализации процесса создания графических интерфейсов, освобождая разработчиков от необходимости изучения особенностей диспетчеров компоновки Swing. Это выполняется путем расширения возможностей конструктора графического интерфейса пользователя для поддержки простой парадигмы "Произвольная структура" с простыми правилами компоновки, понятными и простыми в использовании. В процессе проектирования формы GUI Builder предоставляет визуальные средства поддержки, предлагая расположение и выравнивание компонентов. GUI Builder

способствует переносу пользовательских решений по разработке в функциональный пользовательский интерфейс, реализуемый при помощи диспетчера компоновки GroupLayout и других средств Swing. Благодаря динамической модели размещения компонентов поведение графического интерфейса в GUI Builder во время выполнения соответствует ожидаемому, что позволяет вносить корректировки без изменения установленных взаимосвязей между компонентами. При каждом изменении размеров форм, переключении локалей или применении нового общего стиля графический интерфейс автоматически изменяется в соответствии с новой настройкой вставок и смещений стиля.

Работа имитационной модели обеспечивает создание множества окон со своими вставками визуализации, для чего необходимо произвести первичную инициализацию и создать контейнер. Это происходит следующим образом.

Функционирование имитационной модели

При первом запуске модели папка с исходными файлами в окне "Проекты" содержит пустой узел <default package>. Для продолжения процесса создания интерфейса необходимо создать контейнер Java, в который будут помещены другие требуемые элементы графического интерфейса. В этом действии будет выполнено создание контейнера с использованием компонента JFrame и размещение контейнера в новом пакете.

Добавление контейнера JFrame производится следующим образом.

В окне 'Проекты' щелкнуть правой кнопкой мыши на узел ContactEditor, выберать 'Создать' > 'Форма JFrame'. Также форму JFrame можно найти, выбрав "Создать" > "Другое" > "Формы графического интерфейса Swing" > "Форма JFrame".

Далее необходимо:

- ввести ContactEditorUI в поле имени класса;
- ввести my.numberaddition в поле пакета;
- нажать на окно "Завершить".

Среда IDE создаст форму ContactEditorUI и класс ContactEditorUI в проекте ContactEditorUI.java и откроет форму ContactEditorUI в средстве GUI Builder.

Приветствие «Введите данные» ожидает ввода сообщения и ключа (рисунок 4). При некорректном вводе будет выведено сообщение об ошибке. Прервать моделирование можно на любом этапе нажатием кнопки «Отмена».

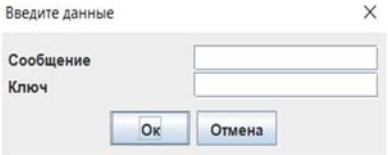

Рис. 4. Окно ввода исходных данных

После ввода данных открывается заглавная страница модели (рисунок 5). На ней отображается вид SP-сети и введенные исходные данные.

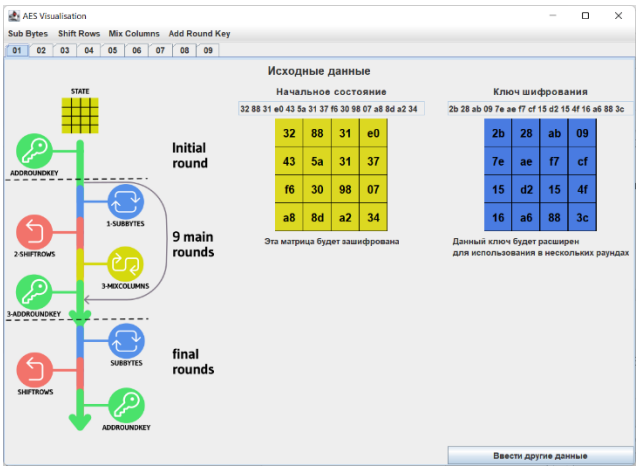

Рис. 5. Заглавная страница имитационной модели

Для просмотра каждого этапа предусмотрены кнопки в верхней части панели. Для изменения исходных данных используется кнопка «Ввести другие данные». Продолжения моделирования вызывает нажатие «Enter» на клавиатуре компьютера. При каждом последующем нажатии «Enter» выполняется последовательный переход от одной раундовой процедуры к другой: Sub bytes - Shift Rows - Mix Columns – Add round key. При нескольких нажатиях «Enter» детально демонстрируется выполнение каждой процедуры и выполняемые вычисления. Это же правило соблюдается при моделировании остальных процедур.

На рисунке 6 отображено выполнение процедуры Sub bytes.

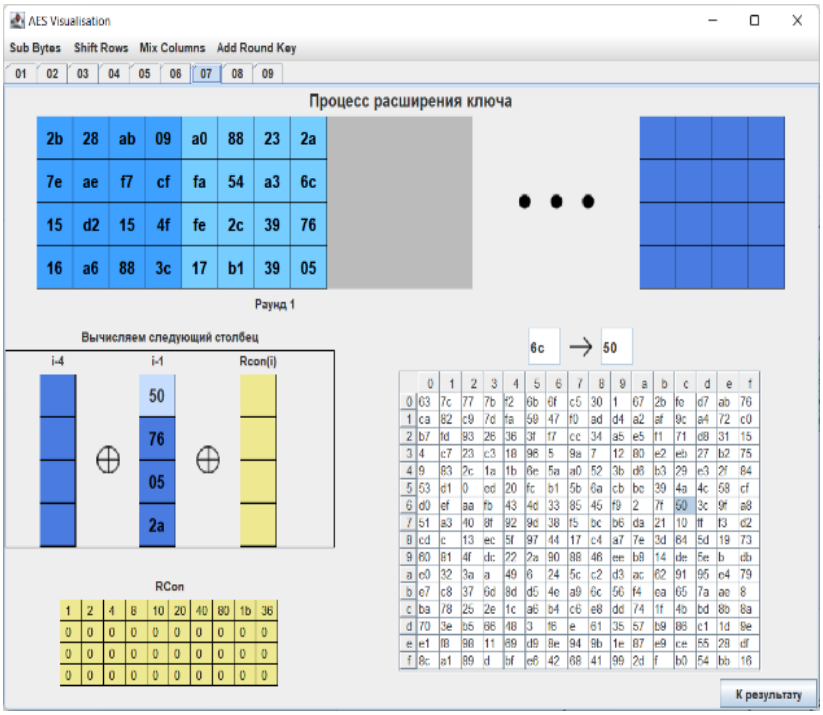

Рис. 6. Визуализация процедуры Sub bytes

Выполнение процедуры Mix Columns отображает рисунок 6.

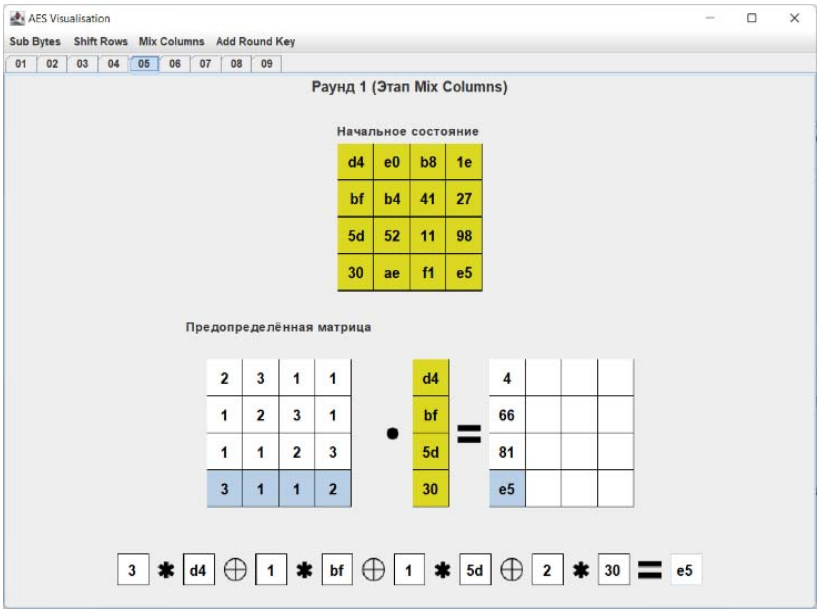

Рис. 6. Визуализация процедуры Mix Columns

Рисунок 7 представляет выполнение процедуры расширения ключа. Еще раз напомним, что при каждом нажатии клавиши «Enter» отображаются выполняемые вычисления.

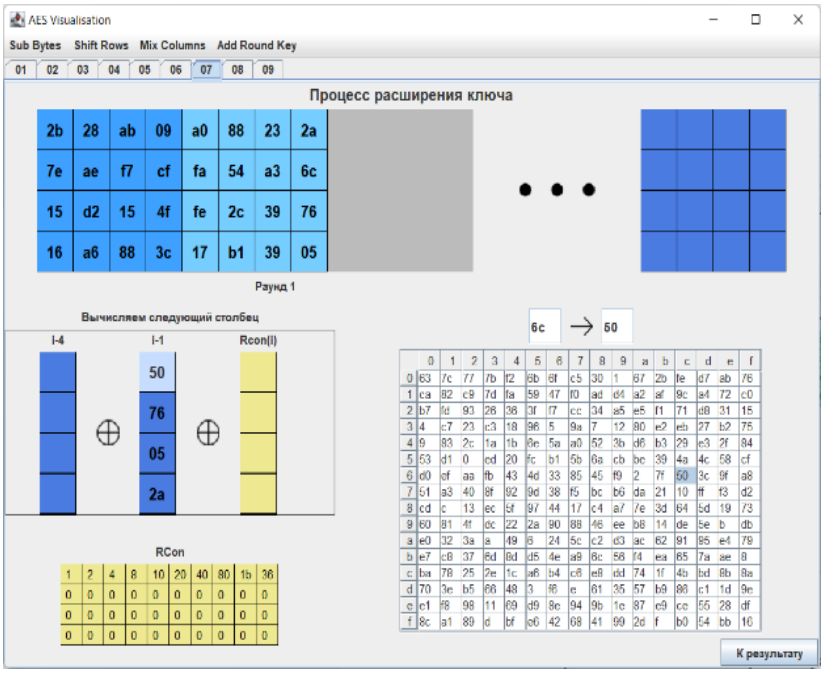

Рис. 7. Детализация процедуры Add round key

Промежуточные и итоговые результаты моделирования отображаются в виде, представленном на рисунке 8.

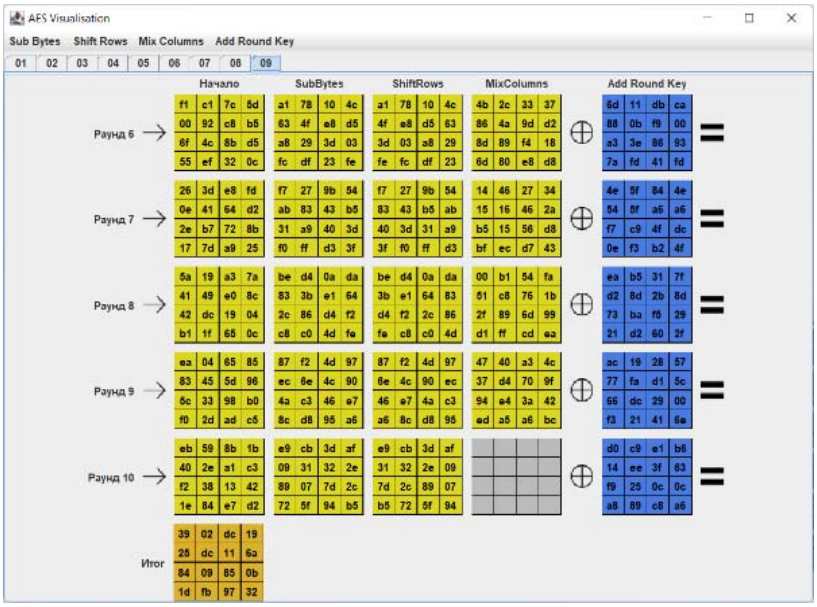

Рис. 8. Визуализация результатов выполнения раундов и итогового результата моделирования

Решения, применимые для моделирования других криптографических систем

При разработке модели возникали проблемы реализации ее графической части. Изначально было решено вести работу в среде разработки Intellij Idea, но инструменты построения формы были неудобны, так как все объекты на форме строго привязаны к сетке. В дальнейшем была рассмотрена среда разработки Apache Netbeans. Она позволяет размещать объекты в любом месте окна без привязки к сетке. Но большая часть модели уже была написана с помощью инструментов изначально выбранной среды разработки, поэтому было решено не переносить проект.

В среду разработки Apache Netbeans интегрирована NetBeans IDE GUI Builder и NetBeans Platform

NetBeans IDE GUI Builder формально известный как проект Matisse инструмент для визуальной разработки интерфейса, позволяющий разработчикам проектировать и составлять интерфейсы к

программам посредством перетаскивания элементов в рабочую область.

NetBeans Platform — платформа для разработки модульных настольных Swing-приложений. NetBeans IDE содержит все, что нужно для разработки плагинов и приложений на основе NetBeans Platform. Приложения могут динамически загружать другие модули. Любое приложение может включить модуль обновления, чтобы позволить пользователям загружать обновления для программ и модулей в работающее приложение.

NetBeans Platform предлагает многократно используемые сервисы и модули для настольных приложений, позволяя разработчикам сфокусироваться на логике приложения. Особенности платформы:

- управление дизайном приложения (меню, всплывающие окна);
- управление настройками пользователя;
- управление хранением данных;
- управление окнами;
- фреймворк для разработки пошаговых мастеров установки;
- NetBeans Visual Library – библиотека визуальных элементов;
- Integrated Development Tools – встроенные инструменты разработки.

В процессе работы для улучшения визуального вида интерфейса модели было принято решение на заглавную страницу интегрировать картинку, демонстрирующею пошаговый процесс шифрования/расшифрования с типом расширения SVG. Основная проблема заключалась в отсутствии встроенной поддержки Java для изображений формата SVG. Для использования таких изображений нужно добавить в проект сторонние библиотеки. Из множества библиотек: Batik, SVGSalamander, Flatlaf, – Flatlaf показался наиболее удобным вариантом, поэтому дальнейшая работа была основана на использовании этой библиотеки. В частности, он предназначен для упрощения интеграции SVG в Java-игры и облегчения для художников разработки 2D-контента игр: от богатых интерактивных меню до диаграмм и графики до сложных анимаций.

Одной из проблем стало создание таблицы S-box с названиями строк и столбцов. Компонент JTable по умолчанию имеет заголовки только для столбцов. Решением стало помещение таблицы в компонент JScrollPane, который позволяет задавать заголовки строк. Чтобы заголовки строк и столбцов были одинаковы, использовался из открытого доступа готовый класс TableWithRowHeader.

Последующей проблемой стало невозможность демонстрации только определенных столбцов. Решением стало использование из открытого доступа уже готового класса TableColumnManager. TableColumnManager обрабатывает запросы на скрытие и отображение столбцов, а затем вызывает соответствующий метод TableColumnModel. Это позволило отображать во время выполнения модели только нужные на данный момент столбцы больших таблиц.

Сложным вопросом была реализация отката изменений таблиц, произведённых во время проигрывания анимаций. Решением стало создание класса AllStatesCopy, в котором хранятся копии исходных значений таблиц. Их можно использовать при сбрасывании анимации слайдов. Во время выполнения вычислений в случае необходимости результат этапа записывается не только в класс AllStates, но и в класс AllStatesCopy. Модели таблиц всегда берут значения из одного источника, который изменяется при проигрывании анимаций, когда таблицу на слайде нужно вернуть к исходному состоянию, в этот источник присваивается значение, сохраненное в классе AllStatesCopy.

Заключение

Представленная имитационная модель SP-сети предоставляет возможность визуально наблюдать выполнение всех процессов и вычислений. Это облегчает восприятие сложного математического аппарата, используемого при создании криптографических систем, и дополнительно позволяет поупражняться с вычислениями, в том числе в поле Галуа.

Использованные при разработке модели решения могут быть полезны при моделировании других криптографических систем и помочь избежать трудностей, с которыми столкнулись авторы.

Имитационная модель реализована в виде кроссплатформенного приложения, нечувствительного к вычислительным средам. Это было направлено на создание условий для использования модели в исследовательских и учебных целях, независимо от состояния материально-технической базы.

Литература

1. *Михалевич И.Ф. Михалевич И.Ф*. Цифровая гигиена информационного общества: влияние пандемии COVID-19 // REDS: Телекоммуникационные устройства и системы, № 3-2022. С. 10-17.

2. *Михалевич И.Ф*. Цифровая трансформация систем управления в условиях пандемии COVID-19 // REDS: Телекоммуникационные устройства и системы, № 4-2021. С. 26-32. (доступ 05.09.2021).

3. Announcing the Advanced Encryption Standard (AES). https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.197.pdf (доступ 11.01.2023).

4. *Joan Daemen, Vincent Rijmen*. AES Proposal: Rijndael. Note on naming. https://csrc.nist.gov/csrc/media/projects/cryptographic-standards-and guidelines/documents/aes-development/rijndael-ammended.pdf#page=1 (доступ 11.01.2023).

5. *Andrey Bogdanov, Dmitry Khovratovich, Christian Rechberger*. Biclique Cryptanalysis of the Full AES. https://www.researchgate.net/publication/221326929_Biclique_Cryptanalysis_of_the_Full_AES (доступ 11.01.2023).

6. Регулярные выражения JAVA. Сайт о программировании https://metanit.com/java/tutorial/7.4.php (доступ 07.10.2022).

7. Java Swing – пример showMessageDialog из JOptionPane. Сайт о программировании https://csharpcoderr.com/5396/ (доступ 07.10.2022).

8. JTable, TableModel, TableCellEditor. Все о JAVA. https://javaonline.ru/swing-jtable.xhtml (доступ 07.10.2022).

9. Пакет java.awt. https://docs.oracle.com/en/java/javase/17/docs/api/java.desktop/java/awt/package-summary.html (доступ 07.10.2022).

ИССЛЕДОВАНИЕ ПОМЕХОУСТОЙЧИВОСТИ СИСТЕМЫ MIMO С ВЫСОКИМИ ПОРЯДКАМИ МОДУЛЯЦИИ

Бен Режеб Софиэн Бен Камель,
Московский Технический Университет Связи и Информатики, студент, Москва, Россия
sbenrezheb@yandex.ru

Аннотация
Анализируется работа модели однопользовательской системы беспроводной связи MIMO при высоких порядках модуляции. Рассмотрены приемники в системе MIMO. Так же представлен график помехоустойчивости 4 видов модуляции, а именно QAM-64, QAM-256, QAM-1024, QAM-4096. По данному графику представлены результаты, так же описаны лучшие и худшие стороны разных видов модуляции.

Ключевые слова: *Радиосвязь, однопользовательская система беспроводной связи, Multiple Input Multiple Output, Zero Forcing Detection, Minimum Mean Squared Error Detection, модуляция QAM-64, QAM-256, QAM-1024, QAM-4096.*

Введение

Усложнение радиоэлектронной, электромагнитной и помеховой обстановки в диапазонах частот сотовой связи ставит новые задачи перед системой радиоконтроля. Количество абонентов и количество передаваемой информации постоянно растет, и для это требуется совершенствование стандартов систем беспроводной связи. Совершенствование может достигаться путем расширения полосы частот канала, либо увеличения отношения сигнал/шум, но так как частотные ресурсы небезграничны и, в основном, уже все выделены, то использовать вариант расширения полосы частот канала без пересечения с другими системами связи, использующие смежные частотные каналы, становится невозможным [1].

Альтернативный способ увеличения показателя скорости передачи информации - применение в системах беспроводной связи технологии MIMO (Multiple Input Multiple Output). Данная технология состоит из нескольких антенн на передающей стороне и нескольких антенн на приемной стороне, что отлично от систем связи, использующие технологию SISO (Single Input Single Output), где одна антенна используется, и на приемной, и на передающей стороне. Использование системы MIMO позволяет улучшить показатели скорости обмена информации путем пространственного мультиплексирования, так же за счет пространственного разнесения повышается помехоустойчивость системы.

Система беспроводной связи MIMO наделена несколькими режимами работы:

- пространственное мультиплексирование, формирование луча,
- пространственно-временное кодирование, прекодирование,
- множественный доступ с пространственным разделением.

В соответствии с выбранным режимом работы беспроводной системы связи могут быть достигнуты такие преимущества, как:

- Повышение помехоустойчивости системы;
- Повышение скорости передачи информации в системе связи [3].

Использование технологии MIMO [14-26] является одной из ключевых стандартов сотовой связи LTE и LTE-Advanced.

Постановка задачи

В данном докладе рассмотрена система связи с несколькими передающими и несколькими приемным антеннами, а именно, система MIMO с высокими порядками модуляции.

Целью статьи является анализ спектральной эффективности системы и исследование её на помехоустойчивость. Исследования проводятся при больших порядках модуляции, а именно будут рассмотрены варианты при QAM-64, QAM-256, QAM-1024, QAM-4096.

Модель однопользовательской системы беспроводной связи

Модель системы MIMO изображена на Рис.1. Мы представляем систему связи с передающими антеннами N_T и приемными антеннами N_R .

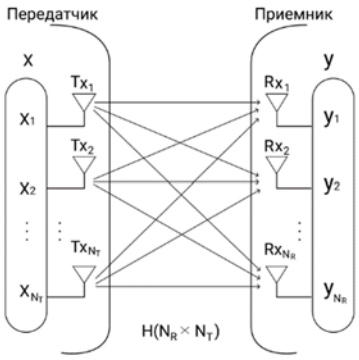

Рис. 1. Модель системы SU-MIMO

Антенны $Tx_1,...,Tx_{N_T}$ посылают сигналы $x_1,...,x_{N_T}$ на приемные антенны $Rx_1,...,Rx_{N_R}$. Каждая приемная антенна объединяет входящие сигналы, которые последовательно суммируются. Принятые сигналы на антеннах $Rx_1,...,Rx_{N_R}$ обозначаются $y_1,...,y_{N_R}$. Выражаем принятый сигнал на антенне $Tx_q;q=1,...,N_R$ в виде:

$$Y_q=\sum_{p=1}^{N_T} h_{qp}\cdot x_p+b_q;q=1,...,N_R \quad (1)$$

Модель канала MIMO с замираниями описывается следующим образом:

$$\mathbf{y}=\mathbf{H}\bullet\mathbf{s}+\mathbf{n} \quad (2)$$

$\mathbf{H}$ — это (NR×NT) комплексная матрица канала размерности $(N_R \times N_T)$, заданная:

$$\mathbf{H}=\begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1N_T} \\ h_{21} & h_{22} & & h_{2N_T} \\ \vdots & & \ddots & \vdots \\ h_{N_R1} & h_{N_R2} & \cdots & h_{N_RN_T} \end{pmatrix},$$

где $h_{qp};p=1,...,N_T;q=1,...,N_R$, — коэффициент усиления комплексного канала, который связывает передающую антенну Tx_p с приемной антенной Rx_q.

- $s=[x_1,...,x_{N_T}]^T$ представляет собой комплексный вектор передаваемого сигнала размерности $(N_T\times1)$.
- $y=[y_1,...,y_{N_R}]^T$ представляет собой комплексный вектор принятого сигнала размерности $(N_R\times1)$.
- $n=[b_1,...,b_{N_R}]^T$ представляет собой вектор комплексного аддитивного шумового сигнала размерности $(N_R\times1)$ [6].

Квадратурная амплитудная модуляция (QAM), включает QAM-16, QAM-32, QAM-64, QAM-128, QAM-256, QAM-512, QAM-1024, QAM-2048 и QAM-4096, где цифры показывают количество точек созвездия и количество возможных состояний, которые могут существовать [2].

Современные беспроводные сети часто требуют более высоких мощностей. При фиксированном отношении сигнал/шум увеличение порядка модуляции QAM увеличивает пропускную способность канала. Стоит отметить, что приращение пропускной способности при низких порядках QAM является

значительным; но при высоких порядках QAM прирост пропускной способности намного меньше. Например, увеличение:

1. С QAM-1024 до QAM-2048 дает прирост пропускной способности на 10,83%.
2. От QAM-2048до QAM-4096 дает прирост пропускной способности на 9,77% [4].

Таблица 1

Увеличения пропускной способности QAM

Bits per symbol	Modulation		Incremental % increase
8	256	QAM	14.56%
9	512	QAM	13.80%
10	1024	QAM	11.98%
11	2048	QAM	10.83%
12	4096	QAM	9.77%

Приемники в системе MIMO

ML Detection (Maximum Likelihood Detection). Детектор максимального правдоподобия

В этом разделе представлено несколько хорошо известных алгоритмов детектирования MIMO, включая детекторы на основе ML и linear (ZF и MMSE).

Представим упрощенную формулу принятого сигнала как:

$$\mathbf{y}=\mathbf{H}\bullet\mathbf{s}+\mathbf{n} \tag{3}$$

где матрица канала, которая, в свою очередь, может быть записана как:

$$\mathbf{H}=[h_1,h_2,\ldots,h_M], \tag{4}$$

где h_m обозначает вектор m- го столбца $\mathbf{H}$.

Суть детектора в MIMO заключается в оценке неизвестного вектора передаваемого сигнала s для заданного вектора принятого сигнала y и коэффициента усиления канала. Невозможно предсказать вектор шума n, но есть знания обо всех возможных комбинациях s, которые рассматриваются как возможные векторы.

Для M передающих антенн с алфавитами сигналов количество возможных векторов задается $|S|^M$, где $|S|$ обозначает размер алфавита. Например, когда передатчик оснащен 2 антеннами и для передачи сигналов используется QAM-4, у нас есть $4^2=16$ возможных кандидатов на s [12].

Linear Detection. Линейное детектирование

Для уменьшения сложности рассматриваются линейные детекторы. С линейными детекторами принятый сигнал фильтруется линейным фильтром, и каждый символ данных декодируется отдельно. Таким образом, роль линейного фильтра заключается в подавлении мешающих сигналов [5].

ZF Detection (Zero Forcing Detection). Детектирование с декорреляцией

Алгоритм ZF задается с помощью:

$$W_{zf}=H(H^HH)^{-1} \tag{5}$$

и с помощью:

$$\tilde{s}_{zf}=W_{zf}^Hy=(H^HH)^{-1}H^Hy=s+(H^HH)^{-1}H^Hn\,. \tag{6}$$

MMSE Detection (Minimum Mean Squared Error Detection). Детектор минимальной среднеквадратичной ошибки.

Чтобы уменьшить воздействие фонового шума, детектор MMSE использует линейный фильтр, который может учитывать шум. Фильтр MMSE может быть найден, минимизировав среднеквадратичную ошибку (MSE):

$$W_{mmse}=\arg_w^{\min}E[\|s-W^Hy\|^2]$$
$$=(E[yy^H])^{-1}E[ys^H]$$
$$=H(H^HH+\frac{N_0}{E_S}I)^{-1} \quad (7)$$

где E_S обозначает энергию символа. Результирующий оценочный вектор символов задается формулой:

$$\tilde{s}_{mmse}=W_{mmse}^Hy$$
$$=(H^HH+\frac{N_0}{E_S}I)^{-1}H^yy \quad (8)$$

и из этого следует, что:

$$C_{mmse}=E[(s-W_{mmse}^Hy)(s-W_{mmse}^Hy)^H]$$
$$=I-H^H(H^HH+\frac{N_0}{E_S}I)^{-1}H$$
$$=(I+\frac{E_S}{N_0}H^HH)^{-1}. \quad (9)$$

MSE для каждого символа детектора MMSE может быть получена из соответствующего диагонального элемента C_{mmse} [9].

Результаты моделирования

Начиная с процесса модуляции QAM в передатчике к приемнику в беспроводной основной полосе частот, каждый символ в созвездии QAM представляет уникальную амплитуду и фазу. Следовательно, их можно отличить от других точек на приемнике.

• QAM-64 или любая другая модуляция применяется к входным двоичным битам.

• Модуляция QAM преобразует входные биты в сложные символы, которые представляют биты путем изменения амплитуды/фазы сигнала во временной области. Использование QAM-64 преобразует 6 бит в один символ на передатчике.

• Преобразование битов в символы происходит в передатчике, в то время как обратное преобразование происходит в приемнике. В детекторе один символ выдает 6 бит в качестве выходных данных.

• Модуляция более высокого порядка требует использования высоколинейного усилителя мощности на передающем конце [7].

Высокий порядок модуляции может обеспечить систему высокой скоростью обмена информации, и более высокому уровню спектральной эффективности для системы радиосвязи. Но стоит заметить, что это огромная разница по цене. Так же модуляции с высоким порядком модуляции очень чувствительны к помехам и шумам [10].

Исходя из графика, можно сделать вывод, что при использовании модуляции с высоким порядком модуляции, скорость передачи информации растет в два раза, в сравнении с популярным видом модуляции QAM-64, но у всего есть своя цена. Система становится сильно чувствительна к помехам и вероятность ошибки сильно растет.

В следствии чего, многие системы связи применяют динамические адаптивные методы модуляции, чтобы получить высокую скорость обмена информации для заданных условий. С возвращением к схеме с низким порядком модуляции связь будет с меньшим коэффициентом ошибок на бит информации.

Выбор правильного порядка модуляции QAM для любой конкретной ситуации и способность динамически адаптировать ее, могут позволить получить оптимальную пропускную способность для условий линии на данный момент. Уменьшение порядка модуляции QAM позволяет достигать более низкой вероятности ошибки на бит. Таким образом, для заданного качества связи будет подобран свой порядок модуляции [11].

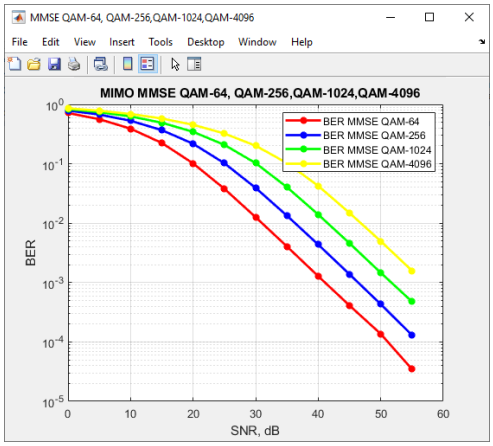

Рис.2. График MIMO MMSE QAM-64, QAM-256, QAM-1024, QAM-4096

Таблица 2

Характеристики модуляции QAM

Характеристики	QAM-256	QAM-1024	QAM-4096
Количество бит на символ	10	11	12
Скорость передачи символов	1/8 бит	1/10 бит	1/12 бит
Общее количество точек в диаграмме созвездия	256	1024	4096
Увеличение пропускной способности по сравнению с QAM-64	33,33%	66.66%	100%

В таких случаях увеличение мощности передачи не приводит к значительному улучшению характеристик линии связи, поскольку это повлияет на линейность систем, например, усилителей. Следовательно, коэффициент усиления системы будет ниже при заданной мощности передачи. Чтобы поддерживать стабильное отношение несущей к помехе, интегрированные усилители мощности также должны работать в линейной области. Это требует уменьшения мощности передачи в схемах, таких как QAM-4096, по сравнению с другими схемами QAM более низкого порядка, такими как QAM-256, QAM-64, QAM-16 и QAM-8.

Другим способом обеспечить более высокое отношение сигнал/шум и сохранить производительность канала связи является переход к модуляции QAM более низкого порядка [13].

Заключение

У каждого порядка модуляции можно найти свои плюсы и минусы. Когда мы достигаем QAM-4096, размер созвездия становится чрезвычайно большим, а точки становятся очень близкими друг к другу. При наличии шума и многолучевых помех существует более высокая вероятность того, что рядом лежащие две точки созвездия могут быть перепутаны друг с другом при приеме. Снижение помехоустойчивости QAM-4096 является не единственной проблемой, хоть и скорость передачи модуляции высокого порядка будет увеличена путем отображения боле 1 бита на одной несущей, для декодирования битов в приемнике потребуется высокое количество децибел. В сравнении с QAM-64, где скорость передачи информации вдвое ниже, при одинаковом значении коэффициента ошибок на бит, отношение сигнал/шум может быть на 15 децибел ниже, чем в случае QAM-4096. Если же рассматривать модуляцию при фиксированном отношении сигнал/шум, увеличение порядка модуляции QAM увеличивает пропускную способность канала. Но здесь тоже есть разумный предел. При увеличении порядка модуляции, количество ошибок на бит будет расти значительно сильнее, чем пропускная способность. Опираясь на конкретную ситуацию и цели, которые необходимо достичь, можно подобрать, какой вид модуляции QAM, с высокими порядками или же с низкими лучше всего подойдет. В заключение следует отметить, что в настоящее время наблюдается устойчивая тенденция к повышению используемых порядков модуляции в современных и перспективных системах связи. Например, в перспективных вариантах технологии Wi-Fi уже предусмотрено использование модуляции QAM-1024 [8].

Литература

1. *Borges D., Montezuma P., Dinis R., Beko M*. Massive MIMO Techniques for 5G and Beyond – Opportunities and Challenges // Electronics 2021, 10, 1667.

2. *Elshokry A., Abu-Hudrouss A*. Performance Evaluation of MIMO Spatial Multiplexing Detection Techniques // Journal of Al Azhar University-Gaza (Natural Sciences), 2012, 14.

3. *Ajay R*. Introduction to MIMO and Massive MIMO. Fundamentals of Network Planning and Optimisation 2G/3G/4G: Evolution to 5G, 2018, 1443.

4. *Shaoshi Y., Lajos H.* Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs. Accepted to appear on IEEE communications surveys & tutorials, 2015, 24.

5. *Hien Quoc N*. Massive MIMO: Fundamentals and System Designs. Division of Communication Systems Department of Electrical Engineering (ISY) Linköping University, 2015, SE-581 83.

6. *Erik G. Larsson, Ove Edfors, Fredrik Tufvesson, Thomas L. Marzetta*. Massive MIMO for Next Generation Wireless Systems, 2014 // IEEE Communications Magazine, (52), 2, pp. 186-195.

7. *Larsson E.* Massive MIMO for Next Generation Wireless Systems // IEEE Communications Magazine, 2014.

8. *Ермолаев В.Т., Флаксман А.Г.* Теоретические основы обработки сигналов в беспроводных системах связи: Монография. Нижний Новгород: Изд-во ННГУ им. Н.И. Лобачевского, 2011. 368 с.

9. *Hampton J*. Introduction to MIMO Communications. Cambridge University Press 2014.

10. *Бакулин М.Г., Крейнделин В.Б., Панкратов Д.Ю.* Технологии в системах радиосвязи на пути к 5G. М.: Горячая линия – Телеком, 2018. 279 с.

11. *Бакулин М.Г., Варукина Л.А., Крейнделин В.Б.* Технология MIMO. Принципы и алгоритмы. М.: Горячая линия – Телеком, 2014. 244 с.

12. *Крейделин В.Б., Смирнов А.Э., Бен Режеб Т.Б.К*. Исследование радиоинтерфейса беспроводных систем межма-шинного взаимодействия M2M. // T-Comm: Телекоммуникации и транспорт. 2014. Т. 8. 6. С. 71-74.

13. *Kreyndelin V., Smirnov A., Ben Rejeb T*. Effective precoding and demodulation techniques for 5G communication systems // 2018 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, 14-15 марта 2018 г.

14. *Бакулин М.Г., Крейнделин В.Б., Панкратов Д.Ю*. Анализ пропускной способности канала MIMO в условиях замираний // Системы синхронизации, формирования и обработки сигналов. 2018. Т. 9. № 2. С. 13-20.

15. *Бакулин М.Г., Крейнделин В.Б*. Проблема повышения спектральной эффективности и емкости в перспективных системах связи 6G // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 2. С. 25-31.

16. *Крейнделин В.Б., Резнёв А.А*. Матрица пространственно-временного кода высокой размерности типа "Голден" // T-Comm: Телекоммуникации и транспорт. 2018. Т. 12. № 6. С. 34-40.

17. *Бакулин М.Г., Крейнделин В.Б., Панкратов Д.Ю*. Алгоритмы нелинейной фильтрации двоичной ЛРП со случайной задержкой и случайной начальной фазой // Системы синхронизации, формирования и обработки сигналов. 2019. Т. 10. № 2. С. 45-51.

18. *Крейнделин В.Б., Григорьева Е.Д*. Анализ быстрого алгоритма умножения матриц и векторов для банка цифровых фильтров // T-Comm: Телекоммуникации и транспорт. 2021. Т. 15. № 1. С. 4-10.

19. *Бакулин М.Г., Бен Режеб Т.Б.К., Крейнделин В.Б., Смирнов А.Э*. Способы минимизации объёма передаваемой информации в обратном канале многоантенных систем MIMO // T-Comm: Телекоммуникации и транспорт. 2021. Т. 15. № 3. С. 17-24.

20. *Бакулин М.Г., Крейнделин В.Б., Панкратов Д.Ю*. Применение технологии MIMO в современных системах беспроводной связи разных поколений // T-Comm: Телекоммуникации и транспорт. 2021. Т. 15. № 4. С. 4-12.

21. *Крейнделин В.Б., Григорьева Е.Д.* Реализация банка цифровых фильтров с пониженной вычислительной сложностью // T-Comm: Телекоммуникации и транспорт. 2019. Т. 13. № 7. С. 48-53.

22. *Панкратов Д.Ю., Степанова А.Г.* Компьютерное моделирование технологии MIMO для систем радиосвязи // T-Comm: Телекоммуникации и транспорт. 2018. Т. 12. № 12. С. 33-37.

23. *Панкратов Д.Ю., Сердюков А.А*. Моделирование системы MIMO в режиме Beamforming. DSPA: Вопросы применения цифровой обработки сигналов. 2021. Т. 11. № 2. С. 12-21.

24. *Бакулин М.Г., Бен Режеб Т.Б.К., Крейнделин В.Б., Панкратов Д.Ю., Смирнов А.Э.* Технология NOMA с кодовым разделением в 3GPP: 5G или 6G? // T-Comm: Телекоммуникации и транспорт. 2022. Т. 16. № 1. С. 4-10.

25. *Бакулин М.Г., Бен Режеб Т.Б.К., Крейнделин В.Б., Миронов Ю.Б., Панкратов Д.Ю., Смирнов А.Э*. Схемы модуляции для систем сотовой связи 5G/IMT-2020 и 6G // T-Comm: Телекоммуникации и транспорт. 2022. Т. 16. № 3. С. 11-17.

26. *Бакулин М.Г., Бен Режеб Т.Б.К., Крейнделин В.Б., Панкратов Д.Ю., Смирнов А.Э.* Схемы NOMA с обработкой на уровне символов // T-Comm: Телекоммуникации и транспорт. 2022. Т. 16. № 5. С. 4-14.

СЕМАНТИЧЕСКИЙ И ВЕРОЯТНОСТНЫЙ ВЕКТОРЫ В ПОИСКОВЫХ ЗАПРОСАХ

Гадасин Денис Вадимович,
Московский Технический Университет Связи и Информатики, зам. зав. каф. СИТиС доцент, к.т.н., Москва, Россия
dengadiplom@mail.ru

Шведов Андрей Вячеславович,
Московский Технический Университет Связи и Информатики, ст. преп. каф. СИТиС, Москва, Россия
a.v.shvedov@mtuci.ru

Вакурин Илья Сергеевич,
Московский Технический Университет Связи и Информатики, аспирант 2 года обучения, Москва, Россия
vort57@mail.ru

Тремасова Лилия Андреевна,
Московский Технический Университет Связи и Информатики, магистрант гр. М092101(75), Москва, Россия
lila.trem@yandex.ru

Аннотация
Статья посвящена тому, как можно оптимизировать кодировку кириллических символов в цифровых системах для ускорения обработки поисковых запросов при помощи морфем. Для этого рассматривается понятие синтаксического и вероятностного вектора. Анализируются частотные и амплитудные характеристики русского языка при помощи ряда Фурье. Сравнивается русский язык и языки, которые имеют латинский алфавит или используют как кириллицу, так и латиницу. Строятся и анализируются графики векторов алфавитов.

Ключевые слова: *Язык, кодирование, синтаксический вектор, морфемы, вероятностный вектор, Unicode.*

Введение

Процесс кодирования информации подразумевает под собой обработку самой информации [1, 2], при которой происходит замена одной последовательности символов другой последовательностью символов с соблюдением двух основных правил:

- кодирование и декодирование должно быть однозначным;
- избыточность должна быть минимальной.

Кодирование применяется, как при хранении, так и при обработке [3, 4] и передаче данных [15-31], с его помощью решаются такие вопросы, как:

- сжатие данных для хранения и передачи;
- обеспечение безопасности данных;
- контроль и исправление ошибок;
- представление данных в нужном виде.

Вопрос представления данных остро встал в начале эпохи зарождения глобальной сети Интернет, когда при передаче файлов от одного устройства к другому, принимающие устройство не могло считать переданный файл, не понимая его кодировки [5]. Кроме того, в мире существует множество языковых систем и алфавитов и не все кодировочные системы могли уместить все символы [6].

Для решения проблемы была создана кодировка Unicode, которая включала в себя все основные письменности мира и подразумевала поэтапное их добавление. Но единицы символьной системы бывают разными и при этом несут в себе разный объём информации.

В русском языке Unicode кодирует каждую букву своим кодом, следовательно, при осуществлении запросе в поисковую систему каждый символ будет иметь вес в 16 бит. Но, например, в японском языке один символ азбуки обозначает не звук, а набор звуков – слог, а Unicode кодирует символ, который

несет в себе больше смысловой нагрузки, умещая в себе несколько отдельных звуков. При этом в японском алфавите слова могут строится не только из символов слоговых азбук катаканы и хираганы, но и из канзи символов, обозначающих целые слова, которые в Unicode кодируются одним кодом. Таким образом, меньшими затратами на кодирование кодируется больше единиц, несущих информацию.

В русском языке есть несколько видов единиц информации. Слова состоят из слогов, но помимо них слово можно разобрать на морфемные составляющие:

- приставку;
- корень;
- суффикс;
- окончание;
- постфикс.

Морфема – наименьшая единица языка, имеющая смысл. Если единица имеет единица имеет смысл, значит она несёт в себе некоторую информацию, которая может быть закодирована.

Морфемные составляющие зачастую так же представляют из себя слоги, но обладают более устойчивой структурой, правилами формирования и применения [7], что в русском языке даёт более благоприятные условия для создания кодовой таблицы.

Словообразование в русском языке

Процесс словообразования в русском языке представлен на рисунке 1.

Рис. 1. Словообразование в русском языке

Как видно из рисунка существует несколько уровней преобразования звуков, которые человек произносит для обозначения предмета или явления окружающей действительности.

Для графической записи звука в современном русском языке используются буквы, которые обозначают один конкретный звук, не считая букв «ъ» и «ь», использующихся для альтерации.

Следующим уровнем идут слоги и морфемы.

Слоги строятся из соображений фонетики и состоит из звуков, чья природа имеет наибольшую акустическую гармонию [8], например, сочетание звуков «ся», «ля», «ом» звучат гармонично и легко произносятся, когда сочетания звуков «нм», «рн» и «вц» произнести сложно, поэтому они и не являются слогами. Слог никак не отражает смысловое формирование слова, формируясь только с позиции эстетики и эргономики, потому и играет обширную роль в поэзии и музыке.

Формирование слов из морфем в свою очередь имеет как раз более весомую смысловую основу [9]. Морфема представляет из себя некую абстракцию, считаясь не знаком, а скорее классом знаков. Реализация морфемы при письме называется морфом, а совокупность морфов одной морфемы с одинаковым звучанием алломорфом. Например, в предложении: «Я бегу, и ты бежишь, а он не бежит» морфема «бег-» есть в трёх формах:

- «бег-» в «бегу»;
- «беж-» в «бежишь»;
- «беж» в «бежишь».

Но алломорфов в неё двое:

- «бег-»;
- «беж-».

Чтобы два морфа относились к одному алломорфу они должны обязательно иметь одинаковый фонемный состав и ударение.

Третьим уровнем преобразования звука в письменную речь идёт слово.

Слово состоит и делится на вышеперечисленные составляющие. Слова склонны появляться и исчезать из употребления, а также менять значение.

Таким образом, самой стабильной единицей является морфема, так как имеет устойчивое формирование и определённые правила применения, которые за всю историю русского языка претерпели лишь незначительные изменения. Поэтому и имеет смысл создания кодов именно для морфем, а не для

Части слова отделяются друг от друга по определенным морфологическим признакам. Закодировать морфему можно исходя из её длинны или частоты её появления в тексте, ускоряя таким образом процесс кодирования и декодирования по такой системе и как следствие снижая необходимое время на обработку запроса.

Принципы и механизмы работы Unicode

Unicode как система имеет несколько принципов на основе которых она строится и развивается.

Принцип гарантии стабильности обеспечивает однозначность дешифровки текста, закодированного Unicode. Если открыть файл, созданный в 2010 году в 2022 году, то его содержимое отобразится корректно ввиду того, что ни один символ не может быть исключён из кодировки или у него не может быть изменён номер.

Принцип динамической компоновки отвечает за то, что символы можно группировать друг с другом, не занимая при этом новые коды. Например, сочетание умлаута «точек над буквой» и буквы «е» даёт букву «ё».

Принцип универсальности подразумевает, что Unicode не кодирует символы из языковых систем, где:

- мало информации, необходимой для надежного и однозначного кодирования символов;
- нет каких-либо стандартов и правил;
- используется пиктографика.

Принцип унификации гласит, что если символ, например, «w» присутствует в английском(дабл-ю), немецком(ви) и французском(дубль-вэ) алфавите, то для него используется одна кодовая ячейка. Исключения составляют схожие символы из разных письменностей.

Опираясь на данные принципы, консорциум Unicode пытается осуществлять оптимальное добавление символов в кодовую таблицу из различных символьных систем, отвечающих принципам Unicode.

Общий вид Unicode таблицы представлен на рисунке 2.

Рис. 2. Unicode таблица

Столбцы нумеруются в шестнадцатеричной системе счисления от 0 до F, строки от 0000 до FFFF в каждой плоскости.

Плоскости в Unicode – это сегменты, которые представляют из себя таблицу. Указателем на плоскость является первое число кода, например, код U+1A32D относится ко второй плоскости. Плоскости в Unicode делятся на блоки.

Плоскость в Unicode представляет из себя непрерывный диапазон из 65 536 кодовых позиций. Существует 17 плоскостей, обозначенных числами от 0 до 16, что соответствует возможным значениям 00—1016 первым двум шестнадцатеричным цифрам в шестизначном формате номера кодовой позиции (U+hhhhhh). Последняя кодовая позиция в Unicode — последняя кодовая позиция в плоскости 16, U+10FFFF. Плоскость 0 называется «Основная многоязычная плоскость» (Basic Multilingual Plane, BMP), которая содержит наиболее часто используемые символы. В версии Unicode 13.0 задействованы кодовые позиции семи плоскостей, при этом две из них предназначены для частного использования.

Ограничение в 17 плоскостей обусловлено кодировкой UTF-16, в которой могли быть закодированы 220 кодовых позиций (16 плоскостей) и BMP. Кодировка UTF-8 была разработана с гораздо большим

лимитом в 231 (2 147 483 648) кодовых позиций (32 768 плоскостей) и могла задействовать 221 (2 097 152) кодовых позиций (32 плоскости) даже при лимите 4 байта.

Первая плоскость содержит кодовые точки от U+0000 до U+FFFF, то есть наиболее часто используемые символы. Остальные шестнадцать плоскостей (U+010000 → U+10FFFF) называются дополнительными или астральными. Плоскости с 5 по 13 в настоящие время не используются.

Каждая плоскость содержит блоки с шестнадцатеричной нумерацией от 00 до FF, блоки делятся по формату письменности и географическим регионам, где зародилась языковая система, использующая символы блока, например, существует блок латинской письменности, а также нелатинской европейской письменности.

На рисунке 3 представлена нулевая плоскость Unicode.

Рис. 3. Нулевая плоскость Unicode

Как уже отмечалось нулевая плоскость является основной и содержит основные и наиболее часто используемые символы Unicode. Блоки не разделены между собой однородно, так как разные символы, относящиеся к разным группам, были добавлены с разними версиями, а согласно одному из базовых принципов Unicode, если символ получил один код, то он не может поменять его. Не смотря на это общая структура группировки все же прослеживается. Самой крупной группой является Идеограммы ККЯ, куда относятся китайские иероглифы, а также языковые системы на их основе.

На рисунке 4 представлена дополнительная многоязычная плоскость.

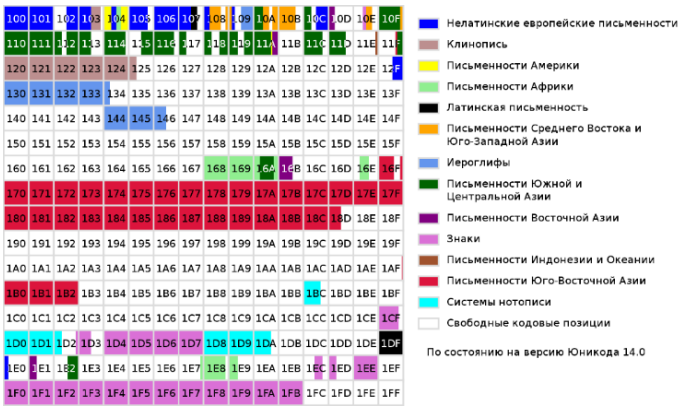

Рис. 4. Дополнительная многоязычная плоскость

Во второй по счёту плоскости №1 содержаться символы, не попавшие в первую по счёту плоскость. Свободного места в дополнительной многоязычной плоскости ещё 60%, что говорит о том, что в стандарте Unicode ещё достаточно места для внесения новых символов или мировых алфавитов.

Кодирование морфем

Как уже отмечалось ранее, при морфемном разборе, без учёта корня, слово делится на:

– приставку;
– суффикс;

– окончание;
– постфикс.

Для примера закодируем слово «переходящий» по стандарту UTF-16BE.

Всего в слове 11 букв, которые кодируются последовательностями весом в 16 бит, вычислим по формуле 1 вес слова «переходящий» в кодировке UTF-16BE.

$$I = k \cdot i = 11 \cdot 16 = 176 \ бит \quad (1)$$

Для кодирования морфем требуется выделить непосредственно части слова, разбор на морфемы показан на рисунке 5.

Рис. 5. Разбор слова на морфемы

Всего в этом слове четыре морфемы, корень будем кодировать по буквам, а приставке, корню и окончанию присвоим значения из диапазона 126-12A, результат показан в таблице 1.

Таблица 1

Присвоение кодов

№	Буква	Код
1.	пере	U+1254F
2.	ящ	U+125A8
3.	ий	U+12A41

Таким образом, по формуле 2 рассчитаем вес слова при таком способе кодирования.

$$I = (k_б + k_m) \cdot i = (3+3) \cdot 16 = 96 \ бит \quad (2)$$

где:

- $k_б$ – количество букв в корне;
- k_m – количество морфем без учёта корня.

Сравнение двух методов по объёму слова представлено в графике на рисунке 6.

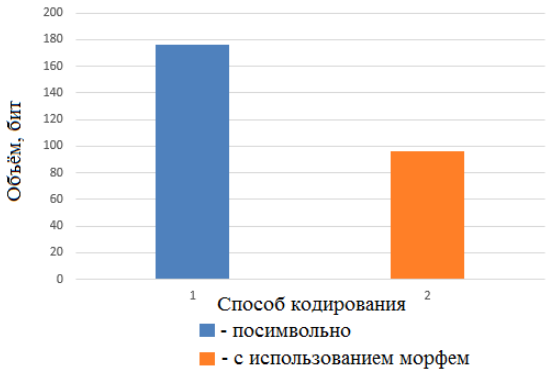

Рис. 6. Сравнение методов по объёму слова

Из графика можно увидеть, вес слова, закодированного при помощи разбиения на морфемы ниже, чем вес слова, закодированного по буквам.

В кодировании показателем того, на сколько тот или иной метод кодирования является эффективным, служит информационная энтропия [10]. Информационная энтропия показывает сколько смысловой информации умещается в каждом символе алфавита [11].

Информационная энтропия рассчитывается по формуле Хартли [12]:

$$i = \log_2 K \quad (3)$$

где:

K – мощность алфавита;

Рассчитаем информационную энтропию для кодирования по одной букве по формуле 4, мощность алфавита возьмем равную 33 по количеству букв в русском языке.

$$i = \log_2 K = \log_2 33 = 5.044 \ бит / символ \quad (4)$$

Рассчитаем информационную энтропию для кодирования с использованием морфем по формуле 5, мощность алфавита возьмем равную 609 по количеству морфем в русском языке без учёта корня.

$$i = \log_2 K = \log_2 33 = 9.250 bit / symbol \quad (5)$$

Сравнение двух методов по энтропии представлено в графике на рисунке 7.

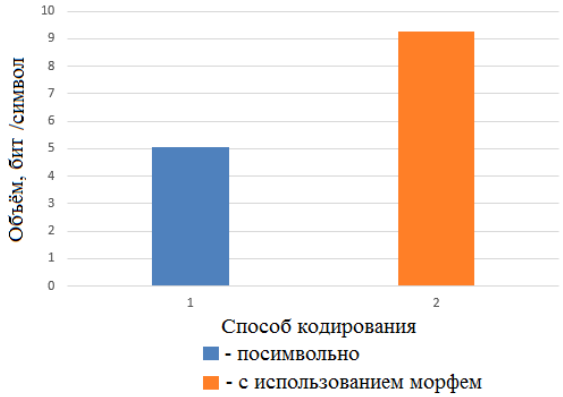

Рис. 7. Сравнение методов по объёму информации на символ

Таким образом, можно заключить, что при кодировании с разделением на морфемы информации в одном символе содержится больше, чем при кодировании по буквам, что в конечном счёте положительно сказывается на эффективности, путём кодирования большей смысловой нагрузки меньшим количеством бит.

В других языках также, как и в русском, тоже присутствуют морфемы, но они конечно же могут являться более значимыми и нести в себе большую информационную ценность.

Например, в английском языке слово «переходящий» звучит, как «transitory» и разбирается по составу, как показано на рисунке 8.

Рис. 8. Разбор слова на морфемы

Как видно из рисунка в английском варианте слова морфем гораздо меньше, но тут стоит отметить, что суффикс -ящ сам по себе без окончания существовать не может и для того, чтобы из существительного «переход» в русском языке сделать прилагательное «переходящий» нужно применить две морфемы разного вида, которые друг без друга в этом слове не имеют смысла, когда в английском варианте слова при помощи только одного суффикса -ory есть возможность сделать из существительного «transit» прилагательное «transitory».

Рассчитаем ценность информации, содержащуюся в суффиксе, учитывая, что в русском языке суффикс без окончания не несёт вообще никакой информации.

$$H(X_1) = -(x+a)^n = \\ = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) = 4(p(x_1) \log_2 p(x_1)) \quad (6)$$

$$H(X_2) = -(x+a)^n = \\ = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) = 3(p(x_1)\log_2 p(x_1)) \quad (7)$$

Соответственно, получаем, что:

$$H(X_1) = 4(p(x_1)\log_2 p(x_1)) = -4(0.25\log_2 0.25) = 8bit \quad (8)$$

$$H(X_2) = 3(p(x_1)\log_2 p(x_1)) = -3(0.33\log_2 0.33) = 4.5bit \quad (9)$$

Таким образом, получаем, что в русском варианте слова на один символ приходится 2 бита, когда в английском языке на один символ приходится 1.5 бита на один символ, что говорит о том, что в разных языковых системах для обозначения одного и того же действия может понадобится разное количество информации.

Синтаксические векторы

Синтаксический вектор – это совокупность позиции лингвистической единицы (буквы, слога, морфемы) и её фонетических признаков в алфавите или тексте, при котором можно сформировать наиболее ёмкое по информации сообщение [13].

Алфавит представляет из себя некоторый набор символов в таблице, чаще всего выстроенный по некоторым устоявшимся историческим правилам [14], например, можно взять три алфавита и выстроить между ними некоторые соответствия путём сравнения позиций разных букв, задав каждой букве координаты, для примера возьмем три алфавита, проведя условный географический вектор с востока на запад, как базовый алфавит возьмем Сербский, так как там присутствуют кириллический и латинский варианты алфавита, Русский и Английский, как отклонения на восток и запад.

Как можно увидеть, что на позиции (1,1) чаще всего находится буква, читающаяся, как [а] и обозначается она всегда символом А, как и на позиции (2,2) находится буква, обозначающая звук [б], следовательно, для языков можно выстроить некоторые вектора, где какие-то фонетические точки будут иметь сходство, а какие-то различия. Для этой задачи можно применить формулу 10, конечно же полная сходимость векторов невозможна, учитывая языковую специфику и разное количество букв в алфавитах.

$$\overline{g} = \{(i,i);(i+n,i+n)\} \quad (10)$$

Составим три синтаксических вектора по координатам, которые имеют одинаковые звук:

$$\overline{g}_a = \{(1.1);(1.2);(н);(2.1);(н);(1.4);(н);(н);(н);(1.6);\\ (1;7);(н);(1.5);(2.3);(2.4);(2.5);(н);(2.6);(2.7);(н);(3.1);\\ (3,2);(3.4);(3.5);(н);(3.6);(3.7);(4.4);(4.5);(н)\} \quad (11)$$

$$\overline{g}_c = \{(1.1);(1.2);(1.3);(1.4);(1.5);(1.6);(1.7);\\ (2.1);(2.2);(2.3);(2.4);(2.5);(2.6);(2.7);(3.1);(3.2);\\ (3.3);(3.4);(3.5);(3.6);(3.7);\\ (4.1);(4.2);(4.3);(4.4);(4.5);(4.6);(4.7);(5.1);(5.2)\} \quad (12)$$

$$\overline{g}_p = \{(1.1);(1.2);(4.3);(4.4);(н);(1.5);(н);(н);\\ (1.6);(4.1);(н);(4.2);(2.3);(н);(2.5);(2.6);(н);\\ (2.7);(3.1);(н);(3.2);(3.3);(3.4);(3.5);(н);\\ (3.6);(3.7);(5.4);(2.2);(2.1)\} \quad (13)$$

где $н$ – отсутствие соответствия в позиции.

Из получившихся векторов видно, что они имеют максимальное различие в начале и конце, в середине различия практически нивелируются, а в центральных позициях начинают совпадать, что говорит о том, что существует определённый набор звуков, встречающийся в каждом алфавите.

На рисунке 9 представлен график сравнения полученных векторов.

Рис. 9. График сравнения полученных векторов

В графике вектор Сербского языка выступает, как некоторая нормаль, где присутствуют все символы, Русский и Английский являются некоторым отклонением. Однако из графика видно, что на некоторых этапах существует сильное сходство между фонетическим набором в Русском и Английском языках.

Для того, чтоб рассчитать степень их похожести воспользуемся так называемой косинусной мерой и рассчитаем значения косинуса между векторами в трёх сечениях.

Возьмем три сечения на графике из рисунка 10 и нанесём на график рисунка 10.

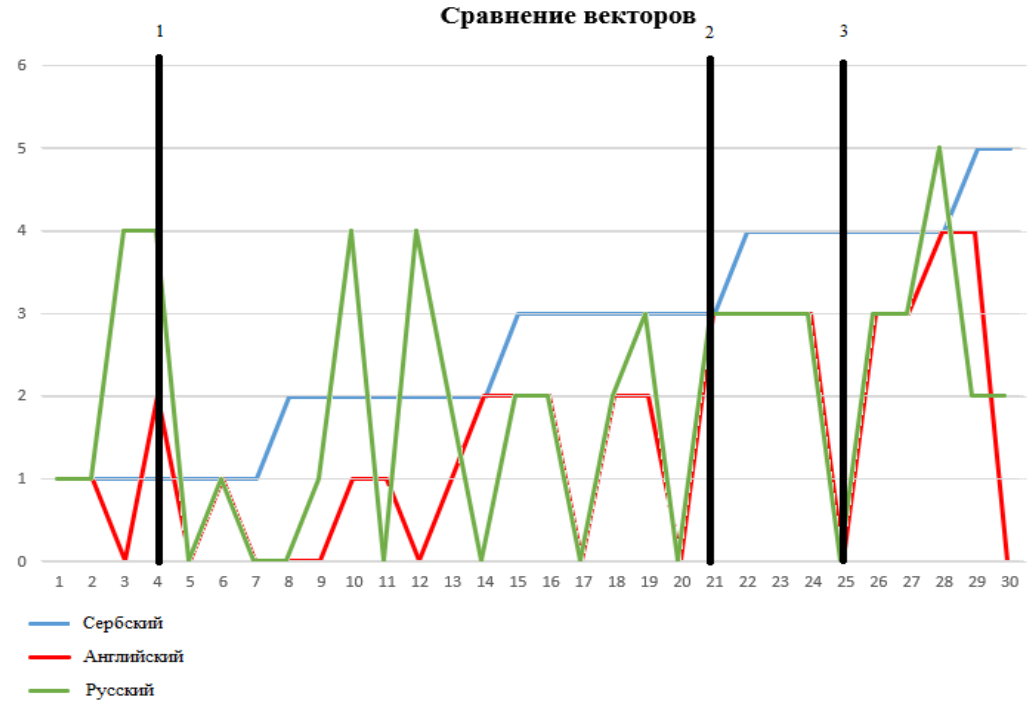

Рис. 10. График векторов с сечениями

Для простоты расчёта косинусной меры аппроксимируем графики, аппроксимированные графики показаны на рисунке 11.

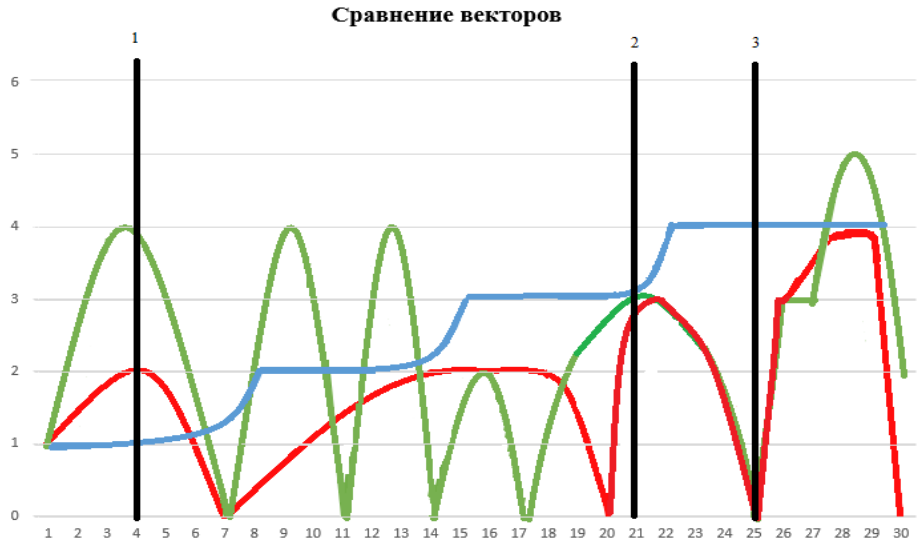

Рис.11. Аппроксимированные графики

Для расчёта косинусного расстояния воспользуемся формулой:

$$\cos(\varphi) = \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}} \tag{14}$$

Рассчитаем значения в отношении вектора Сербского языка к Английскому и к Русскому в сечении 1 по следующим формулам, соответственно:

$$\cos(\varphi)_{se1} = \frac{2 \cdot 1 + 1 \cdot 4}{\sqrt{2^2 + 1^2}\sqrt{1^2 + 4^2}} = 0.65 \tag{15}$$

$$\cos(\varphi)_{sr1} = \frac{1 \cdot 4 + 4 \cdot 4}{\sqrt{1^2 + 4^2}\sqrt{4^2 + 4^2}} = 0.8 \tag{16}$$

Рассчитаем значения в отношении вектора Сербского языка к Английскому и к Русскому в сечении 2 по следующим формулам, соответственно:

$$\cos(\varphi)_{se2} = \frac{3 \cdot 3 + 7 \cdot 1}{\sqrt{3^2 + 7^2}\sqrt{3^2 + 1^2}} = 0.66 \tag{17}$$

$$\cos(\varphi)_{sr2} = \frac{3 \cdot 3 + 7 \cdot 2}{\sqrt{3^2 + 7^2}\sqrt{3^2 + 2^2}} = 0.84 \tag{18}$$

В сечении номер 3 вычислить косинусную меру нельзя так, как только в Сербском языке вектор имеет отличные от нуля координаты.

В сечении 1 значения косинуса меньше, чем в сечении 2, что говорит о том, что мера схожести во втором сечении выше, чем в первом, что можно увидеть и в разнице значений косинусов для каждого сечения, она показана в формулах 19 и 20.

$$\cos(\varphi)_{sr1} - (\varphi)_{se1} = 0.8 - 0.65 - 0.15 \tag{19}$$

$$\cos(\varphi)_{sr2} - (\varphi)_{se2} = 0.84 - 0.66 = 0.18 \tag{20}$$

Разница в косинусах во втором сечении больше разницы косинусов в первом сечении, что говорит о том, что разница между векторами во втором сечении ниже, чем в первом, так как значение косинуса

стремится к единице, как и в сечении номер три, где значения нулевые, а косинус в нуле принимает значение единицы.

Если взять координаты звуков, которые присутствуют во всех трех алфавитах и составить из них квадратную матрицу, то можно будет рассчитать строчную детерминанту, взяв строчные координаты:

$$\begin{pmatrix} (1.1) & (1.2) & (1.5) \\ (1.6) & (2.3) & (2.5) \\ (2.6) & (2.7) & (3.2) \end{pmatrix} \tag{21}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 2 & 3 \end{pmatrix} = 1\cdot 2\cdot 3 + 1\cdot 2\cdot 2 + 1\cdot 1\cdot 2 - 1\cdot 2\cdot 2 - \tag{22}$$
$$1\cdot 2\cdot 2 - 1\cdot 1\cdot 3 = 6 + 4 + 2 - 4 - 4 - 3 = 1$$

Так как строковая детерминанта равна 1, то получившиеся вектора имеют высокую семантическую идентичность в точках соприкосновения на графиках.

Поскольку алфавит сам по себе представляет некоторый семантический вектор, а каждая буква представляет собой звук, который характеризуется фазой, частотой и амплитудой, то алфавит можно представить в виде ряда Фурье, в котором при добавлении каждого элемента будет увеличиваться сходимость.

Построим ряд Фурье, используя данные характеристики, приняв, что начальная фаза 0 для всех звуков:

$$\begin{aligned} f(x) = \frac{a_0}{2} + \sum_{n=0}^{31}(A_n \cos w) = (86\cos 20) + \\ +(167\cos 7) + (75\cos 10) + (92\cos 12) + \\ +(83\cos 8) + (75\cos 4) + (78\cos 10) + \\ +(92\cos 10) + (81\cos 20) + (200\cos 20) + \\ +(160\cos 15) + (200\cos 20) + (78\cos 20) + \\ +(70\cos 5) + (86\cos 12) + (8400\cos 4) + \\ +(1300\cos 20) + (8600\cos 15) + (1900\cos 7) + \\ +(2800\cos 20) + (2800\cos 20) + (270\cos 20) + \\ +(78\cos 20) + (240\cos 20) + (75\cos 20) \end{aligned} \tag{23}$$

На рисунке 12 изображен график ряда Фурье для Русского языка:

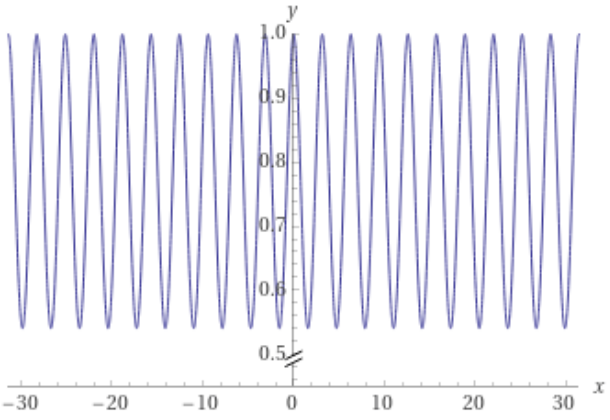

Рис. 12. График ряда Фурье для Русского языка

График имеет достаточно равномерную структуру, что говорит о хорошей сходимости ряда, а следовательно, и высокой семантической взаимосвязи элементов.

Вероятностный вектор

Каждый язык содержит в себе набор слов, их количество постоянно меняется, но основа языка, как правило остаётся неизменной, что позволяет просчитать вероятность появления каждой буквы в тексте или звука в устной речи, так как каждая буква обозначает звук.

Вероятность появления букв в Русском языке представлена на рисунке 13.

Символ	Вероятность	Символ	Вероятность	Символ	Вероятность
Пробел	0,175	К	0,028	Ч	0,012
О	0,090	М	0,026	Й	0,010
Е	0,072	Д	0,025	Х	0,009
А	0,062	П	0,023	Ж	0,007
И	0,062	У	0,021	Ю	0,006
Н	0,053	Я	0,018	Ш	0,006
Т	0,053	Ы	0,016	Ц	0,004
С	0,045	З	0,016	Щ	0,003
Р	0,040	Ъ	0,014	Э	0,003
В	0,038	Б	0,014	Ф	0,002
Л	0,035	Г	0,013		

Рис. 13. Вероятность появления букв в русском языке

Очевидно, что не все буквы в Русском языке обозначают звук или набор звуков, который зависит от позиции буквы в слове, как например, буква ё. Буквы ъ и ь применяются для альтерации и работают в паре с другими символами, поэтому их в дальнейшем анализе они учитываться не будут.

На рисунке 14 представлен график распределения звуков в Русском языке:

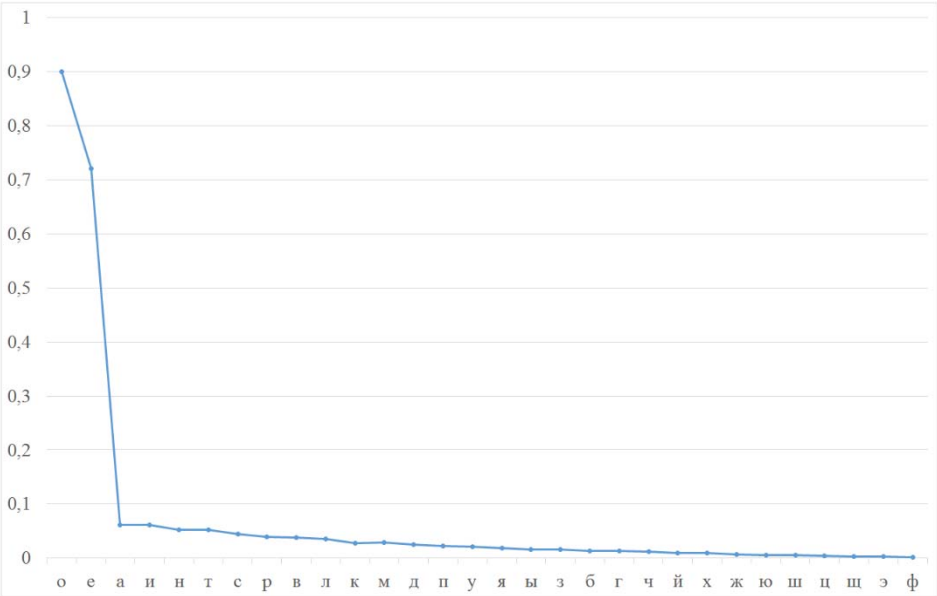

Рис. 14. График распределения звуков

Как видно график имеет характер нормального распределения вероятностей. Построим на его основе вероятностные вектора для слова «нос» и слова «сон».

На рисунке 15 представлен вектор слова «сон».

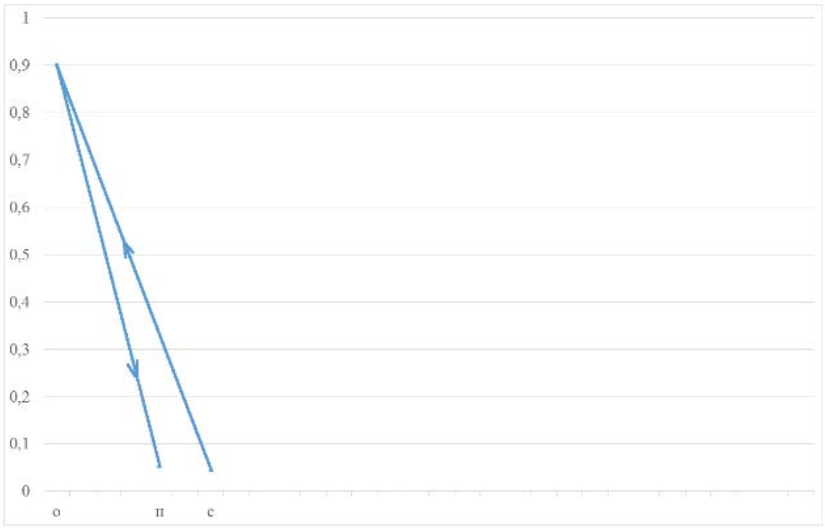

Рис. 15. Вектор слова «сон»

На рисунке 16 представлен вектор слова «нос».

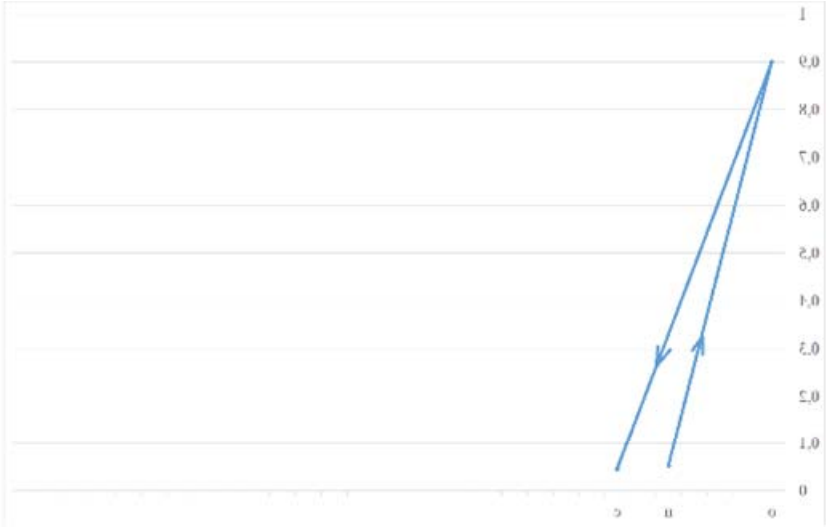

Рис. 16. Вектор слова «нос»

Как можно увидеть из рисунков 16 и 17 вектора слов «сон» и «нос» являются противоположно направленными, но они занимают в векторном пространстве одинаковые позиции, что позволяет сэкономить физическое место на их размещение, например, в базе данных.

Поскольку в базе ячейки со словами «сон» и «нос» находятся в одной области памяти, то потребуется меньше вычислительных мощностей на получение доступа к ним, что позволит снизить нагрузку на телекоммуникационное оборудование.

Заключение

Ряд Фурье удобен для разложения вектора на составляющие и представляет их в виде синуса или косинуса, что позволяет легко сравнивать значения каждого элемента. В поисковых запросах это может быть использовано для создания паттерна для обработки запроса, особенно, если запрос был передан при помощи голосового ввода.

Вероятностный вектор может существенно ускорить доступ к ячейкам памяти базы данных, путём составления векторов, которые занимают в пространстве идентичные позиции.

При кодировании слов с разделением на более простые смысловые единицы-морфемы, можно достичь более оптимального кодирования путём преобразования первоначальной информации в более информационно- ёмкие формы. Это позволит достичь следующих результатов:

- снижение нагрузки на узловое оборудование;
- снижение нагрузки на оконечное оборудование;
- повышение эффективности кодирования;
- повышение экономии места на носителях при хранении данных;
- увеличении скорости передачи данных;
- увеличение скорости обработки информации;
- увеличение скорости переформатирования информации.

Литература

1. *Гадасин Д.В., Шведов А.В., Пантелеева К.А.* Предобработка информации для систем машинного обучения // Актуальные проблемы и перспективы развития экономики : Труды XXI Международной научно-практической конференции, Симферополь-Гурзуф, 20-22 октября 2022 года. Симферополь: Крымский федеральный университет им. В.И. Вернадского, 2022. С. 268-269. EDN QVIOMF

2. Свидетельство о государственной регистрации программы для ЭВМ № 2022662724 Российская Федерация. Программное приложение "Анализатор текстовых данных" ("Text Data Analyzer" - на английском языке) для выполнения лабораторных работ студентами вузов по дисциплине "Мультимедийные информационные системы" : № 2022661351 : заявл. 21.06.2022 : опубл. 07.07.2022 / В. А. Докучаев, В. В. Маклачкова, Д. В. Гадасин [и др.] ; заявитель Общество с ограниченной ответственностью Фирма «ТЕЛЕСОФТ». – EDN DVURCM.

3. *Шведов А.В., Коноплева М.И.* Применение алгоритмов и способов преобразования речевого сигнала в цифровую информацию // Актуальные проблемы и перспективы развития экономики : Труды XXI Международной научно-практической конференции, Симферополь-Гурзуф, 20-22 октября 2022 года. Симферополь: Крымский федеральный университет им. В.И. Вернадского, 2022. С. 290-292. EDN HPOTTL

4. *Иванова О.В., Иванов П.В., Смелов М.Н*. Проблемы и алгоритмы поиска информации в глобальных компьютерных сетях // T-Comm: Телекоммуникации и Транспорт. 2010. С. 23-25.

5. *Гадасин Д.В., Шведов А.В., Ермалович А.В*. Концепция "туманные вычисления" – эволюционный этап развития инфокоммуникационных технологий // Технологии информационного общества : Сборник трудов XII Международной отраслевой научно-технической конференции, Москва, 14-15 марта 2018 года. Том 2. М.: Издательский дом Медиа Паблишер, 2018. С. 96-99. EDN XUPRRB

6. *Калинина М.А., Захарова М.А*. Векторы семантической деривации заимствованной лексики в русском языке // Известия ВГПУ. Самара. 2020. С. 24-37.

7. *Гадасин Д.В., Шведов А.В., Пантелеева К.А.* Предобработка информации для систем машинного обучения // Актуальные проблемы и перспективы развития экономики : Труды XXI Международной научно-практической конференции, Симферополь-Гурзуф, 20-22 октября 2022 года. Симферополь: Крымский федеральный университет им. В.И. Вернадского, 2022. С. 268-269. EDN QVIOMF

8. *Gadasin D.V., Shvedov A.V., Vakurin I.S*. Determination of Semantic Proximity of Natural Language Terms for Subsequent Neural Network Training // 2022 Systems of Signals Generating and Processing in the Field of on Board Communications, SOSG 2022 - Conference Proceedings, Moscow, 15-17 марта 2022 года. Moscow, 2022. DOI 10.1109/IEEECONF53456.2022.9744290. EDN LASMDY

9. *Шведов А.В., Савин В.А., Мартынов М.Д.* Разработка приложения для синтаксического анализа сформированных в базу структурированных данных // Технологии информационного общества : Сборник трудов XVI Международной отраслевой научно-технической конференции, Москва, 02-03 марта 2022 года. М.: Издательский дом Медиа Паблишер, 2022. С. 167-169. EDN AUXGNF

10. *Гадасин Д.В., Шведов А.В., Вакурин И.С*. Определение семантической близости текстов с использованием алгоритма сравнения сущности графов // REDS: Телекоммуникационные устройства и системы. 2022. Т. 12. № 4. С. 11-19. EDN PVJKQJ.

11. *Яковенко Н.В., Шведов А.В., Пантелеева К.А., Гадасин Д.Д.* Средства реализации поисковых и контекстных механизмов для работы с большими данными // REDS: Телекоммуникационные устройства и системы. 2022. Т. 12. № 3. С. 56-63. EDN TBDOXX

12. *Яремко О.Э., Яремко Н.Н., Могилева Е.С.* Кратные ряды Фурье и интегралы Фурье с неразделяющимися переменными // Известия высших учебных заведений. Поволжский регион. Самара, 2020. С. 24-37.

13. *Lane H., Cole H., Hannes H*. Natural Language Processing in Action. Manning, 2019, pp. 44-57.

14. *Кабальнов Ю.С., Максимов С.В., Калентьева М.Б*. Контекстно-словарное сжатие текстовой информации на основе лексических правил // Вестник УГАТУ. Уфа: УГАТУ, 2007. С. 98-105.

15. *Гадасин Д.В*., *Юдина А.А*. Кластеризация в крупномасштабных сетях // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2020. № 1. С. 19-26. EDN OYSXON

16. *Гадасин Д.В., Шведов А.В., Кузин И.А*. Трехмерная реконструкции объекта по одному изображению с использованием глубоких свёрточных нейронных сетей // T-Comm: Телекоммуникации и транспорт. 2022. Т. 16. № 7. С. 29-35. DOI 10.36724/2072-8735-2022-16-7-29-35. EDN YTLCNW

17. *Гадасин Д.В., Смальков Н.А., Кузин И.А.* Использование метода роя частиц для балансировки нагрузки в сетях Интернета вещей // Системы синхронизации, формирования и обработки сигналов. 2022. Т. 13. № 2. С. 17-23. EDN LIUWNT

18. *Гадасин Д.В., Кольцова А.В., Полякова А.Н.* Модель построения кластера для пограничных вычислений // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2020. № 1. С. 86-92.

19. *Shvedov A.V., Gadasin D.V., Alyoshintsev A.V*. Segment routing in data transmission networks // T-Comm. 2022. Vol. 16. No. 5. P. 56-62. DOI 10.36724/2072-8735-2022-16-5-56-62. EDN VAYLJQ

20. *Шведов А.В., Гадасин Д.В., Клыгина О.Г.* Организация взаимодействия туманных вычислений и сегментной маршрутизации для предоставления сервисов IoT в smart grid // Системы синхронизации, формирования и обработки сигналов. 2022. Т. 13. № 3. С. 40-49. EDN TRRYZN

21. *Назаров М.Д., Шведов А.В.* Корреляция атрибутов соглашения об уровне обслуживания с основными параметрами QoS в корпоративных сетях // Телекоммуникации и информационные технологии. 2020. Т. 7. № 2. С. 73-79. EDN VQHDTJ

22. *Kalmykov N.S., Dokuchaev V.A.* Segment routing as a basis for software defined network // T-Comm. 2021. Т. 15. № 7. С. 50-54.

23. *Dokuchaev V.A., Maklachkova V.V., Statev V.Yu.* Classification of personal data security threats in information systems // T-Comm. 2020. Т. 14. № 1. С. 56-60.

24. *Докучаев В.А., Маклачкова В.В., Статьев В.Ю.* Цифровизация субъекта персональных данных // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 6. С. 27-32.

25. *Pavlov S.V., Dokuchaev V.A., Mytenkov S.S*. Model of a fuzzy dynamic decision support system // T-Comm. 2020. Т. 14. № 9. С. 43-47.

26. *Pavlov S.V., Dokuchaev V.A., Maklachkova V.V., Mytenkov S.S*. Features of supporting decision making in modern enterprise infocommunication systems // T-Comm. 2019. Т. 13. № 3. С. 71-74.

27. *Гадасин Д.В., Кольцова А.В., Гадасин Д.Д., Полякова А.Н*. Оценка вероятности формирования виртуального кластера // Системы синхронизации, формирования и обработки сигналов. 2021. Т. 12. № 1. С. 4-12.

28. *Кузин И.А., Гадасин Д.В*. Модель контейнера данных для минимизации трафика при передаче субъективных характеристик объектов на изображении трехмерной сцены // Телекоммуникации и информационные технологии. 2021. Т. 8. № 2. С. 96-100.

29. *Усачева Д.И., Шишкин М.О., Гадасин Д.В., Гузеев А.В.* Применение OLAP-технологий для анализа многомерных данных в контакт-центре // Телекоммуникации и информационные технологии. 2019. Т. 6. № 1. С. 142-149.

30. *Гадасин Д.В., Кузин И.А.* Модель представления цветовых и глубинометрических характеристик объектов на изображении // DSPA: Вопросы применения цифровой обработки сигналов. 2021. Т. 11. № 1. С. 31-38.

31. *Гадасин Д.В., Нестерова Е.А*. Особенности проведения практических занятий по дисциплине мультимедийные информационные системы для стадии "исследование и обоснование создания информационной системы" // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2021. Т. 10. № 1. С. 15-21.

ОЦЕНКА ГОТОВНОСТИ ВЫЯВЛЕНИЯ УГРОЗЫ ПРИ ПРОВЕДЕНИИ ФИШИНГ АТАКИ. РАЗРАБОТКА МЕТОДИКИ ПОВЫШЕНИЯ КАЧЕСТВА ДЕТЕКТИРОВАНИЯ ФИШИНГ ПИСЕМ

Елецкий Андрей Евгеньевич,
Московский Технический Университет Связи и Информатики, Москва, Россия
eletandel@gmail.com

Югай Руслан Сергеевич,
Московский Технический Университет Связи и Информатики, Москва, Россия

Кудряшов Всеволод Владимирович,
Московский Технический Университет Связи и Информатики, аспирант, Москва, Россия
v.v.kudryashov@mtuci.ru

Аннотация

Угроза информационной безопасности - одна из самых обсуждаемых проблем последних лет, каждый день множество людей по всему миру становятся жертвами киберпреступников из-за своего низкого уровня компьютерной грамотности. Исследование посвящено оценке готовности выявления угрозы при проведении фишинг атаки, а также разработке методики, которая способна повысить уровень компьютерной грамотности среди людей. Анкетирование студентов России позволило оценить их способность верно детектировать фишинг письма, понять, как они оценивают свой уровень компьютерной грамотности, уделяют ли они время с целью поднять данный показатель, а также дало возможность оценить эффективность разработанной методики.

Ключевые слова: *Киберпреступления, фишинг, информационная безопасность, компьютерная грамотность, мошенничество, персональные данные.*

Введение

Киберпреступления с каждым годом всё больше набирают популярность. Стремительное развитие Интернет-технологий дало толчок не только развитию электронной коммерции и различным онлайн-сервисам, но и электронному мошенничеству, кибератакам [1]. На 2021 год мировой ущерб от действий киберпреступников оценивается в 6 триллионов долларов, это сопоставимо с ВВП Франции и Италии вместе взятых и по прогнозам эта сумма будет только расти. Из всех способов доставки вредоносных программ до пользователя самым распространенным является доставка через почтовые сервисы, 94% всех вредоносных программ доставляется именно через почту. Также важным является факт, что более 80% всех событий кибербезопасности связано с фишинговыми атаками. В следствие данных фактов было принято решение провести исследование именно на тему фишинговых атак. На данный момент существуют инструменты, предупреждающие пользователя от том, что ссылка, по которой он пытается перейти, введет на поддельный сайт [2]. Но они не гарантируют полную защиту от фишинга из-за их принципа работы. Следовательно, необходимо повышать уровень компьютерной грамотности самих пользователей компьютерными системами.

На сегодняшний день существуют исследования на данную темы, однако в данном исследовании было проведено анкетирование студентов России, также был проведен анализ результатов, разработана методика по повышению уровня компьютерной грамотности, в этом заключается новизна данного исследования

Предположим, что фишинг атаки являются эффективным методом получения конфиденциальной информации в наши дни, из-за низкой “компьютерной грамотности” пользователей компьютерными системами, и множество людей попадаются в руки мошенников, использующих фишинг, в частности студенты. Докажем или опровергнем данную гипотезу

Используемые методы исследования: наблюдение, анализ, поиск информации в глобальной сети интернет, чтение учебной и научно-популярной литературы, анкетирование.

Цели и задачи исследования

Целью исследования является оценка готовности выявления угрозы при проведении фишинг атаки, разработка методики повышения уровня компьютерной грамотности, а также изучение структуры фишингового письма и способов его детектирования.

Для достижения цели исследования определены следующие задачи:

1) Изучение существующих исследований на данную тему
2) Изучение структуры фишингового письма, исследование методов их детектирования
3) Проведение анкетирования среди студентов
4) Анализ результатов анкетирования
5) Разработка и применение методики повышения компьютерной грамотности
6) Проведение повторного анкетирования студентов на которых была применена ранее разработанная методика
7) Анализ результатов анкетирования
8) Подведение итогов

Ключевое понятие и его толкование

Ключевое понятие для нашего исследования – фишинг атака

Фишинг атака – это атака, когда мошенники, используя почтовые сервисы, пытаются скомпрометировать данные пользователей сети интернет

Компания “Лаборатория Касперского” приводит такое толкование этого термина: “ Фишинг (англ. phishing, от *fishing* – рыбная ловля, выуживание и *password* – пароль) – вид интернет-мошенничества, цель которого – получить идентификационные данные пользователей. Сюда относятся кражи паролей, номеров кредитных карт, банковских счетов и другой конфиденциальной информации.

Фишинг представляет собой пришедшие на почту поддельные уведомления от банков, провайдеров, платежных систем и других организаций о том, что по какой-либо причине получателю срочно нужно передать/обновить личные данные. Причины могут называться различные. Это может быть утеря данных, поломка в системе и прочее” [3].

Структура современного фишинг письма

Современные фишинг письма достаточно тяжело распознать неопытному пользователю, так как мошенники научились мимикрировать их под настоящие. Способы, которые используют мошенники, чтобы сделать фишинг письмо похожим на настоящее:

1) Использование похожего адреса электронной почты отправителя. Для этого мошенники могут зарегистрировать домен, который очень похож на официальный домен какой-либо компании. К примеру есть домен mtuci (...@mtuci.ru), мошенник регистрирует домен mtucl (...@mtucl.ru) и использует его, чтобы отправлять письма от лица университета неопытным пользователям. Также есть множество сервисов, как платных, так и бесплатных, которые позволяют отправить письмо под видом любого адреса электронной почты, в данном случае человек никак не отличит настоящий адрес от поддельного
2) Копирование дизайна оригинального письма. Мошенники полностью копируют дизайн оригинального письма или же используют похожий стиль при создании своего.
3) Создание похожего URL адреса на сайт или документ. Мошенник создает сайт с похожим URL адресом и вставляет эту ссылку в своё письмо. К примеру есть сайт https://mtuci.ru, мошенник создает сайт https://mtucl.ru, неопытный пользователей не заметит отличий и перейдёт на этот него, после чего запустится скачивание вируса, или же мошенник попытается мимикрировать свой сайт под оригинальный сайт университета МТУСИ и пользователь оставит там свой логин и пароль.

Исследования на тему актуальности применения фишинг атак

Рассмотрим исследование, проведенное компанией Group-IB. В своей работе они проанализировали российский сегмент интернета на предмет фишинга. По результатам их анализа за первые девять месяцев 2022 года в российском сегменте интернета было зафиксировано около 18 тыс. мошеннических фишинговых сайтов.

Для сравнения: за аналогичный период прошлого года компании удалось выявить лишь 15 тыс. фишинговых доменов в зонах .ru и .рф. Эксперты Group-IB также заявили, что рост будет продолжаться и далее [4].

«Фишинг остается наиболее массовой угрозой для пользователей в интернете, и его масштабы неуклонно растут. Именно фишинговые сайты составляют 98-99% заблокированных ресурсов

киберпреступников. Остальной процент от общего числа заблокированных страниц составляют сайты с вредоносным ПО», – заявил руководитель группы по защите от фишинга CERT-GIB Иван Лебедев [5].

Intel Security также провела своё исследование и опубликовала его в 2015 году [6]. Компания провела тест проверку знаний пользователей и их умения распознавать электронные письма, отправленные мошенниками с целью получения доступа к логинам, паролям и другим конфиденциальным данным.

В исследовании приняли участие около 19 000 человек из 144 стран. Им было предложено изучить 10 сообщений, специально подготовленных Intel Security. Некоторые образцы содержали угрозы кражи информации, т.е. фишинговые атаки. Только 3% из всех опрошенных смогли точно определить, можно ли доверять тому или иному посланию, тогда как 80% респондентов посчитали безопасным как минимум одно из писем с угрозой.

Проведение исследования

Для оценки готовности выявления угрозы при проведении фишинг атаки было принято решение провести анкетирование. За основу была взята методика исследования, разработанная компанией Intel Security. Причиной послужили:

1) Простота, их методика легка в реализации
2) Репрезентативность, данную методику можно применить на любую группу людей
3) Эффективность, данная методика показала отличные результаты в исследованиях Intel Security

Методика заключается в анкетировании группы лиц с целью проверки знаний пользователей и их умения распознавать фишинг письма.

Для проведения анкетирования было создано 6 электронных писем. Три письма являлись копией настоящих писем. Причины, почему мы выбрали именно эти письма:

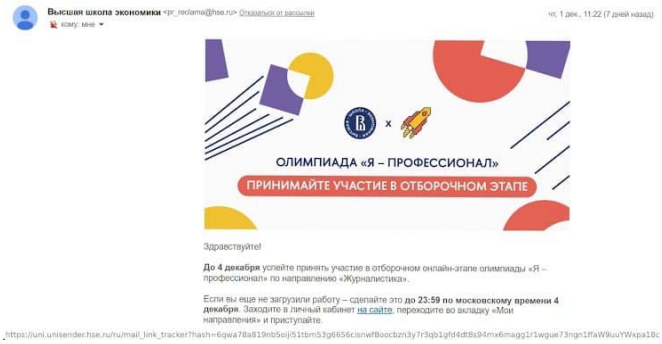

Рис. 1. электронное письмо из анкеты

У данного письма, представленного на рисунке 1, необычный почтовый адрес отправителя, у некоторых пользователей он может вызвать подозрения. Проверить достоверность данного письма можно в сети интернет через домен из адреса электронного письма или ссылки, изучив, где он зарегистрирован и кем.

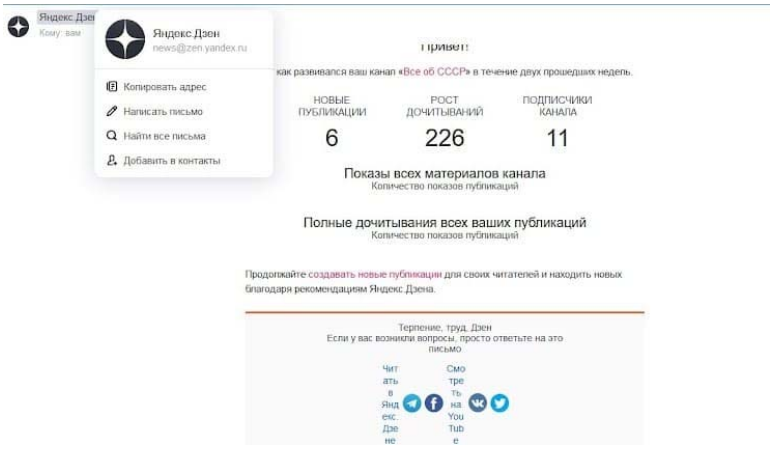

Рис. 2. электронное письмо из анкеты

Письмо, представленное на рисунке 2, сделано не очень качественно, у многих может сложиться впечатление, что Яндекс не мог создать такое письмо. Проверить достоверность данного письма можно в сети интернет через домен из адреса электронного письма, изучив, где он зарегистрирован и кем.

Рис. 3. Электронное письмо из анкеты

Странный дизайн у письма, представленного на рисунке 3, ссылка в данном письме может вызвать подозрение у пользователей. Проверить достоверность данного письма можно в сети интернет через домен из адреса электронного письма или ссылки, изучив, где он зарегистрирован и кем.

Также было создано три фишинговых письма. За основы были взяты настоящие письма, но были изменены адреса электронной почты отправителя, а также ссылки в данных письмах вели на совсем другие ресурсы:

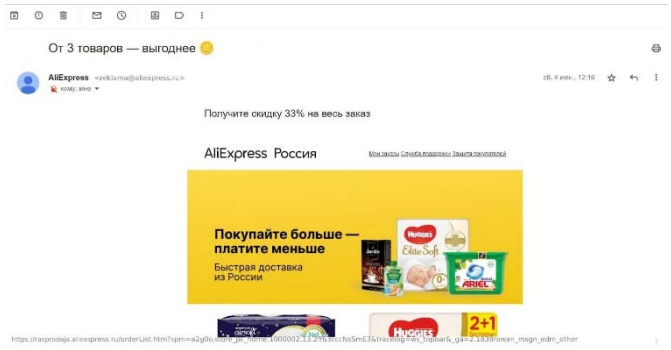

Рис. 4. Электронное письмо из анкеты

Ссылка в письме, представленном на рисунке 4, ведёт на совершенно другой ресурс, что заметно даже по виду ссылки.

Рис. 5. Электронное письмо из анкеты

Письмо, представленное на рисунке 5, OZON использует почту с корпоративным доменом, а не почту с доменом mail. Также ссылки в письме ведут на совершенно другой ресурс, что заметно даже по виду ссылки.

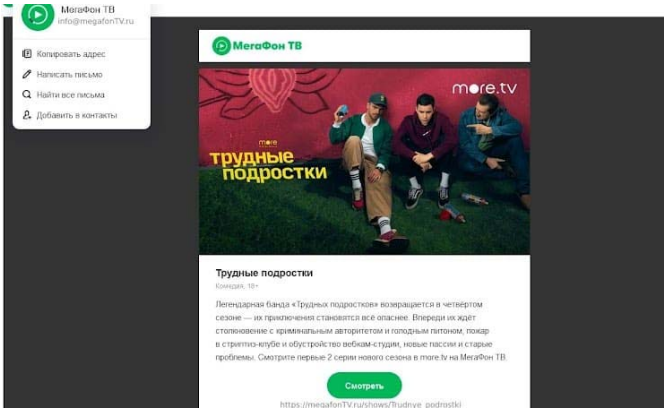

Рис. 6. Электронное письмо из анкеты

Письмо, представленное на рис. 6, МегаФон использует другой корпоративной домен, ссылка ведёт на совершенно другой ресурс.

Анкетирование было решено провести дистанционно для студентов. Дистанционный формат анкетирования был выбран для того, чтобы опросить большое количество студентов в короткие сроки.

Анкетирование представляло из себя 9 вопросов:

1-6) Определите тип письма: 1. Настоящее 2. Фишинговое

7) Попадались ли вы когда-нибудь на фишинг атаку?: 1.Да 2.Нет

8) Оцените ваш уровень компьютерной грамотности: 1.Ужасный 2.Ниже среднего 3.Средний 4.Выше среднего 5.Отличный

9) Проходили ли вы курсы компьютерной грамотности ? 1.Да 2.Нет

За каждый правильный ответ на вопросы с 1 по 6 начислялся один балл.

Проведение анкетирования среди студентов

Анкетирование проводилось в две группы среди студентов различных вузов, направлений, курсов и возрастов. Первая группа состояла из студентов добровольцев, которые согласились протестировать на себе методику повышения уровня компьютерной грамотности, количество участников - 15 человек. Вторая группа состояла из студентов, которые не изъявили желания протестировать на себе методику повышения компьютерной грамотности, количество участников – 208 человек.

Анализ результатов анкетирования

Рис. 7. График зависимости среднего балла за тест от количества человек

Выборка дошла до релевантных значений (рис. 7).

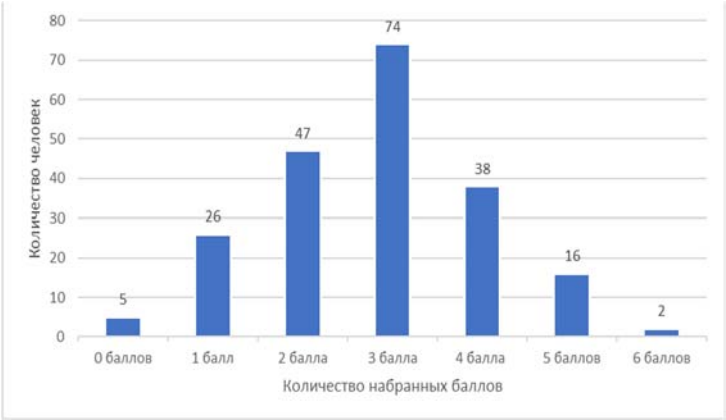

Рис. 8. Количество баллов, набранных респондентами

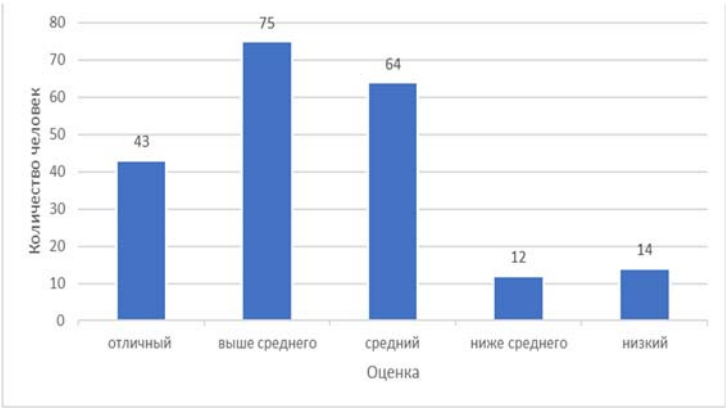

Рис. 9. Как анкетируемые оценивают свой уровень компьютерной грамотности

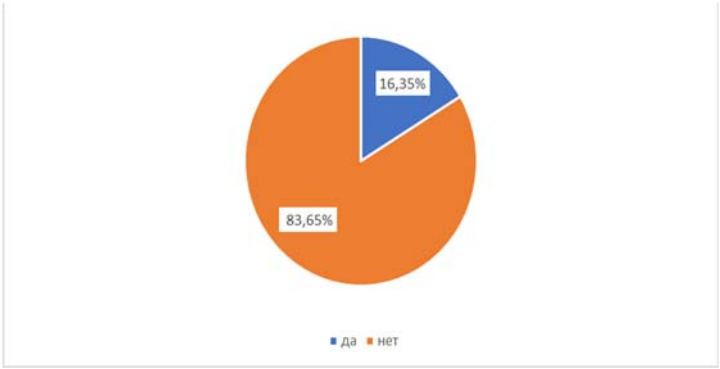

Рис. 10. Проходили ли респонденты курсы компьютерной грамотности

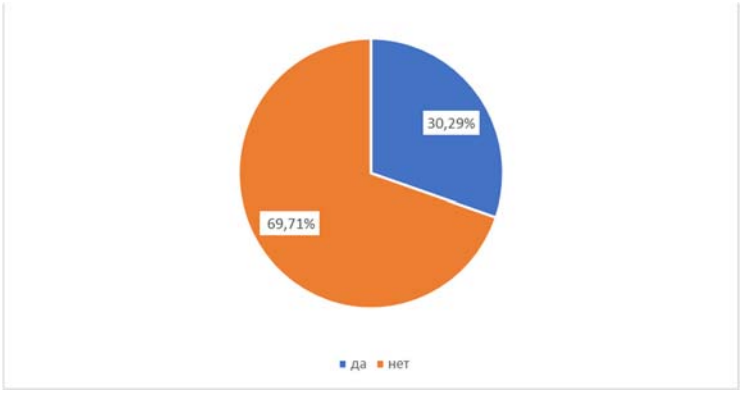

Рис. 11. Попадались ли анкетируемые на фишинг атаку

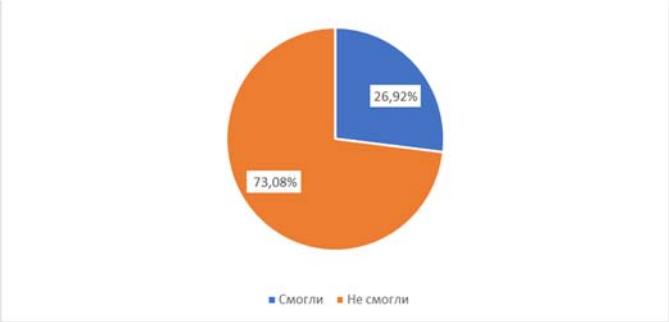

Рис. 12. Количество участников, которые смогли верно детектировать больше половины писем

Можно заметить, что большинство респондентов набрали меньше 4 баллов за тест (рис. 8), следовательно смогли верно детектировать меньше половины писем, что наглядно видно из рисунка 12, представленных в анкете, при этом оценивая свой уровень компьютерной грамотности, как средний и выше среднего (рис. 9), поисковая информационная деятельность отсутствует. 73% студентов не смогли верно детектировать и половину фишинг писем. При этом курсы по повышению уровня компьютерной грамотности проходили всего 16% анкетируемых (рис. 10), а попадались на фишинг атаку целых 30% (рис. 11).

Данные результаты наглядно показывают необходимость разработки и внедрения методики повышения уровня компьютерной грамотности и качества детектирования фишинг писем.

Разработка методики повышения уровня компьютерной грамотности

С целью повысить качество детектирования фишинг писем было принято решение разработать методику. При разработке методики ставились следующие задачи:

1) Сделать её в формате видеоурока, чтобы студенты могли ознакомиться с ней в любое удобное им время
2) Сократить время, которое требуется для ознакомления, чтобы в короткие сроки студенты получили максимум полезной информации.

Видеоурок состоял из нескольких частей:

1) Что такое фишинг, что к нему относится. В данной части изложены основы, которые нужны для общего понимания, что же такое фишинг и фишинг письмо, в частности.
2) Примеры качественного детектирования фишинг писем на примере писем из анкеты. В данной части изложено подробное объяснение, как можно было грамотно обнаружить, что письмо из анкетирования было фишинговым
3) Как защитить себя от фишинг мошенников, что делать в случае, если вы попались на фишинг атаку. Краткое подведение итогов с целью закрепить изложенный материал

Видеоурок “Методика детектирования фишинг писем” https://drive.google.com/file/d/1f0tg_bTuuDg-L0SiKharnIAKbVS7pxEy/view?usp=sharing

Проведение повторного анкетирования

Для проведения повторного анкетирования было разработано 6 новых электронных письма. 3 письма являлись копией настоящих писем. Причины, почему мы выбрали именно эти письма:

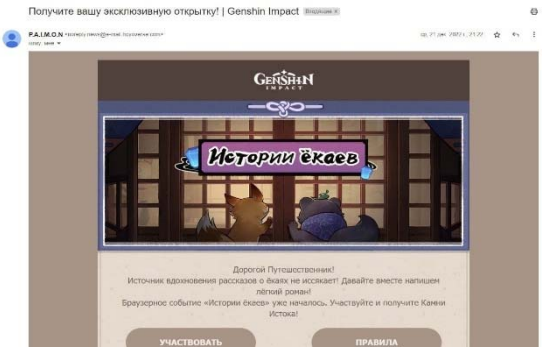

Рис. 13. Электронное письмо из анкеты

У данного письма, представленного на рисунке 13, необычный почтовый адрес отправителя, у некоторых пользователей он может вызвать подозрения. Проверить достоверность данного письма можно в сети интернет через домен из адреса электронного письма, изучив, где он зарегистрирован и кем.

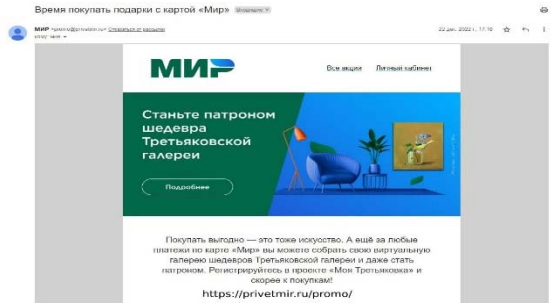

Рис. 14. Электронное письмо из анкеты

Ссылка в данном письме, представленном на рисунке 14, может вызвать подозрение у пользователей. Проверить достоверность данного письма можно в сети интернет через домен из адреса электронного письма или ссылки, изучив, где он зарегистрирован и кем.

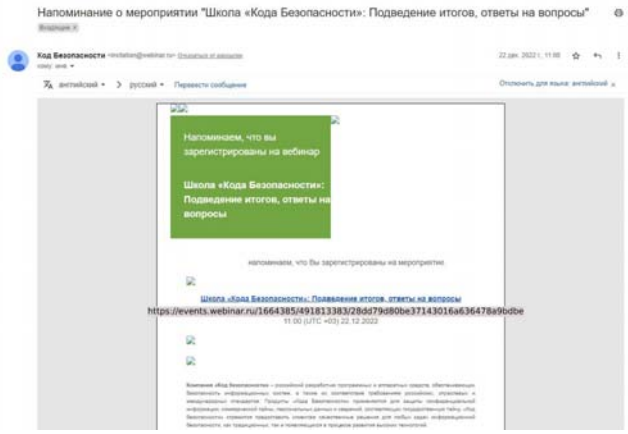

Рис. 15. Электронное письмо из анкеты

Странный дизайн письма, представленного на рисунке 15, ссылка в данном письме может вызвать подозрение у пользователей. Проверить достоверность данного письма можно в сети интернет через домен из адреса электронного письма или ссылки, изучив, где он зарегистрирован и кем.

Также было создано 3 фишинговых письма. За основы были взяты настоящие письма, но были изменены адреса электронной почты отправителя, а также ссылки в данных письмах вели на совсем другие ресурсы.

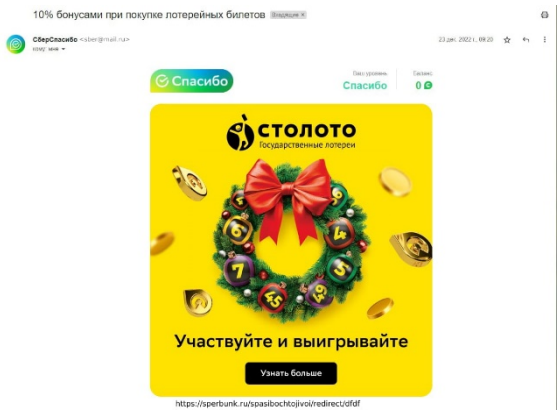

Рис. 16. Электронное письмо из анкеты

Письмо, представленное на рисунке 16, Сбер использует почту с корпоративным доменом, а не почту с доменом mail. Также ссылки в письме ведут на совершенно другой ресурс, что заметно даже по виду ссылки

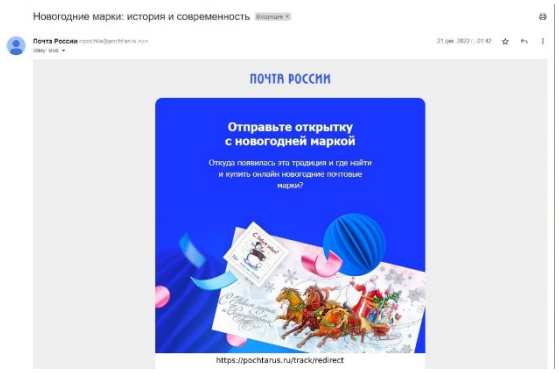

Рис. 17. Электронное письмо из анкеты

Письмо, представленное на рисунке 17, Почта России использует другой корпоративной домен, ссылка ведёт на совершенно другой ресурс.

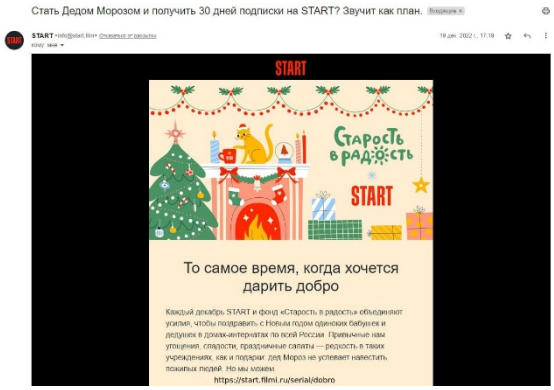

Рис. 18. Электронное письмо из анкеты

Ссылка в письме, представленном на рисунке 18, ведет на совершенно другой ресурс, что заметно даже по виду ссылки.

Вопросы в новой анкете соответствовали по сложности вопросам из предыдущей анкеты. Анкетирование было решено провести дистанционно для студентов. Дистанционный формат анкетирования был выбран для того, чтобы опросить большое количество студентов в короткие сроки.

Анкетирование представляло из себя 6 вопросов:

1-6) Определите тип письма: 1. Настоящее 2. Фишинговое

За каждый верный ответ на вопрос начислялся 1 балл.

Анализ результатов анкетирования после применения разработанной методики повышения качества детектирования фишинг писем

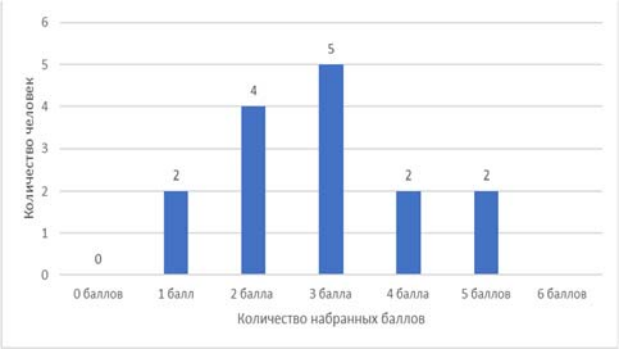

Рис. 19. Количество набранных баллов респондентами до применения методики

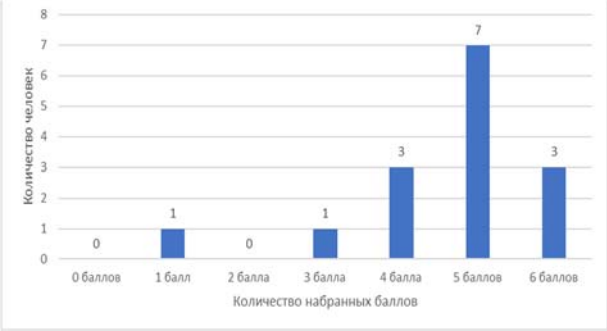

Рис. 20. Количество набранных баллов респондентами после применения методики

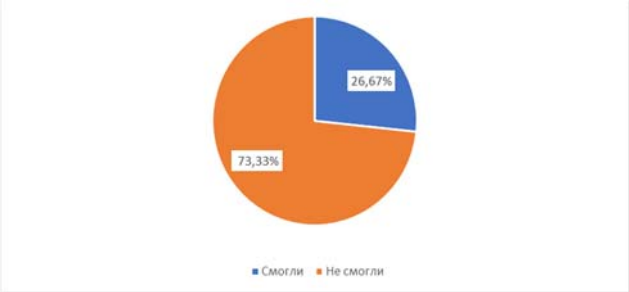

Рис. 21. Количество респондентов, которые смогли верно детектировать больше половины писем до применения методики

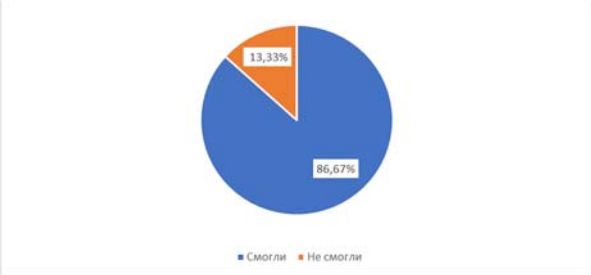

Рис. 22. Количество респондентов, которые смогли верно детектировать больше половины писем после применения методики

Можно заметить существенный прирост качества детектирования фишинг писем среди респондентов (рис. 19, 20). Средний балл за тест вырос с 2,87 до 4,6 прирост на 60,28%. Количество респондентов, которые смогли верно детектировать больше половины писем увеличилось с 26,67% до 86,67% (рис. 21, 22). Методика доказала свою эффективность.

Заключение

Результаты проведенного исследования позволяют сделать выводы:

1. До применения методики студенты могли верно детектировать только 50% фишинг писем, после применения результат составил 80%+-10%. Очевидно, что при таком уровне компьютерной грамотности, какой был до применения методики, мошенникам не составит труда заполучить конфиденциальную информацию путем фишинг атаки. Это ставит под угрозу информационную безопасность нашей страны, ведь студенты – это будущие директора и сотрудники как в частных компаниях, так и в государственных организациях.
2. Респонденты преувеличивают свой уровень компьютерной грамотности, большинство студентов не проходят курсы повышения компьютерной грамотности. Данная самоуверенность в собственных силах идёт на руку мошенникам и из-за этого студенты активно попадаются на фишинг атаки.

3. Разработанная методика повышения качества детектирования фишинг писем позволила улучшить показатели студентов в тесте на определение фишинг писем. Наблюдается крайне высокая степень освоения материала и способность применить полученные знания на практике

Наша гипотеза подтвердилась, фишинг атаки являются эффективным методом получения конфиденциальной информации в наши дни, из-за низкой “компьютерной грамотности” пользователей компьютерными системами, и множество людей попадаются в руки мошенников, использующих фишинг, в частности студенты. Представители опрашиваемых нами фокус-групп не умеют качественно детектировать фишинг письма. Наше исследование свидетельствует о том, что сегодня крайне необходимо усилить борьбу с фишинг мошенниками. Студенты способны в короткие сроки научиться качественно детектировать фишинг письма на практике. Следовательно, необходимо внедрять методики повышения уровня компьютерной грамотности в вузах. Информировать граждан об опасности фишинг атак. Это позволит уменьшить угрозу информационной безопасности нашей стране. Данное исследование поможет развитию платформ security awareness, которые помогают с обучением сотрудников компаний.

Литература

1. *Гуськова А.М*. Фишинг как основной метод социальной инженерии в схемах финансового мошенничества // Сборник докладов международной научной конференции «Исследования молодых учёных». Казань. 2019. С. 3-6.

2. *Мартынюк Р.А*. Механизм распознавания фишинговых сайтов по косвенным признакам // Международный научный журнал «Молодой учёный». 2020. С. 19-22.

3. Лаборатория Касперского// Что такое “фишинг”//текст -электронный/url: https://encyclopedia.kaspersky.ru/knowledge/what-is-phishing/

4. Group-IB// Фиши прилетели: Group-IB выявила рекордное число мошеннических ресурсов в Рунете в 2022 году/текст - электронный//url: https://www.group-ib.ru/media-center/press-releases/phishing-runet-2022

5. ТАСС//Число фишинговых сайтов в РФ за девять месяцев 2022 года выросло на 15%/ текст - электронный// url: https://tass.ru/ekonomika/16298647

6. security affairs/ New Intel Security study shows that 97% of people can’t identify phishing emails//текст - электронный// url: https://securityaffairs.co/wordpress/36922/cyber-crime/study-phishing-emails-response.html

АНАЛИЗ РАБОТЫ ПРОТОКОЛА MQTT

Кириллов Кирилл Алексеевич,
Московский Технический Университет Связи и Информатики, Москва, Россия
deadvil_kirill@mail.ru

Маликова Елена Егоровна,
Московский Технический Университет Связи и Информатики, доцет каф. СС и СК, к.т.н., Москва, Россия
emalikova@gmail.com

Аннотация
В данной статье рассматриваются протоколы передачи данных для устройств Интернета вещей (IoT), углублено изучена архитектура протокола MQTT и продемонстрирована реализация протокола в программном обеспечении Mosquitto. Также в среде имитационного моделирования AnyLogic спроектирована модель обслуживания заявок, поступающих от датчиков на сервер.

Ключевые слова: *Брокер, клиент, подписка, протокол MQTT, протокол CoAP, Протокол HTTP/2, качество обслуживания (QoS), IoT, программа Mosquitto.*

Введение

В настоящее время все чаще пользователи используют технологию Интернет Вещей [1], будь то промышленная отрасль, продукты питания, умные дома и др. Множество данных от различных устройств поступает на облачные сервера, используя протоколы IoT [8-11]. Основными можно считать протоколы MQTT, CoAP, HTTP/2, которые собирают и передают данные между устройствами и серверами. Каждый протокол имеет свои преимущества и недостатки. В статье рассмотрены такие протоколы как HTTP/2, CoAP, MQTT, продемонстрирована реализация работы протокола MQTT в программном обеспечении Mosquitto. Также в работе использовался инструмент имитационного моделирования Anylogic. В данном программном обеспечении спроектирована модель обслуживания заявок, поступающих от датчиков на сервер для их последующей обработки.

Сравнение протоколов передачи данных для устройств IoT

Для обеспечения успешной работы приложений IoT необходимо обеспечить работу протоколов маршрутизации. В настоящее время используются большое количество протоколов обмена данными. Рассмотрим самые распространенные протоколы IoT.

Протокол HTTP/2 (*HyperText Transfer Protocol v.2.0)* – обновленная версия протокола HTTP [2]. Данный протокол изначально предназначен для выхода в глобальную сеть, взаимодействие между датчиками скорее реализовано как дополнительный функционал, и в отличии от других протоколов, специализирующихся для устройств IoT, потребляет много оперативной памяти, больше энергии и буферного пространства.

Протокол CoAP (Constrained Application Protocol) разработан на основе HTTP, работает на прикладном уровне. Передача данных происходит благодаря протоколу UDP, что позволяет передать небольшой размер служебных данных. Протокол предназначен для взаимосвязи двух узлов между собой, основываясь на клиент-серверной архитектуре [3]. CoAP удобен для получения малых данных с датчиков, но уступает MQTT в надежности, так как у MQTT имеются 3 уровня качества обслуживания.

Message Queue Telemetry Transport (MQTT) – предназначенный для передачи сообщений протокол с шаблоном передачи сообщений типа издатель/подписчик, при этом протоколом определяется формат сообщения, но не его содержимое, поэтому в системе, где используется данный протокол, очень важно чтобы и издатель и подписчик знали что слушать и что публиковать [4]. MQTT протокол рассчитан на простоту в использовании, небольшую нагрузку на каналы связи, работу в условиях постоянной потери связи или «легкую» встраиваемость в любую систему.

Протокол используется для работы с телеметрией, работы различных датчиков, для обмена данными между устройствами IoT. Протокол функционирует на стеке TCP/IP, но существуют и версии протокола для работы с другими стеками протоколов. MQTT разработан для устройств с ограниченными ресурсами и сетей с ограниченной пропускной способностью, например, с микропроцессорной системой.

Основная цель при создании протокола – обеспечение сбора данных с большого количества устройств для последующей передачи данных в ИТ-системы. Протокол имеет небольшую транспортную часть пакета (минимум 2 байта) и минимизированную часть формата посылки для данных самого протокола для снижения нагрузки на сеть.

Протокол имеет различные спецификации, такие как:

- MQTT v.3.1.1 – текущая версия протокола;
- MQTT-SN v.1.2 – более легкая версия протокола для систем датчиков, не поддерживающих стек TCP/IP, например для стандарта ZigBee.

Если говорить о надёжности протоколов CoAP и MQTT, то для критических коммуникаций лучше выбирать MQTT, так как он обеспечивает качество обслуживания (QoS), гарантирующее доставку и хранение сообщения на брокере. С другой стороны, CoAP можно использовать для сбора данных с датчиков, а QoS может быть реализован на стороне приложения.

Рассмотрим более подробно протокол MQTT.

Архитектура протокола MQTT

MQTT является протоколом с шаблонной передачей данных типа издатель/подписчик и имеет два типа подключаемых устройств. Первым является Клиент (Client), который является либо издателем, публикующим данные в сеть, либо подписчиком, получающим данные с сервера. Один и тот же Клиент может быть как издателем, так и подписчиком.

Функции Client:

- установление соединения с сервером;
- публикация сообщений или подписка на сообщения других клиентов.

Центром данного обмена сообщений является Брокер (Broker). Клиенты обмениваются данными не на прямую, а через Брокера. Благодаря данной концепции, клиенты не могут знать о существовании других клиентов, и издатель не может определить, подписан ли кто-то на его рассылку сообщений или нет. Брокер обрабатывает все запросы на подписку/отписку от клиент-устройств (рис. 1).

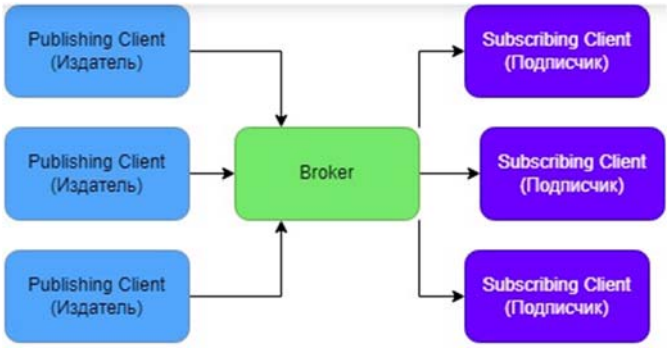

Рис. 1. Функции брокера

Функции Broker:

- является посредником между клиентами;
- допускает соединения от клиентов;
- получает сообщения, опубликованные клиентами;
- обработка запросов на подписку и отказ от нее;
- перенаправляет далее сообщения, которые совпадают с подписками клиента.

Прокол MQTT поддерживает только 14 типов пакетов сообщений определенных спецификаций. Самыми популярными являются запросы типа CONNECT, PUBLISH, SUBSCRIBE, DISCONNECT. Публикуемые клиентом данные идентифицируются при помощи специальных уникальных строк – topic. Брокер использует строку «topic» для рассылки сообщений подписчикам [5].

Полезные данные пакета сообщений не определены в спецификации MQTT, следовательно, данные могут быть представлены в любом стандартизированном виде, например, JSON, CSV, XML. Формат кодировки полностью определяется приложением.

На рисунке 2 можно увидеть, как издатель отправляет сообщение брокеру с определенным топиком, после этого брокер выполняет рассылку этого сообщения всем клиентам, подписанным на данный топик (рис. 2).

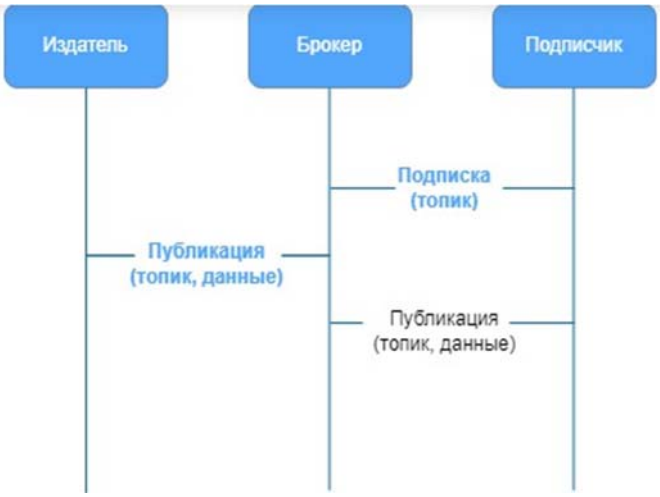

Рис. 2. Архитектура Издатель/Подписчик

Протокол является надежным. Ниже приведены функции протокола MQTT:

- протокол использует механизм QoS (Quality of Service), который позволяет определять, доставлен ли пакет по назначении;
- Last Will and Testament – это сообщение, которое будет отправлено брокером от имени клиента в случае потери связи с последним;
- Keep Alive Time – сообщение, которое служит для определения наличия клиента в сети. Отправляется брокеру с заданным промежутком времени в тех случаях, когда отправка других сообщений не выполняется, чтобы оповестить брокера, что клиент все еще в сети.
- Store Session State – данная функция позволяет сохранять всю информацию о сессии;
- Clean Session – при значении «false» брокер сохранит последнюю информацию, например, о подписках, недоставленных сообщениях в случае потери связи с клиентом. При значении «true» данные не будут сохранены.

В плане обеспечения функции безопасности в протоколе имеются следующие возможности:

- имеется возможность аутентификации по логину и паролю при подключении клиента к брокеру.
- MQTT не поддерживает механизмы шифрования, при этом рекомендуется:
  - использовать дополнительные способы шифрования полезных данных;
  - использовать TLS (Transport Layer Security) для установки зашифрованного соединения по TCP/IP.

MQTT поддерживает три уровня качества обслуживания сообщения (QoS), которые влияют на надежность доставки сообщений.

QoS 0 – отправка сообщений с максимальной производительностью. При этом нет проверки успешности доставки сообщения. В данном случае некоторые сообщения могут быть потеряны. Например, при использовании датчика температуры с циклическим обновлением данных, брокер все равно получит какое-либо из показаний. Публикации сообщений однонаправленные, что позволяет добиться значительного роста производительности (рис. 3).

Рис. 3. Уровень надежности QoS 0

QoS 1 – Гарантированная доставка. При использовании этого уровня качества гарантируется доставка сообщения брокеру, при этом возможно появление дубликатов. Получив дубликат, брокер отправляет снова сообщение получателю и сообщение о получении отправителю (рис.4).

Рис. 4. Уровень надежности QoS 1

QoS 2 – Гарантированная доставка одного сообщения. При использовании этого уровня качества получателю будет доставлено одно и только одно сообщение, исключается возможность появления дубликатов. Достаточно сложный метод в реализации, оказывающий негативное влияние на производительность, ввиду использования большого количества пакетов для доставки одного сообщения. (рис. 5).

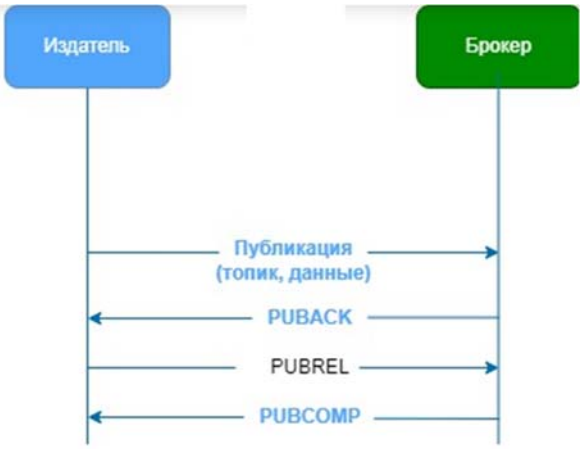

Рис. 5. Уровень надежности QoS 2

Структура пакета сообщения протокола MQTT

Архитектура пакета протокола MQTT представляет собой фиксированный заголовок длинной два байта. Первый байт содержит информацию о топике пакета сообщения и флаги, которые могут меняться в зависимости от типа сообщения. Второй байт состоит из оставшихся данных. Некоторые пакеты сообщений могут содержать переменный заголовок и полезные данные. В переменном заголовке помещаются такие данные, как идентификатор пакета, имя протокола и т.д. (рис. 6-7).

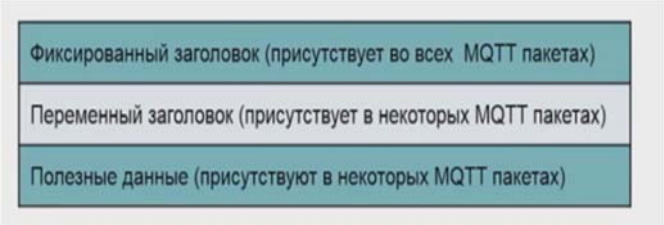

Рис. 6. Структура пакета сообщения

Бит	3	2	1	0
Байт 1	Флаги, специфичные для каждого типа пакета MQTT (QoS, DUP, Retain)			
Байт 2	Оставшаяся длина данных			

Бит	7	6	5	4
Байт 1	Тип пакета MQTT(например: CONNECT, SUBSCRIBE, PUBLISH и другие)			
Байт 2	Оставшаяся длина данных			

Рис. 7. Архитектура пакета

Сообщение содержит ID клиента, имя пользователей, пароль для подключения к брокеру, параметры Last Will, параметр Keep Alive.

Примеры применения протокола MQTT

Рассмотрим примеры применения протокола MQTT:

- для обмена некритичными по времени данными между производственными объектами и облачными сервисами корпоративных сетей;
- для сбора диагностических данных сетевых ресурсов для анализа использования и планирования своевременного обслуживания;
- для передачи и хранения данных о качестве процессов производственной линии (например, данные, связанные с качеством позиционирования при сварке, уровнем заполнения бутылки, степенью затяжки винта и т.д.);
- передачи данных производственного анализа диагностики и качества в единой стандартизированной системе;
- для согласованности данных многие параметры могут быть отправлены в одном PUSBLISH пакете;
- часто отправка данных осуществляется только в одном направлении.

Различные устройства, являющиеся MQTT клиентами, отправляют свои сообщения брокеру. Брокер располагается в облачном сервисе корпоративной сети, откуда он отправляет полученные с производственного участка данные в приложение на мобильном устройстве (рис. 8) [5].

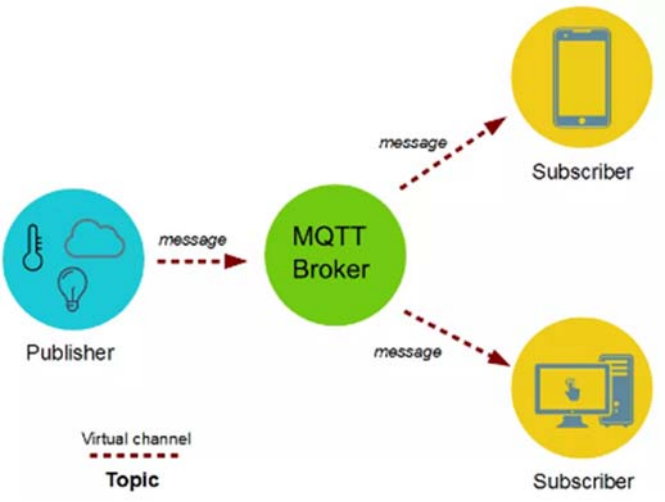

Рис. 8. Схематический вид применения

Реализация работы протокола MQTT с помощью программы Mosquitto

Для углубленного изучения протокола MQTT реализуем его работу, используя программу Mosquitto, которая позволяет работать как в режиме клиента, так и брокера (сервера) [6]. Данная программа предназначена для GPS-слежения, датчиков окружения. Мы установим соединение между брокером на одной рабочей станцией и клиентом на другой (рис. 9).

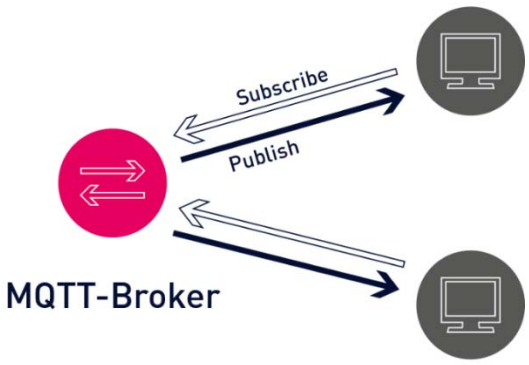

Рис. 9. Схема будущего подключения

После установки брокера, запускаем Mosquitto.exe, который внешне представляет собой командную строку. Откроем вторую консоль и проверим работу брокера. В первом окне мы подпишемся на топик “topic1” (рис. 10) и со второго окна будем передавать сообщения на первый (рис. 11).

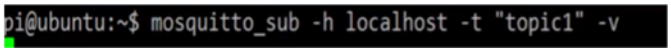

Рис. 10. Подписка на топик

Рис. 11. Отправка сообщения подписчикам

Из рисунка 12 видно, что сообщение со второй консоли поступило на первую.

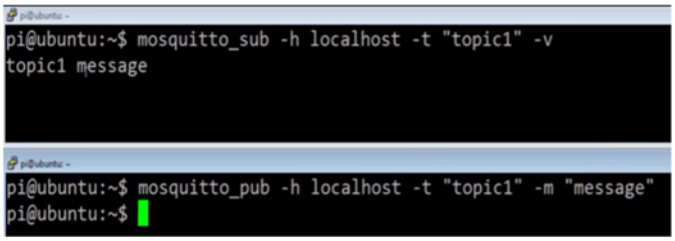

Рис. 12. Работа протокола MQTT

Однако, это работает, потому что наша машина является и клиентом, и брокером. Следующий шагом запустим клиент со стороннего приложения, а именно с расширения для браузера Google Chrome – MQTTLens.

Для настройки клиента нужно заполнить:

- Connection name – уникальное имя клиента для идентефикации в сети;
- Hostname – адрес подключения к брокеру. Указывается ip адрес брокера устройства. Если брокер и клиент развертываются на одном устройстве, указываем localhost.

При неоходимости заполняем дополнительные поля, например пароль. В ином случае, брокер будет принимать все отсылаемые ему сообщения (рис. 13).

Рис. 13. Страница настройки клиента

Так как мы в предыдущем пункте ввели IP-адрес нашей машины, то можно теперь написать топик в брокере и получить сообщение в клиенте. Создадим топик «/releу1/» и подпишемся на него (рис. 14). Теперь ждем сообщение от брокера.

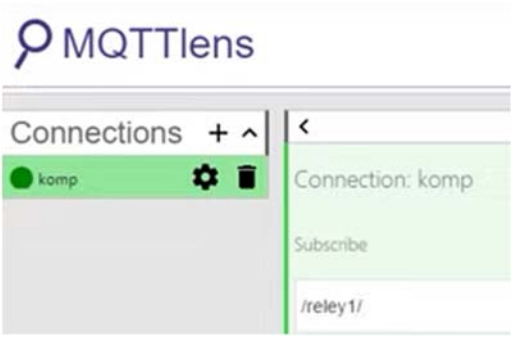

Рис. 14. Подписка на топик из клиента MQTTlens

Как видно на рисунке 15 клиенту пришли два сообщения Message 0 и Message 1, из чего следует, что была установлена связь между клиентом и брокером.

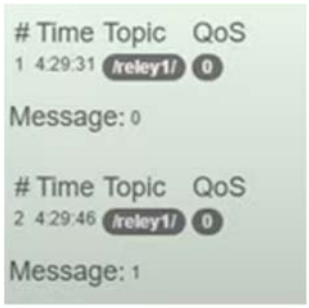

Рис. 15. Полученное сообщение от брокера

Соответственно, если в устройстве запрограммировано действие командой /releу1/, например, переключение, то на всех устройствах, подписанных на топ, будет совершено переключение. Данная технология сильно упрощает настройку и управление различных датчиков, так как можно подписать их на определенные топики и посылать каждой отдельной группе сообщения – команды.

Моделирование нагрузки на сервер в среде AnyLogic

Anylogic – программа для имитационного моделирования [7]. В данном программном обеспечении будет спроектирована модель обслуживания заявок поступающих сообщений от датчиков на сервер и будут проанализированы результаты моделирования.

Смоделируем ситуацию, где от датчиков на нескольких газопроводов поступают сообщения диспетчеру об утечке газа. Сигнал от группы датчиков с помощью протокола MQTT передается по каналу связи с определенной пропускной способностью, далее поступает на контроллер. Контроллер передает сообщение на общий канал, который доставляет данные прямо на центральный сервер и соответственно диспетчеру.

Структурная схема этого процесса представлена на рисунке 16.

Логическая схема данной сети приведена на рисунке 17. В качестве источника (source) выступают сообщения, которые отправляются датчиками.

Вторым элементом является queue, который соответствует максимальной ширине канала связи. Данный элемент моделирует очередь заявок на обработку.

Третий объект delay – моделирует время обработки сервером запроса.

Последний объект sink – выводит обработанные заявки из модели. В случае занятости канала связи, заявка не поступает на обработку и удаляется.

Различные датчики (source1- 4) отправляют пакеты по каналам связи, если канал свободен (queu1, queue2), затем передаются на контроллер домена (delay1,2), который обрабатывает поступившие заявки и перенаправляет их на центральный канал связи (queue3), доставляющий сообщения на центральный сервер и, соответственно, диспетчеру(delay3).

Смоделируем ситуацию, в которой имеется четыре группы датчиков (от четырех газопроводов), в каждой из которых находится четыре датчика. Они передают сообщения об утечке газа, отсылая по одному сообщению в секунду каждый в течении пяти минут. Пропускная способность составляет три заявки в секунду, ширина канала обслуживания 500 заявок.

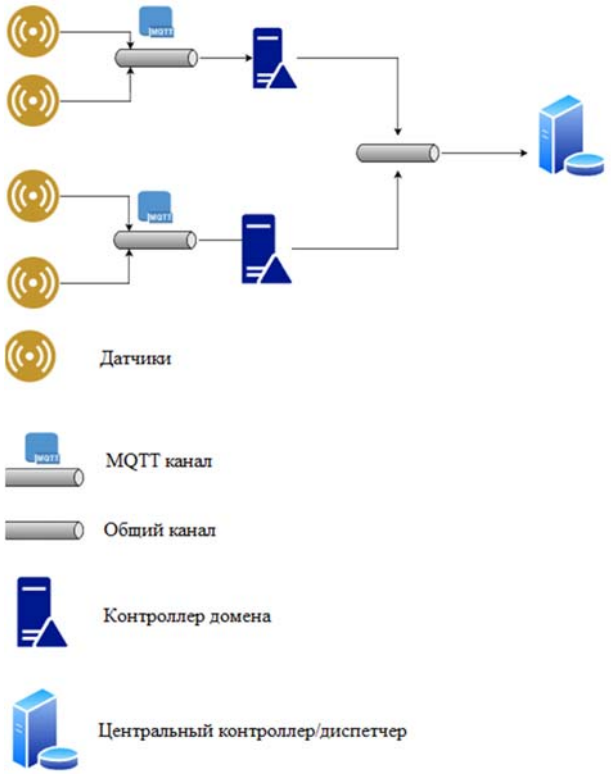

Рис. 16. Структурная схема сети

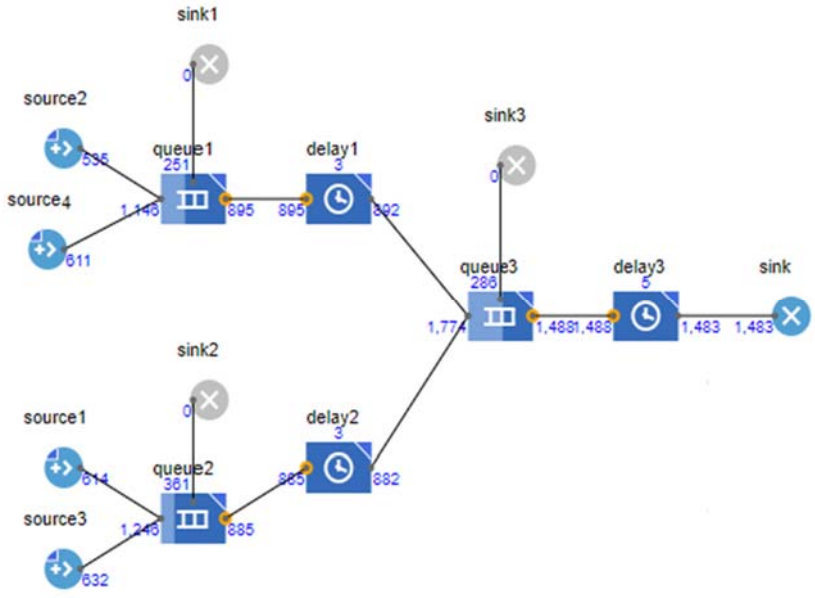

Рис. 17. Логическая схема сети

На рисунке 18 представлены результаты моделирования. На оси абсцисс показано время в секундах, а по оси ординат общее количество заявок.

Результаты моделирования показывают, что при вместимости очереди (блоки queue) равным 500 заявок и обработки запросов сервером (delay) равным три заявки в секунду и число необслуженных заявок сводится к минимуму.

Рис. 18. Результаты моделирования

Данные результаты демонстрирует, что все заявки были успешно обработаны, соответственно все газовые утечки были обнаружены, что способствует увеличению безопасности на объекте.

Заключение

В данной статье были рассмотрены протоколы передачи данных для устройств IoT, углублено изучена архитектура протокола MQTT и продемонстрирована реализация протокола в программном обеспечении Mosquitto. Также продемонстрировано моделирование внештатной ситуации в среде имитационного моделирования AnyLogic, которое иллюстрирует работу протокола MQTT.

Литература

1. *Ли П.* Архитектура интернета вещей / пер. с анг. М. А. Райтмана. М.: ДМК Пресс, 2019. 454 с.
2. http://book.itep.ru/4/45/http2.htm (дата обращения: 15.01.2023).
3. ITU-T/ The Constrained Application Protocol (CoAP) / RFC 7252 – Proposed Standard. 2014.
4. MQTT Version 3.1.1 OASIS Standard. 2014.
5. Публичные облачные сервера для IoT устройств // kotyara12 URL: https://kotyara12.ru/iot/cloud_services/ (дата обращения: 23.01.2023).
6. https://santehnika-terra.ru/articles/mqtt-broker-mosquito-windows-nastroyka.html (дата обращения 15.01.2023).
7. *Боев В.Д.* Компьютерное моделирование: Пособие для практических занятий, курсового и дипломного проектирования в AnyLogic. Спб.: ВАС, 2014. 432 с.
8. *Antonova V.M., Malikova E.E., Panov A.E., Spichek I.V., Malikov A.Y.* Implementation of IoT technology for data monitoring via cloud services // T-Comm. 2021. Т. 15. № 2. С. 46-53.
9. *Степанов С.Н., Степанов М.С., Маликова Е.Е., Цогбадрах А., Ндайикунда Ж.* Построение и анализ обобщенной модели разделения ресурса для lte технологий с функциональностью NB-IOT // T-Comm: Телекоммуникации и транспорт. 2018. Т. 12. № 12. С. 71-77.
10. *Маликова Е.Е., Пшеничников А.П.* Особенности преподавания перспективных инфокоммуникационных технологий на кафедре сети связи и системы коммутации в МТУСИ // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2018. Т. 7. № 3. С. 35-42.
11. *Маликова Е.Е., Пшеничников А.П., Пелевин И.И.* О роли образования в реализации программы "Цифровая экономика Российской Федерации" // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2018. Т. 7. № 1. С. 32-36.

АНАЛИЗ АЛГОРИТМОВ ПО УЛУЧШЕНИЮ КАЧЕСТВА ТЕКСТА С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ EYE-TRACKING ПРИ ПОТОКОВОЙ ПЕРЕДАЧЕ ВИДЕОКОНТЕНТА

Петров Андрей Валерьевич,
Московский Технический Университет Связи и Информатики, Москва, Россия
petrovavch04@gmail.com

Дрибний Дарья Яковлевна,
Московский Технический Университет Связи и Информатики, Москва, Россия
dasha.dribniy@bk.ru

Бутовская Дарья Андреевна,
Московский Технический Университет Связи и Информатики, Москва, Россия
dashabutovsay@gmail.com

Родионов Алексей Анатольевич,
Московский Технический Университет Связи и Информатики, Москва, Россия
doomsong95@gmail.com

Егоров Дмитрий Аркадьевич,
Московский Технический Университет Связи и Информатики, Москва, Россия
d.a.egorov@mtuci.ru

Аннотация
В настоящее время существует проблема ухудшения качества текста при потоковой передаче видеоконтента. В данной статье проведен анализ существующих алгоритмов распознавания и увеличения качества текста с использованием новых технологии, нейросетей. Также в работе представлена методология, позволяющая оптимизировать работу существующих алгоритмов распознавания на основе нейросетей с использованием нового оборудования пользователей. Интерес к данному сектору увеличивается с каждым годом. Эта работа будет полезна при создании дополнительных тестовых материалов, а также для исследования работы алгоритмов по улучшению качества текста с использованием технологии eye-tracking при потоковой передаче видеоконтента.

Ключевые слова: *нейросеть, eye-tracking, качество, текст, входные данные.*

Введение

Текст является одним из наиболее выразительных средств коммуникации и может быть встроен в документы в качестве средства передачи информации [1].

Существуют различные алгоритмы улучшения качества изображения и видео. Такие алгоритмы получают широкое распространение в настоящее время. Например, индустрия и разделы науки, связанные с цифровыми оптическими приборами, такие как: микроскопия, астрономия, оптическое приборостроение. На современном этапе развития технологий системы кодирования видео показывают высококачественные и полностью удовлетворительные результаты [2, 11-17]. Однако различные искажения текста могут присутствовать ещё на этапе исходного видео. Это актуально на данный момент так как трафик видео растет с каждым годом [3, 4].

В данной статье рассматриваются алгоритмы распознавания текста, выявлены проблемы данной области и предложен путь решения. Хотя распознавание текста порождает множество задач, основная цель состоит в определении, есть ли текст на представленном изображении или в видео, и есть ли возможность, обнаружить, локализовать и распознать данный текст. Одним из новых способов улучшения качества изображения и видео являются нейросети.

Нейросети работают следующим образом: информация подается на входной слой, который передает информацию на скрытый слой Взаимосвязи между двумя слоями случайным образом присваивают веса каждому входному сигналу. Также вычисляется смещение, добавляемое к каждому

входному сигналу после умножения весов на каждый сигнал по отдельности. Далее взвешенная сумма передается в функцию активации. Функция активации определяет, какие узлы необходимо запускать для извлечения объектов. Модель применяет прикладную функцию к выходному слою для получения выходных данных. Веса корректируются, и выходные данные передаются обратно, чтобы свести к минимуму ошибку.

Анализ с нейросетей

Feed Forward Neural Networks (FFNNs) [5]: это тип искусственных нейронных сетей, в которых информация проходит только в одном направлении, от входа к выходу, через скрытые слои. Это одна из самых простых и наиболее часто используемых архитектур нейронных сетей.

В FFNNs используются входные нейроны для получения входных данных, затем скрытые слои обрабатывают данные с помощью функций активации и весов для получения выходного сигнала. Выходные данные сравниваются с желаемым выходом, после чего ошибка передается по обратному распространению для обновления весов и уменьшения ошибки. Этот процесс повторяется до тех пор, пока ошибка не достигнет приемлемого уровня.

К преимуществам FFNNs относят способность решать широкий спектр задач, таких как классификация, регрессия и распознавание образов. FFNN также хорошо масштабируются, что позволяет создавать глубокие нейронные сети с большим количеством скрытых слоев для решения сложных задач.

FFNNs также хорошо настраиваются, позволяя добавлять несколько скрытых слоев и настраивать различные гиперпараметры, такие как скорость обучения и количество нейронов на слой.

Также важно учитывать доступные ресурсы и вычислительную мощность, поскольку для обучения FFNNs может потребоваться большой объем данных и вычислений, особенно для глубоких сетей.

Однако у FFNNs есть и некоторые ограничения. Одним из главных недостатков является то, что данные нейросети не очень хорошо подходят для задач, требующих сохранения пространственных или временных связей между входами. Другой недостаток заключается в том, что сети FFNNs могут страдать от проблемы исчезающего градиента, когда градиенты становятся настолько малы, что уже не пригодны для эффективного обновления весов. Кроме того, сети FFNNs могут быть склонны к чрезмерной подгонке, когда сеть становится слишком сложной и запоминает данные обучения, но не может обобщить старые на новые данные.

Radial Basis Function Neural Networks [6]: Это особый класс нейронных сетей с прямой связью, состоящий из трех слоев: входного слоя, скрытого слоя и выходного слоя. Это принципиально отличается от большинства архитектур нейронных сетей, которые состоят из многих слоев и обеспечивают нелинейность за счет рекуррентного применения нелинейных функций активации. Входной слой получает входные данные и передает их скрытому слою, где происходит вычисление. Скрытый слой нейронной сети радиальных базисных функций является самым мощным и сильно отличается от большинства нейронных сетей. Выходной слой предназначен для задач прогнозирования, таких как классификация или регрессия.

Нейронные сети с радиальными базисными функциями – это обычно используемые искусственные нейронные сети, используемые для задач аппроксимации функций и классификации вспомогательных векторов, они позволяют аппроксимировать многомерные функции с помощью линейных комбинаций терминов, основанных на одной одномерной функции.

RBFNN состоит из входного, скрытого и выходного слоев. RBFNN строго ограничен наличием только одного скрытого слоя. Этот скрытый слой называется вектором признаков.

Дальше применяется нелинейная передаточная функция к вектору признаков, прежде чем переходить к проблеме классификации. Когда увеличивается размерность вектора признаков, увеличивается линейная разделимость вектора признаков.

Нелинейная разделимая проблема (проблема классификации образов) лучше разделима в пространстве высокой размерности, чем в пространстве низкой размерности.

Процесс обучения включает в себя выбор таких параметров: Прототип, коэффициент бета для каждого нейрона, матрица выходных весов между нейронами и выходными узлами.

Выходные веса можно обучить с помощью градиентного спуска. Преимущества заключаются в: простом дизайне, хорошем обобщении, более быстром обучении, только одном скрытом слое и прямой интерпретация значения или функции каждого узла в скрытом слое.

К минусам данного метода можно отнести: ограниченная выборка, фиксированная ошибка оценки модели.

Applications of Recurrent Neural Networks [7]: Разработанная для сохранения выходных данных слоя, рекуррентная нейронная сеть подается обратно на входные данные, чтобы помочь в прогнозировании результата слоя. Первый уровень обычно представляет собой нейросеть прямой связи, за которой следует слой рекуррентной нейронной сети, где некоторая информация, которая была у нее на предыдущем временном шаге, запоминается функцией памяти. В этом случае реализовано прямое распространение. Там хранится информация, необходимая для его дальнейшего использования. Если результат неверен, нейросеть использует ресурсы, чтобы уменьшить веса для обработки информации. Данная нейронная сеть может использоваться для обработки и локализации текста, выявления грамматических ошибок.

Формулировка возможных проблем обнаружения текста

Сложность среды, гибкие стили получения изображений и вариативность содержания текста создают различные проблемы, которые классифицированы в Таблице 1 и анализируются следующим образом.

Сложность сцены: В естественной среде появляются многочисленные искусственные объекты, такие как здания, символы и картины, которые по своей структуре и внешнему виду похожи на текст. Сам текст обычно выкладывается для удобства чтения. Сложность сцены заключается в том, что окружающая обстановка затрудняет различение текста и не текста.

Неравномерное освещение: при съемке изображений в естественных условиях часто встречается неравномерное освещение, обусловленное освещенностью и неравномерной реакцией сенсорных устройств. Неравномерное освещение приводит к искажению цвета и ухудшению визуальных характеристик, и, следовательно, к ложным результатам обнаружения, сегментации и распознавания.

Размытие и ухудшение качества: при живой съёмке происходит расфокусировка и размытие изображений, содержащих текст. Процедуры сжатия и декомпрессии изображений или видео также ухудшают качество текста, в частности, графического видеотекста. Типичное влияние расфокусировки, размытия и ухудшения качества заключается в том, что они снижают четкость символов и вносят изменения, что затрудняет выполнение основных задач, таких как сегментация.

Искажение перспективы: Искажение перспективы возникает, когда оптическая ось камеры не перпендикулярна текстовой плоскости (рис. 1).

Рис. 1. Пример искажения перспективы

Границы текста теряют прямоугольную форму, а символы искажаются, что снижает производительность моделей распознавания, обученных на неискаженных образцах.

Шрифты: Символы курсивных и прописных шрифтов могут накладываться друг на друга, что затрудняет выполнение сегментации. Символы различных шрифтов имеют большие внутриклассовые вариации и образуют множество подпространств, что затрудняет точное распознавание, когда число классов символов велико.

Изменение соотношения сторон: Текст, такой как дорожные знаки, может быть кратким, в то время как другой текст, такой как титры к видео, может быть гораздо длиннее. Другими словами, текст имеет разное соотношение сторон. Чтобы обнаружить текст, необходимо рассмотреть процедуру

поиска с учетом местоположения, масштаба и длины текста, что приводит к высокой вычислительной сложности.

Многоязычная среда: Хотя большинство латинских языков насчитывают десятки символов, в таких языках, как китайский, японский и корейский, существуют тысячи классов символов. В арабском языке есть соединенные символы, которые меняют форму в зависимости от контекста. Хинди объединяет алфавитные буквы в тысячи формы, которые представляют слоги. В многоязычной среде оптическое распознавание символов в отсканированных документах остается исследовательской проблемой, в то время как распознавание текста в комплексных изображениях является более сложной задачей [8].

Таблица 1

Категория	**Подкатегория**
Окружающая среда	Сложность сцены
	Неравномерное освещение
Получение изображение	Размытие и ухудшение качества
	Искажение перспективы
Текстовое содержание	Шрифты
	Искажение сторон
	Многоязычная среда

Новая метология обнаружения текста с улучшением качества текста с использованием новой технологии eye-tracking при потоковй передачи видеоконтента

В данном разделе рассматривается технология распознавания текста с помощью устройств отслеживания взгляда. Данная технология является достаточно новой, как и само направление, которое существует около 40 лет [9]. Технология используется в научных исследованиях, медицине, психологии, маркетинге, а также встраивается во множество пользовательских приборов в последние несколько лет. Одно из важных применений устройства – исследование когнитивных задач, таких как чтение текста, анализ диаграмм, оптических и зрительных иллюзий.

Технология отслеживания взгляда работает по определённому алгоритму (рис. 2).

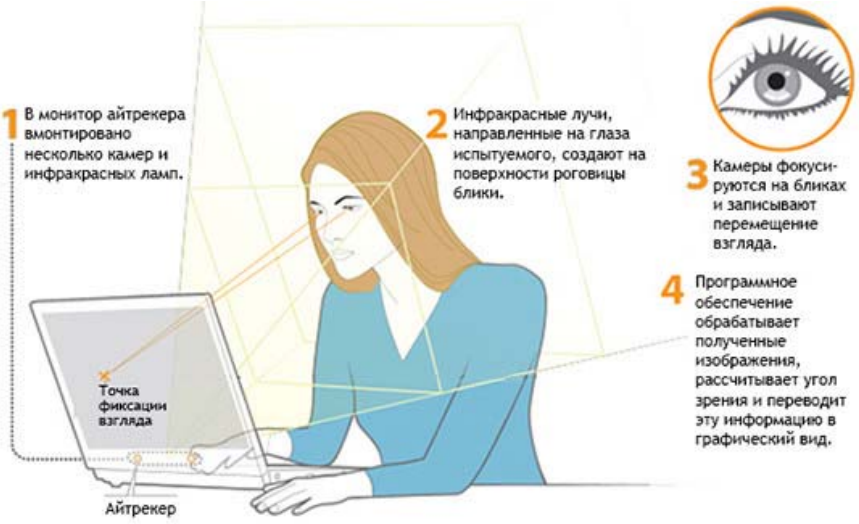

Рис. 2. Схема работы технологии eye-tracking

Устройство излучает ближний инфракрасный свет, далее свет отражается в глазах человека. Эти отражения улавливаются камерами eye-tracker. Благодаря калибровки и фильтрации, специальными алгоритмами и вычислениям система устройства идентифицирует взгляд пользователя, или, другими словами, создает визуальные карты с регионом интереса пользователя. [10]. Регион интереса (РОИ) – это определённая область, на которую человек сконцентрирован во время просмотра видео. В компьютерном зрении и оптическом распознавании символов РОИ определяет границы рассматриваемого объекта. При использовании данных РОИ считываемая зона уменьшается до области, необходимой для анализа, благодаря чему можно оптимизировать алгоритмы по улучшению качества текста с использованием технологии eye-tracking при потоковой передаче видеоконтента, или, другими словами, сократить датасет для обучения нейросетей.

Данная технология помогает отслеживать взгляд человека на экране. Зная куда смотрит человек, нейросети могут анализировать именно данную область для выявления возможного наличия текста. Если нейросеть находит текст, в регионе интереса то происходит обработка и выводится регион интереса с лучшим качеством.

Работа FFNNs зависит от характера задачи, количества и качества имеющихся данных. Если задача требует сохранения пространственных или временных связей между входными данными, то FFNNs не удовлетворяет условиям для использования в оптимизации работы алгоритмов по улучшению качества текста с использованием технологии eye-tracking при потоковой передаче видеоконтента. Если данные ограничены или зашумлены, FFNNs могут работать не корректно, и в сочетании с FFNNs могут потребоваться другие методы, такие как выбор признаков или уменьшение размерности.

К минусам RBFNN можно отнести: ограниченная выборка, фиксированная ошибка оценки модели, что также не удовлетворяет условиям для использования в оптимизации работы алгоритмов по улучшению качества текста с использованием технологии eye-tracking при потоковой передаче видеоконтента. RBFNN имеет производительность сети в ограниченном обучающем наборе данных, т.е. плохую способность к обобщению.

Радиальные базисные функции – это функции с действительными значениями, которые используют машинное обучение с учителем для работы в качестве нелинейного классификатора. Значение данной функции зависит от расстояния между входом и некоторой фиксированной точкой. RBFNN используются в искусственных нейронных сетях, используемых для задач аппроксимации функции и классификации вспомогательных векторов. Согласно, приведенному выше описанию и анализу можно сказать, что Radial Basis Function Neural Networks может подойти для оптимизации работы алгоритмов по улучшению качества текста с использованием технологии eye-tracking при потоковой передаче видеоконтента.

Заключение

На основании проведённого исследования, выявлено, что процесс обработки текста с использованием нейросетей на данном этапе развития технологий довольно сложный, так как имеется мало наборов данных для обучения, из-за чего происходит затрата большого количества времени и ресурсов. Нейросети показывают отличную производительность для решения поставленной задачи, и могут применяться для улучшения качества изображения и видео. Хотя нельзя не отметить тот факт, что нейронные сети — не до конца изученная технология, которая будет исследоваться ещё множество лет, и, при создании алгоритма может возникнуть множество трудностей, основными будут распознавание и обработка текста. В данной работе представлена методология решения данной проблемы, позволяющая оптимизировать количество данных с использованием нового оборудования пользователей.

Исследование выполнено за счет гранта Российского научного фонда № 23-29-00302 и гранта Российского научного фонда № 23-29-00299.

Литература

1. Давыдова А.А., Можаева А.И., Хурани В.Х. и др. Методология исследования идентичности восприятия регионов интересов пользователей при просмотре потокового видео содержащего различный контент и артефакты сжатия // Системы синхронизации, формирования и обработки сигналов. 2022. Т. 13. № 6. С. 42-51. EDN IZCJOE
2. Mozhaeva A., Streeter L., Vlasuyk I., Potashnikov A. Full Reference Video Quality Assessment Metric on Base Human Visual System Consistent with PSNR // 2021 28th Conference of Open Innovations Association (FRUCT), 2021, pp. 309-315.
3. *Mozhaeva A., Mazin V., Cree M.J., Streeter L*. Video Quality Assessment Considering the Features of the Human Visual System. In: Yan, W.Q., Nguyen, M., Stommel, M. (eds) Image and Vision Computing. IVCNZ 2022. Lecture Notes in Computer Science, 2023, vol 13836. Springer, Cham. https://doi.org/10.1007/978-3-031-25825-1_21
4. Nezhivleva K.I., Davydova A.A., Drebuzhan A.M., Mozhaeva A.I., Balobanov A. Comparing of Modern Methods Used to Assess the Quality of Video Sequences During Signal Streaming with and Without Human Perception // 2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), 2022, pp. 1-6, doi: 10.1109/SYNCHROINFO55067.2022.9840983.
5. *Anne-Johan Annema*. Feed-foorward neural networks Vector Decomposition Analysis, Modeling and Analog Implementation.
6. Radial Basis Functions: Theory and Implementations. Cambridge Monographs on Applied and Computational Mathematics, Series Number 12.
7. *Medsker L.R., Jain L.C.* Recurrent Neural Networks: Design and Applications // International Series on Computational Intelligence.
8. Ye AQ., Doermann D. Text detection and recognition in imagery: A survey // IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 7, pp. 1480–1500, Jul. 2015.
9. *Rayner K.* Eye Movements and Visual Encoding During Scene Perception. Psychological Science, 2009. Vol. 20, No. 1, pp. 6-10.
10. *Vyatkin M., Potashnikov A., Selivanov V., Vlasuyk I., Nezhivleva K., Mozhaeva A.,* Method of preventing leakage of personal data through eyetracking modules of user devices // T-Comm: Телекоммуникации и транспорт, 2022. Т. 16. № 7. С. 44-51.
11. *Mozhaeva A., Vashenko E., Selivanov V., Potashnikov A., Vlasuyk I., Streeter L*. Analysis of current video databases for quality assessment // T-Comm. 2022. Т. 16. № 2. С. 48-56.
12. *Власюк И.В., Пашковская А.Р., Мясникова В.С., Никольская Д.И., Можаева А.И.* Анализ стоимости создания современных баз данных видеопоследовательностей с субъективной оценкой качества // DSPA: Вопросы применения цифровой обработки сигналов. 2022. Т. 12. № 1. С. 4-11
13. *Егоров Д.А., Федоров В.Д., Лейман В.В., Власюк И.В.* Методика оценки пространственно-частотной характеристики камер на основе генеративных случайных последовательностей // DSPA: Вопросы применения цифровой обработки сигналов. 2021. Т. 11. № 1. С. 47-53.
14. *Иванчев В.В., Калужских Е.А., Власюк И.В*. Разработка локального метода сжатия динамического диапазона // DSPA: Вопросы применения цифровой обработки сигналов. 2021. Т. 11. № 2. С. 31-43.
15. *Можаева А.И., Власюк И.В., Поташников А.М., Стритер Ли*. Эталонная объективная метрика оценки качества видео совместимая с PSNR учитывающая частотные и периферическую характеристики зрения человека // DSPA: Вопросы применения цифровой обработки сигналов. 2021. Т. 11. № 2. С. 44-54.
16. *Valitskaya N.S., Vlasyuk I.V., Potashnikov A.M*. Video compression method on the basis of discrete wavelet transform for application in video information systems with non-standard parameters // T-Comm. 2020. Т. 14. № 3. С. 47-53.
17. *Поташников А.М., Власюк И.В.* Метод построения равноконтрастного цветового пространства для заданной системы отображения информации и условий контроля // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 4. С. 15-22.

МЕТОД ГЛАВНЫХ КОМПОНЕНТ КАК СПОСОБ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ

Шульпина Полина Дмитриевна,
Московский технический университет связи и информатики, Москва, Россия

Докучаев Владимир Анатольевич,
Московский технический университет связи и информатики, Заведующий кафедрой СИТиС, д.т.н., профессор, Москва, Россия,
v.a.dokuchaev@mtuci.ru

Аннотация
Обучение модели с учителем – будь то традиционная модель или модель глубокого обучения – включает в себя несколько этапов. Первый – это подача данных через модель, в результате чего формируются прогнозы. Второй – сравнение этих прогнозов с фактическими значениями, которые также называются эталонными. Третий – оптимизация модели на основе минимизации некоторой объективной функции. В ходе этого итерационного процесса модель становится лучше. Иногда входная выборка содержит множество столбцов. Использование каждого столбца в модели машинного обучения означает проблемы, так называемое проклятие размерности. В этом случае придется выборочно обрабатывать признаки. В этой рассматривается Метод главных компонент (Principal Component Analysis), который является одним из таких способов.

Ключевые слова: *Метод главных компонент, PCA, машинное обучение, предварительная обработка данных, глубокое обучение, извлечение функций, масштабирование функций.*

Введение

Происходящая в настоящее время цифровая трансформация затрагивает все стороны жизни человека: экономическую, политическую, социальную и т.д. [1-4]. В условиях роботизации, автоматизации и информатизации технологических процессов и бизнеса все большее значение приобретает правильность принимаемых решений и сокращение необходимого для этого времени. Поэтому актуальным становится внедрение систем поддержки принятия решений с использованием концепции ИИ. В то же время возникает необходимость передачи и обработки большого количества данных [15-21] для принятия правильного решения с учетом новых потенциальных рисков [5-8]. Одним из способов решения этой задачи является использование анализа главных компонент.

Метод главных компонент (PCA) — это способ уменьшить размерность данных, при этом потеряв наименьшее количество информации [9].

Цель PCA: найти набор векторов (главных компонент), которые лучше всего описывают разброс и направление данных, что позволяет выбрать некоторые для уменьшения размерности пространства признаков.

Машинное обучение и проклятие размерности

При обучении модели с учителем мы следуем трехэтапному итерационному процессу. Первым шагом в обучении модели является **подача набора данных в модель**. Для каждой выборки генерируются прогнозы.

На втором этапе **сравниваются прогнозы и эталонные значения (реальные цели)**. Это сравнение дает значение ошибки или потери.

Третий шаг - **улучшение модели**. В случае нейронных сетей градиенты вычисляются с помощью обратного распространения, а затем используются оптимизаторы для изменения внутренних компонентов модели. Затем начинается все сначала. После оптимизации модели, прогнозы становятся немного лучше. Необходимо просто продолжать итерации, пока мы не будем удовлетворены результатами. В задачах машинного обучения обычно требуется огромное количество обучающих данных, чтобы обеспечить наличие нескольких образцов с каждой комбинацией значений [10].

Поскольку увеличение размерности равносильно возрастающей потребности в большем количестве данных, необходимо уменьшить количество размерностей в наборе данных. Это называется уменьшением размерности [11].

Сокращение размерности: Отбор признаков и Извлечение признаков

В области уменьшения размерности существует два основных подхода: Отбор признаков и Извлечение признаков.

Отбор признаков включает процесс выбора подмножества релевантных признаков для использования при построении модели. Этот подход используется, когда большая часть дисперсии в наборе данных может быть объяснена несколькими переменными. Если остальные действительно не важны, то их можно легко отбросить без потери большого количества информации.

Извлечение признаков начинается с исходного набора измеренных данных и строит производные значения, которые должны быть информативными и не избыточными. В результате получается набор данных более низкой размерности, объясняющий большую часть дисперсии, сохраняя при этом относительную простоту. Особенно в случае, когда каждое измерение вносит равный вклад, извлечение признаков может быть предпочтительнее, чем выбор признаков. То же самое верно, если нет представления о вкладе каждой переменной в предсказательную силу модели.

Этапы PCA

1. Разброс набора данных выражается в ортонормальных векторах - главных направлениях.
2. Сортируя эти векторы в порядке важности, можно найти измерения данных (наибольшую дисперсию).
3. Сокращение числа измерений до самых важных.
4. Проецирование набора данных на новые измерения (главные компоненты), выполнив снижение размерности без потери значительной части информации.

На первом этапе разложение набора данных на векторы может быть выполнено двумя разными способами - с помощью разложения ковариационной матрицы по собственным векторам или разложения по сингулярным значениям [12].

Выражение разброса набора данных в векторах и их сортировка

После стандартизации набора данных можно представить направления разброса в виде пары двух векторов, как показано на рисунке 1.

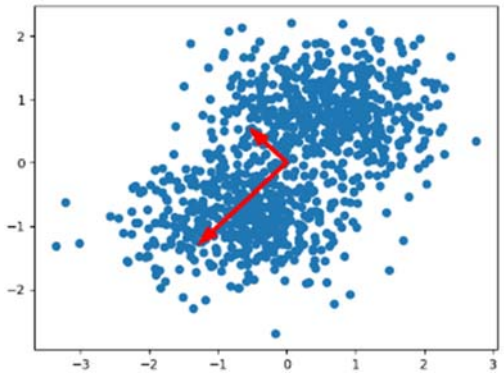

Рис. 1. Выражение разброса набора данных

Если внимательно посмотреть на набор данных, то можно увидеть, что он в основном распространяется в двух направлениях. Это направления от правого верхнего угла к левому нижнему углу и от правой нижней середины к левой верхней середине. Эти направления отличаются от направлений осей, которые ортогональны друг другу: оси x и y имеют угол 90 градусов.

Эти векторы называются **собственными векторами**. Их длина представлена **собственным значением**. Собственные значения объясняют дисперсию данных вдоль новых осей признаков [14]. Общий вклад собственных векторов в разброс для нашего примера выглядит следующим образом: $[0.76318124; 0.23681876]$·

Для разброса данных с помощью собственных пар используются два метода: **Разложение по собственным векторам** ("EIG") и **разложение по сингулярным значениям** ("SVD"). Их можно применять для выражения разброса набора данных в собственных парах для уменьшения числа измерений проецированием набора данных на наиболее важные из них, главные компоненты.

PCA-EIG: Разложение по собственным векторам

Одним из способов выполнения PCA является разложение по собственным векторам (EIG). Можно использовать ковариационную матрицу *N*-мерного набора данных и разложить ее на *N* собственных пар:

1. **Стандартизация набора данных**. PCA на основе EIG работает хорошо, только если набор данных центрирован и имеет среднее нулевое значение ($\mu = 0.0$).
2. **Вычисление ковариационной матрицы переменных.** Ковариационная матрица показывает, сколько дисперсий имеет каждая отдельная переменная, и насколько сильно переменные движутся вместе.
3. **Разложение ковариационной матрицы на собственные пары.** Математически возможно переписать ковариационную матрицу так, чтобы получить набор собственных векторов и собственных значений.
4. **Сортировка собственных пар в порядке убывания важности** для основных направлений, вносящих наибольший вклад в распространение.
5. **Выбор дисперсионного вклада главных направлений и выбор *n* главных компонент**. Происходит снижение размерности только *n* главных компонент, вносящих наибольший вклад.
6. **Построение проекционной матрицы** для проецирования исходного набора данных на главные компоненты.

Шаги 2 и 3 специфичны для PCA-EIG и представляют собой суть того, что делает PCA на основе разложения собственных векторов уникальным.

Использование многомерного набора данных Iris

Бесплатная библиотека машинного обучения для языка программирования Python Scikit-learn предоставляет набор данных Iris, который можно использовать для классификации трех групп цветков ириса на основе четырех характеристик: длины чашелистика (0), ширины чашелистика (1), длины лепестка (2) и ширины лепестка (3). Как видно из рисунка 2, два цветка ириса не могут быть линейно разделены, но эта группа может быть отделена от другого цветка ириса.

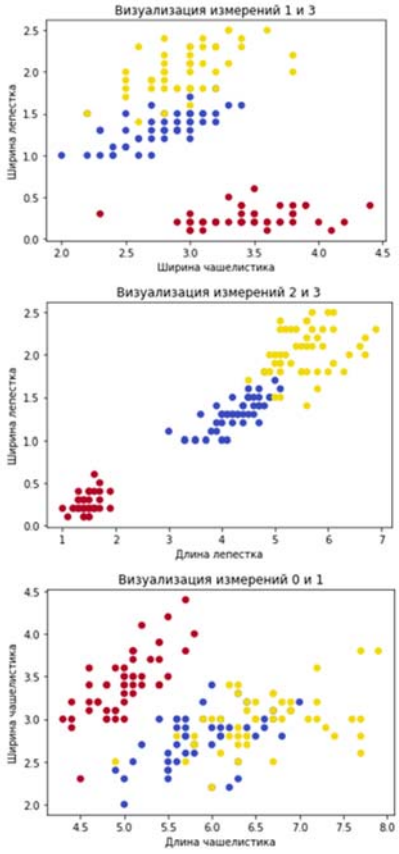

Рис. 2. Визуализация измерений

Мы приравниваем стандартное отклонение к единице, а среднее значение последовательности числовых данных к нулю ($\mu = 0.0$, $\sigma = 1.0$), выполнив $x = \frac{x - \mu}{\sigma}$ для каждого измерения. Так мы изменили способ отображения значений в пространстве модели [15].

Вычисление ковариационной матрицы переменных

Ковариационная матрица — это квадратная матрица, дающая ковариацию между каждой парой элементов данного случайного вектора.

Переменная. X – это математическое представление одного измерения набора данных.

Среднее значение переменной. Вычисляется как сумма всех значений, деленная на сумму значений.

Дисперсия. Описывает «разброс» данных вокруг переменной – сумма квадратов разницы между каждым числом и средним, т.е. сумма $(x - \mu)^2$ для каждого числа.

Ковариация. Описывает совместную изменчивость двух переменных. Для каждой пары чисел ковариация вычисляется как $Cov(x, y) = (x - \mu_x)(y - \mu_y)$.

Ковариационная матрица для *n* переменных. Ковариационная матрица для двух измерений X и Y выглядит следующим образом: $\begin{bmatrix} Cov(X,X) & Cov(X,Y) \\ Cov(Y,X) & Cov(Y,Y) \end{bmatrix}$

Свойства ковариационной матрицы: $Cov(X,X) = Var(X)$, $Cov(X,Y) = Cov(Y,X)$.

Следовательно, наша ковариационная матрица является симметричной и квадратной, $n \times n$ матрицей и, следовательно, может быть записана следующим образом:

$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{bmatrix}$$

Разложение ковариационной матрицы на собственные векторы и собственные значения

Особенность EIG-PCA заключается в том, что мы можем **разложить ковариационную матрицу на собственные векторы и собственные значения** следующим образом: $C = VLV^T$, где V – это матрица собственных векторов, где каждый столбец – собственный вектор, L – диагональная матрица с собственными значениями и V^T - транспонирование V. В результате получаем следующее:

$$\begin{bmatrix} 0.52103086 & -0.37921152 & -0.71988993 & 0.25784482 \\ -0.27132907 & -0.92251432 & 0.24581197 & -0.12216523 \\ 0.57953987 & -0.02547068 & 0.14583347 & -0.80138466 \\ 0.56483707 & -0.06721014 & 0.63250894 & 0.52571316 \end{bmatrix}$$

Собственные значения: 2,91912926, 0,91184362, 0,144265, 0,02476212.

Каждое главное измерение вносит свой вклад в объяснение дисперсии: $[0.72978232; 0.2279609; 0.03606625; 0.00619053]$. Вклад первого главного измерения составляет 73%, а второго – 23%. При построении матрицы проекций, исходные данные проецируются на главные компоненты, как показано на рисунке 3.

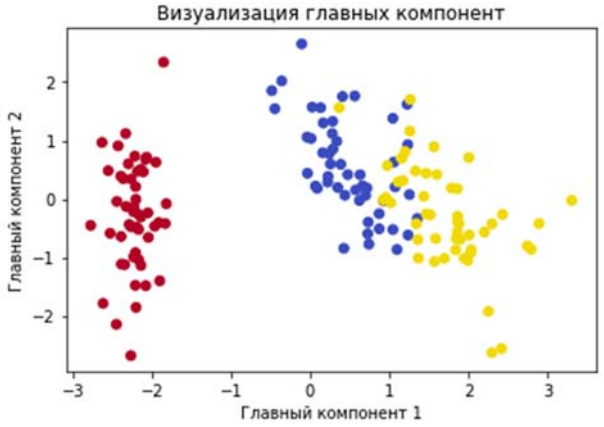

Рис. 3. Визуализация главных компонент

PCA-SVD: Сингулярное разложение значений

Один из наиболее численно-устойчивых методов является использование **разложения по сингулярным значениям** для матрицы данных вместо разложения по собственным векторам для ее ковариационной матрицы. В варианте SVD вычисляются **сингулярные значения матрицы данных**. В SVD матрица разлагается на три компонента: унитарные массивы U, векторы с сингулярными значениями s и унитарные массивы vh.

В EIG-варианте PCA мы вычислили ковариационную матрицу нашего набора данных, а затем выполнили разложение по собственным векторам для этой матрицы, чтобы найти собственные векторы и собственные значения. Теперь есть возможность использовать их для сортировки наиболее важных и проецирования набора данных на наиболее важные из них, т.е. на главные компоненты.

Перевод результатов SVD в векторы и значения

Выполнив SVD на матрице данных, мы можем получить те же результаты, что и при подходе PCA-EIG. При таком подходе собственные векторы ковариационной матрицы получаются следующими:

$$\begin{bmatrix} 0.52103086 & -0.37921152 & -0.71988993 & 0.25784482 \\ -0.27132907 & -0.92251432 & 0.24581197 & -0.12216523 \\ 0.57953987 & -0.02547068 & 0.14583347 & -0.80138466 \\ 0.56483707 & -0.06721014 & 0.63250894 & 0.52571316 \end{bmatrix}$$

Теперь сравним их с выходом vh:

$$\begin{bmatrix} 0.52106591 & -0.26934744 & 0.5804131 & 0.56485654 \\ -0.37741762 & -0.92329566 & -0.02449161 & -0.06694199 \\ 0.71956635 & -0.24438178 & -0.14212637 & -0.63427274 \\ 0.26128628 & -0.12350962 & -0.80144925 & 0.52359713 \end{bmatrix}$$

За исключением знака, столбцы vh равны строкам собственных векторов на основе EIG.

Выбор n компонентов и построение матрицы проекций

Принимается n = 2, а размерность уменьшается с 4 до 2. Строим матрицу проекций, спроецировав данные на главные компоненты и построить график проекции. Окончательный результат изображен на рисунке 4.

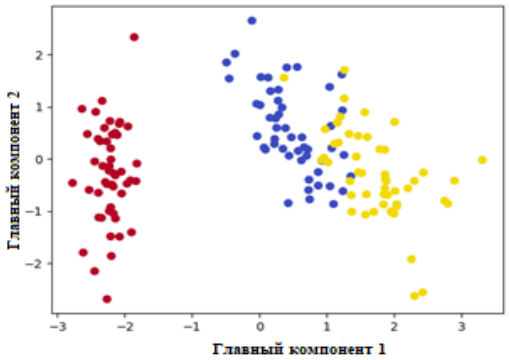

Рис. 4. Визуализация главных компонент с помощью PCA-SVD

Заключение

Таким образом, определив контекст применения PCA, появляется возможность вычислить собственные векторы и собственные значения и отсортировать их, чтобы найти главные направления в наборе данных. После создания проекционной матрицы для этих направлений мы можем отобразить набор данных на эти направления. Для получения собственных векторов существует два метода: использование разложения по собственным векторам (EIG) и более обобщенное разложение по сингулярным значениям (SVD).

Подходы к извлечению признаков, такие как PCA, которые пытаются построить пространство признаков более низкой размерности на основе исходного набора данных, могут помочь уменьшить это проклятие. Так можно воссоздать пространство признаков с меньшим числом измерений и с минимальной потерей информации.

Литература

1. *Павлов С.В., Леонович Е.В., Маклачкова В.В., Докучаев В.А*. Сети 2030: перспективы и проблемы // REDS: Телекоммуникационные устройства и системы. 2022. Т. 12. № 2. С. 17-23. EDN UZNLKQ

2. *Докучаев В.А., Кальфа А.А., Маклачкова В.В*. Архитектура центров обработки данных. М.: Горячая линия – Телеком, 2020. 240 с. ISBN 978-5-9912-0849-9. EDN BHARSE

3. *Dokuchaev V.A., Gorban E.V., Maklachkova V.V*. The System of Indicators for Risk Assessment in High-Loaded Infocommunication Systems // 2019 Systems of Signals Generating and Processing in the Field of on Board Communications, SOSG 2019, Moscow, 20-21 марта 2019 года. Moscow, 2019. P. 8706726. DOI 10.1109/SOSG.2019.8706726. EDN WFULUV

4. *Базаев А.Е., Докучаев В.А.* Проблемы настройки и администрирования маршрутизаторов CISCO серии ASR // Телекоммуникации и информационные технологии. 2021. Т. 8. № 2. С. 69-73. EDN AOERZX

5. *Kalmykov N.S., Dokuchaev V.A.* Segment routing as a basis for software defined network // T-Comm. 2021. Vol. 15. No 7. P. 50-54. DOI 10.36724/2072-8735- 2021-15-7-50-54. EDN LYVZCV

6. *Докучаев В.А., Калмыков Н.С*. Анализ возможности применения segment routing в программно-конфигурируемой сети // Тенденции развития интернет и цифровой экономики : труды IV Всероссийской международным участием научно-практической конференции, Симферополь – Алушта, 03-05 июня 2021 года. Симферополь: Крымский федеральный университет им. В.И. Вернадского, 2021. С. 31-33. EDN RBFDCX

7. *Калмыков Н.С., Докучаев В.А.*Анализ протоколов реализующих технологию программно-конфигурируемой сети // Телекоммуникации и информационные технологии. 2020. Т. 7. № 1. С. 19-25. EDN CPKZCY

8. *Shvedov A.V., Gadasin D.V., Alyoshintsev A.V.* Segment routing in data transmission networks // T- Comm. 2022. Vol. 16. No. 5. P. 56-62. DOI 10.36724/2072-8735-2022-16-5-56-62. EDN VAYLJQ

9. Zhu X., Dong H., Rossi P.S., Landro M. Выбор характеристик на основе анализа главных компонент для локализации подводных источников с помощью глубокого обучения. Норвежский университет науки и технологии, 2020. С. 1-15. DOI: 10.3390/rs13081486

10. Взаимосвязь между SVD и PCA. Как использовать SVD для выполнения PCA, StackExchange, [Электронный ресурс]. Режим доступа: https://stats.stackexchange.com/questions/134282/relationship-between-svd-and-pca-how-to-use-svd-to-perform-pca.

11. *Ефимов В.М., Ефимов К.В., Ковалева В.Ю.* Метод главных компонент и его обобщения для последовательностей любого типа // Вавиловский журнал генетики и селекции, 2019. С. 1032-1036. DOI: 10.18699/VJ19.584

12. *Рашка С*. Принципиальный компонентный анализ, [Электронный ресурс]. Режим доступа: https://sebastianraschka.com/Articles/2015_pca_in_3_steps.html.

13. Ламберс Д.В. Математические основы PCA, [Электронный ресурс]. Режим доступа: https://www.math.usm.edu/lambers/cos702/cos702_files/docs/PCA.pdf.

14. *Cristianini N., Shawe-Taylor J.* Введение в опорные векторные машины и другие методы обучения на основе ядра. Издательство Кембриджского Университета. С. 687-689. DOI: 10.1017/CBO9780511801389

15. *Гадасин Д.В., Шведов А.В., Кузин И.А*. Трехмерная реконструкции объекта по одному изображению с использованием глубоких свёрточных нейронных сетей // T-Comm: Телекоммуникации и транспорт. 2022. Т. 16. № 7. С. 29-35. DOI 10.36724/2072-8735-2022-16-7-29-35. EDN YTLCNW

16. *Гадасин Д.В., Кольцова А.В., Полякова А.Н.* Модель построения кластера для пограничных вычислений // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2020. № 1. С. 86-92.

17. *Dokuchaev V.A., Maklachkova V.V., Statev V.Yu.* Classification of personal data security threats in information systems // T-Comm. 2020. Т. 14. № 1. С. 56-60.

18. *Докучаев В.А., Маклачкова В.В., Статьев В.Ю.* Цифровизация субъекта персональных данных // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 6. С. 27-32.

19. *Pavlov S.V., Dokuchaev V.A., Mytenkov S.S.* Model of a fuzzy dynamic decision support system // T-Comm. 2020. Т. 14. № 9. С. 43-47.

20. *Pavlov S.V., Dokuchaev V.A., Maklachkova V.V., Mytenkov S.S.* Features of supporting decision making in modern enterprise infocommunication systems // T-Comm. 2019. Т. 13. № 3. С. 71-74.

21. *Гадасин Д.В., Нестерова Е.А.* Особенности проведения практических занятий по дисциплине мультимедийные информационные системы для стадии "Исследование и обоснование создания информационной системы" // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2021. Т. 10. № 1. С. 15-21.

ИСПОЛЬЗОВАНИЕ ПОДВИЖНЫХ РАДИОРЕЛЕЙНЫХ СТАНЦИЙ ДЛЯ ОРГАНИЗАЦИИ РАДИОСВЯЗИ МЕЖДУ ПОЛЕВЫМИ АЭРОДРОМАМИ

Сухопаров Павел Евгеньевич,
Краснодарское высшее военное авиационное училище лётчиков имени Героя Советского Союза А.К. Серова, г. Краснодар, Россия

Афонин Илья Евгеньевич,
Краснодарское высшее военное авиационное училище лётчиков имени Героя Советского Союза А.К. Серова, г. Краснодар, Россия

Решетникова Ирина Витальевна,
Северо-Кавказский филиал ордена Трудового Красного Знамени ФГБОУ ВО «Московский технический университет связи и информатики», г. Ростов-на-Дону, Россия

Аннотация
Статья посвящена обоснованию актуальности использования радиорелейных станций радиосвязи для обеспе-чения устойчивого обмена информацией между взаимодействующими и подчиненными авиационными частями и подразделениями при их рассредоточении по разветвленной сети полевых аэродромов. Акцентировано внима-ние на то, что именно наличие устойчивой радиосвязи является обязательным условием для решения задач управ-ления силами и средствами воздушно-космических сил. Рассмотрены особенности функционирования радиоре-лейной связи между полевыми аэродромами и приведены требования, предъявляемые к антеннам радиорелейной связи.

Ключевые слова: *полевой аэродром, система связи, автоматизированная система управления, радиорелейная связь.*

В настоящее время в ряде ведущих стран, в первую очередь в США, активно развиваются новые высокотехнологичные средства вооруженной борьбы, в том числе высокоточное оружие (ВТО), которое уже находится в арсеналах целого ряда государств и широко используется ими при решении своих военно-политических и экономических проблем. Анализ опыта локальных войн и вооруженных конфликтов последних лет с участием стран блока НАТО позволяет сделать вывод о том, что средствам ВТО отводится существенная роль при решении тактических и стратегических задач.

В случае начала вооруженного конфликта противник массово применяет ВТО по командным пунктам (КП) системы государственного и военного управления, элементам системы противовоздушной обороны (ПВО) и другим объектам критической государственной инфраструктуры (крупные промышленные объекты, электростанции, узлы железнодорожного и воздушного транспорта и т.д.), в том числе по аэродромам базирования авиации воздушно-космических сил (ВКС) [1].

Боеготовность и боеспособность авиации ВКС, как одной из составляющих ВС РФ, во многом определяется качеством системы управления, позволяющей своевременно и в полном объеме решать сложные и ответственные задачи, связанные с поддержанием на должном уровне обороноспособности государства. В соответствии с этим на систему боевого управления и связи возлагаются задачи, основной из которых является своевременный и качественный обмен информацией в виде приема и передачи приказов и распоряжений нижним звеньям управления, получение донесений из подчиненных соединений и частей, обеспечение надежного взаимодействия между соединениями и частями ВКС в мирное и, особенно, в военное время.

В условиях угрозы нанесения противником массированного применения средств воздушно-космического нападения авиация ВКС, для повышения ее живучести, может быть рассредоточена по разветвленной сети полевых аэродромов. Однако такие аэродромы в силу своего географического положения и дефицита времени развертывания средств связи и радиотехнического обеспечения не всегда могут быть не оснащены полноценной системой связи и автоматизированной системой управления (АСУ) войсками и оружием. В этом случае могут быть использованы мобильные средства связи и АСУ.

В свою очередь системы связи и АСУ должны соответствовать требованиям [5], предъявляемым к ним системой управления ПВО по таким параметрам, как:

своевременность – способность обеспечивать прохождение и обработку всех видов информации в заданные сроки или в реальном масштабе времени (при обеспечении требуемых уровней достоверности и безопасности связи);

достоверность – способность обеспечивать воспроизведение передаваемых сообщений в пунктах приема с заданной точностью;

безопасность – способность противостоять несанкционированному получению, уничтожению и (или) изменению информации, передаваемой (принимаемой, хранимой, обрабатываемой, отображаемой) с использованием технических средств связи и автоматизации управления, а также способность ликвидировать нарушение обмена информацией, возникшее вследствие всех видов воздействий на систему связи и АСУ, их элементы.

Эффективность функционирования средств связи и АСУ при ведении боевых действий может существенно снижаться в условиях применения противником средств РЭБ. Низкая эффективность применения каналов радиосвязи при постановке противником помех обусловлена тем, что резко ухудшается качество функционирования приемо-передающих устройств. Воздействие помех на приемо-передающие устройства системы, входящей в состав системы связи и АСУ аэродрома, затрудняет или исключает возможность приема сигналов и команд с вышестоящих звеньев управления и, следовательно, значительно затрудняет выполнение поставленной перед авиационным полком (дивизией) боевой задачи.

В случае рассредоточения средств авиационных частей и соединений по полевым аэродромам с одной стороны повышается их живучесть, но с другой стороны возникает ряд проблем, связанных с устойчивостью управления и взаимодействия по причине снижения качества передаваемой информацией в каналах радиосвязи, обусловленного ухудшением электромагнитной обстановки. Также определенные ограничения могут накладывать особенности рельефа местности, большие расстояния между аэродромами, а также удаленность пунктов управления (ПУ).

Возможности существующих полевых узлов связи авиационного полка по организационно-техническому построению, функциональной принадлежности элементов и специальным возможностям аппаратных станций могут существенно затруднять их готовность к обеспечению необходимым перечнем и качеством услуг связи должностных лиц вышестоящих пунктов управления, а также информационное обеспечение (поддержку) в принятии ими решений.

В таких условиях становится актуальным изыскание путей улучшения качества связи с взаимодействующими и подчиненными авиационными частями, рассредоточенными по разветвленной сети полевых аэродромов, с целью повышения устойчивости управления авиационными частями и подразделениями.

Одним из возможных направлений преодоления указанных трудностей является использование радиорелейных линий, что позволяет качественно обеспечивать решение поставленных задач как стационарными, так и подвижными КП [2]. Кроме того, такие каналы могут также использоваться в качестве резервных каналов связи для волоконно-оптических систем. При этом указанные системы связи сочетают в себе высокую надежность, возможность цифровой технологии передачи данных и высокую пропускную способность [3]. При этом возможность организации радиорелейной связи с заданными требованиями по обеспечению необходимого уровня живучести, помехоустойчивости, и рабочего сектора во многом определяется характеристиками антенн, используемых в системах радиорелейной связи и АСУ.

Необходимость существенного улучшения параметров радиотехнических систем управления войсками или создание новых перспективных средств связи ВКС диктует и повышение требований, предъявляемых к антенным системам, таких как уменьшение веса и габаритов антенн, заданная полоса пропускания, простота в эксплуатации, высокий энергетический потенциал радиолинии, надежность и т.д.

В соответствии с тем, что при организации взаимодействии с вышестоящими звеньями управления, а также с соседними полками в условиях территориального рассредоточения средств РТО и связи авиационных частей и соединений, прием приказов, сигналов и донесений формируется с различных азимутальных направлений, что приводит к необходимости обеспечения оперативной смены положения приемо-передающей антенны. Кроме того, для расширения зоны уверенного приема. Это, в свою очередь диктует необходимость использования специальных довольно громоздких антенно-мачтовых

опорно-поворотных устройств, которые для расширения зоны уверенного приема должны иметь значительную высоту, что неизбежно приводит к снижению показателя оперативности при разворачивании и сворачивании пунктов связи на полевых аэродромах.

В условиях активного радиопротиводействия противника антенно-фидерные устройства радиорелейных систем связи и АСУ должны обладать высокой помехозащищенностью, что может быть достигнуто за счет снижения уровня бокового и заднего излучения

Особое влияние на помехоустойчивость каналов радиорелейной связи, электромагнитную совместимость и их разведзащищенность от средств радиотехнической разведки оказывают характеристики излучения и, в первую очередь, диаграмма направленности. Так, повышение помехоустойчивости и разведзащищенности связано с возможностью формирования ДН с низким уровнем боковых лепестков (высоким коэффициентом защитного действия), высоким уровнем кросс поляризационной развязки, обеспечивающим возможность работы в системах с поляризационным уплотнением.

Для оценки влияния диаграммы направленности антенны на помехоустойчивость и разведзащищенность можно воспользоваться уравнением радиолинии I типа [4], на основании которого получено выражение характеризующее отношение сигнал/помеха на выходе антенны в виде

$$Q = \frac{P_{пер}^{(1)} \eta_{пер}^{(1)} G_{пер}^{(1)} R_2^2}{P_{пер}^{(2)} \eta_{пер}^{(2)} G_{пер}^{(2)} R_1^2} \frac{\left| F_{пр}\left(\theta_1, \varphi_1\right) \right|^2}{\left| F_{пр}\left(\theta_2, \varphi_2\right) \right|^2}. \quad (1)$$

В соотношении (1) $P_{пер}^{(1)}$ представляет собой мощность передатчика полезного сигнала; $\eta_{пер}^{(1)}$ - КПД передатчика полезного сигнала; $G_{пер}^{(1)}$ - КНД антенны, используемой для передачи полезного сигнала; $P_{пер}^{(2)}$ - мощность передатчика помехового сигнала; $\eta_{пер}^{(2)}$ - КПД передатчика станции помех; $G_{пер}^{(2)}$ - КНД антенны, используемой для передачи помехового сигнала; $F(\bullet)$ – нормированная ДН рассматриваемой антенны; θ_1 и θ_2 - угол максимума ДН и соответственно угол, которым определяется направление прихода помехи; R_1, R_2 является расстоянием между передающей сигнал и приемной антеннами и антенной станции помех и приемной антенной соответственно.

В то же время, диаграмма направленности антенны, как отмечалось ранее, определяет и разведзащищенность системы связи при передаче команд и сигналов. Это определяется тем, что мощность, излучаемая антенной в направлении антенны системы радиоразведки противника, может быть записана в виде:

$$P_{разв} = \frac{P_{пер} \eta_{пер} G_{пер} G_2 \lambda^2}{4\pi R^2} \left| F_{пер}\left(\theta_{разв}, \varphi_{разв}\right) \right|^2, \quad (2)$$

где $P_{пер}$, $\eta_{пер}$, $G_{пер}$,является мощностью, КПД, КНД передающей антенны соответственно; G_2 - КНД антенны системы радиоразведки противника, $F_{пер}\left(\theta_{разв}, \varphi_{разв}\right)$ диаграмма направленности антенны в направлении антенны системы радиоразведки противника.

Соотношения (1), (2) показывают, что повышение помехоустойчивости и разведзащищенности определяется формированием низкого уровня боковых и задних лепестков ДН.

Кроме того, конструкция антенны в значительной степени определяет коэффициент технической готовности и живучести пункта управления. В частности, с конструкцией антенны связана и возможность оперативного изменения направления связи в зависимости от изменения обстановки.

Таким образом, антенны перспективных станций радиорелейной связи по сравнению с существующими должны удовлетворять следующим требованиям:

- более низкий уровень боковых и задних лепестков ДН;
- возможность оперативной смены направления связи;
- возможность свертывания и развертывания антенных систем в короткие сроки;
- малые массо-габаритные характеристики.

В качестве одного из вариантов построения антенн радиорелейной связи может служить антенная система, включающая в свой состав совокупность кольцевых антенных решеток, размещенных вблизи поверхности металлического цилиндра с сечением в виде звездообразного контура. Форма несущей конструкции определяется тактико-техническими требованиями к углу разворота главного максимума диаграммы направленности (ДН) антенны. Выбор продольно ориентированных излучателей определяется условиями распространения электромагнитных волн над поверхностью земли.

Для улучшения характеристик согласования элементов АР, снижения уровня боковых лепестков ДН возможно использование цилиндрической конструкции, поперечное сечение которой представляет собой контур с определенным образом выбранными параметрами. При этом для обеспечения одинаковых характеристик каждого излучателя изменение геометрических параметров такого контура должно соответствовать шагу их размещения в составе решетки. Одним из вариантов является изменение параметров контура по закону $R = R_0 + \Delta R\cos(N\varphi)$, как показано на рисунке 1 [5]. Параметр N выбирается из условия обеспечения идентичности характеристик излучателей. Для исключения дополнительных главных максимумов в ДН расстояние между излучателями принимается равным 0,5λ … 0,6 λ.

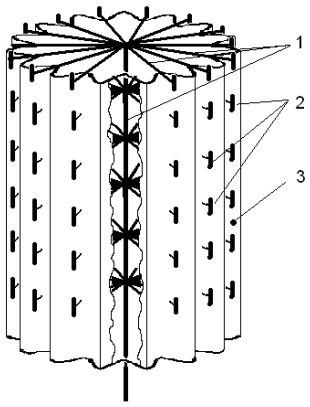

Рис. 1. Цилиндрическая антенна подвижной радиорелейных станций радиосвязи:
1 – несущий каркас; 2 – излучатели; 3 – металлическая фольга или прочная металлическая пленка

Решение вопросов, направленных на повышение живучести средств связи и АСУ, относится, прежде всего, к излучающему раскрыву антенной системы, поскольку аппаратный модуль антенны (СВЧ-тракт, система управления лучом, передающее и приемное устройства и т.д.) может быть размещен в отсеках подвижного ПУ или КП.

Как следует из проведенного анализа, применение радиорелейных линий радиосвязи, обеспечивающих устойчивое функционирование средств связи и АСУ в условиях ведения противником радиоэлектронной борьбы, может повысить устойчивость управления авиационными частями и подразделениями в случае их рассредоточения по разветвленной сети полевых аэродромов.

Литература

1. *Афонин И.Е., Ермаков Д.А.* Некоторые аспекты анализа информационного конфликта в технической сфере // Инновационные технологии в образовательном процессе. Сборник материалов XX Южно-Российской научно-практической конференции. Краснодар: КВВАУЛ, 2019. С. 42-46.

2. *Фролов О.П.* Антенны и фидерные тракты для радиорелейных линий связи. М.: Радио и связь, 2001. 416 с.

3. *Быховский М.А., Кирик Ю.М., Носов В.И.* и др. Основы проектирования цифровых радиорелейных линий связи: учебное пособие; под ред. М. А. Быховского. М.: Горячая линия – Телеком, 2014. 334 с.

4. *Черенкова Е.Л., Чернышев О.В*. Распространение радиоволн: учебник для вузов по специальности «Радиосвязь и радиовещание». М.: Радио и связь, 1984. 272 с.

5. *Габриэльян Д.Д., Звездина М.Ю., Лабунько О.С., Сухопаров П.Е*. Взаимосвязь геометрических параметров и поверхностного импеданса звездного контура // Электромагнитные волны и электронные системы. 2008. Т.13. №5. С. 33-34.

6. *Юхнов В.И., Енгибарян И.А., Сафарьян О.А., Сахаров И.А.* Оценка вариации Аллана при использовании метода статистической стабилизации частоты генераторов // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2019. № 1. С. 147-150.