

НАУЧНЫЙ ЖУРНАЛ

**ТЕЛЕКОММУНИКАЦИИ
И ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ**

№1-2022

(Дата издания: июнь 2022 г.)

Орлов Владимир Георгиевич (*Главный редактор*)

к.т.н., Главный специалист отдела организации научно-исследовательской работы студентов Московского технического университета связи и информатики «МТУСИ», Москва, Россия

Андреев Владимир Александрович

д.т.н., профессор, Поволжский государственный университет телекоммуникаций и информатики, Самара, Россия

Зимин Игорь Викторович

к.т.н., доцент, заведующий кафедрой Телекоммуникаций института Электроники и Телекоммуникаций при Кыргызском государственном технический университете имени И.Раззакова, Бишкек, Кыргызстан

Маркосян Мгер Вардкесович

к.т.н., доцент, Ереванский НИИ средств связи, Ереван, Армения

Нефёдов Виктор Иванович

д.т.н., профессор, Российский технологический университет МИРЭА, Москва, Россия

Самойлов Александр Георгиевич

д.т.н., профессор, заместитель директора института информационных технологий и радиоэлектроники Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых (ВлГУ), Владимир, Россия

Рогачев Александр Александрович

д.т.н., в.н.с., Гомельский государственный университет имени Франциска Скорины, Гомель, Республика Беларусь

Суржиков Анатолий Петрович

д.ф.-м.н., профессор, Национальный исследовательский Томский политехнический университет, Томск, Россия

Титов Евгений Вадимович

к.т.н., профессор, Московский технический университет связи и информатики, Москва, Россия

УЧРЕДИТЕЛЬ:

**ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ СВЯЗИ И ИНФОРМАТИКИ» (МТУСИ)**

РЕДАКЦИОННАЯ ПОДГОТОВКА:

**Отдел организации научно-исследовательской работы студентов
(ОНИРС МТУСИ)**

СОДЕРЖАНИЕ №1-2022

«Информационные технологии, кибернетика и информационная безопасность»

<i>Крейнделин В. Б., Легков Н.А.</i> ЗАЩИТА АУТЕНТИФИКАЦИОННЫХ ДАННЫХ САЙТОВ И WEB-ПРИЛОЖЕНИЙ	6
<i>Кононов Р. В., Шелухин О.И.</i> МНОГОКЛАССОВАЯ КЛАССИФИКАЦИЯ СЕТЕВЫХ АТАК МЕТОДАМИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА	11
<i>Кривенцова А. А., Трунов А.С.</i> АНАЛИЗ ГЛИКЕМИЧЕСКИХ ОТКЛОНЕНИЙ ПО АНАМНЕЗУ И АНАЛИЗУ КРОВИ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ.....	17
<i>Фурсаев К. В., Савельев Н. Д., Белов Н. В.</i> РАЗРАБОТКА МНОГОФУНКЦИОНАЛЬНОЙ СИСТЕМЫ ОЦЕНКИ СИТУАЦИИ	22
<i>Губин А. С., Тутова Н. В.</i> АНАЛИЗ ПОДХОДА К РАЗРАБОТКЕ ПРИЛОЖЕНИЙ С «ЧИСТОЙ» АРХИТЕКТУРОЙ	28
<i>Романова З. И., Кудряшова А. Ю.</i> НОВАЯ СОЦИАЛЬНАЯ СЕТЬ ДЛЯ ЛЮДЕЙ С МЕНТАЛЬНОЙ ИНВАЛИДНОСТЬЮ	38
<i>Ландызин А. Н., Шелухин О.И.</i> МЕТОДИКА ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ НАБОРА ДАННЫХ ДЛЯ БИНАРНОЙ И МНОГОКЛАССОВОЙ КЛАССИФИКАЦИИ АТАК	46
<i>Субботин А. А., Воронова Л. И.</i> ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ ПО РАСПОЗНАВАНИЮ ЖЕСТОВ	58
<i>Юсифов Э. С., Синева И. С.</i> АНАЛИЗ ДАТАСЕТА GOEMOTIONS ДЛЯ ДАЛЬНЕШЕЙ РАЗРАБОТКИ МОДЕЛЕЙ РАСПОЗНАВАНИЯ ЭМОЦИЙ НА БАЗЕ МАШИННОГО ОБУЧЕНИЯ	65

«Цифровые технологии радиосвязи и телерадиовещания»

<i>Гавриков О. В., Мальгин А. Д., Орлов В. Г., Попов О. Б., Самарин Е. В.</i> СНИЖЕНИЕ ИСКАЖЕНИЙ ЗВУКОВОГО СИГНАЛА ПРИ ЕГО ПЕРЕДИСКРЕТИЗАЦИИ	71
<i>Донцов Д. Е., Горгадзе С. Ф.</i> АНАЛИЗ И СРАВНЕНИЕ ВАРИАНТОВ ПОСТРОЕНИЯ УСИЛИТЕЛЕЙ МОЩНОСТИ МОБИЛЬНЫХ СЕТЕЙ 5G	83
<i>Бугаева Г. М., Набиев В. Р., Репинский В. Н.</i> ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ОБНАРУЖЕНИЯ СИГНАЛОВ	91
<i>Припутин В. С., Мартынов С. В.</i> АНАЛИЗ АЛГОРИТМОВ АДАПТИВНОГО ШУМОПОДАВЛЕНИЯ В СИСТЕМЕ БЕЗ ОПОРНОГО СИГНАЛА	100

<i>Ковалевская А. А., Санников В. Г.</i> СИНТЕЗ НОВЫХ СИГНАЛОВ ДЛЯ ФИЛЬТРА БАТТЕРВОРТА ПО КРИТЕРИЮ МАКСИМУМА СРЕДНЕГО ЗНАЧЕНИЯ ЕГО ОТКЛИКА	111
<i>Китаев А. П., Синева И. С.</i> ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ИЗОБРАЖЕНИЙ В ЗАДАЧАХ НЕФТЕГАЗОВОЙ ЛИТОЛОГИИ	116
<i>Демин А. Д., Лобов Е. М.</i> РАЗРАБОТКА ПРОГРАММНЫХ МОДУЛЕЙ ФОРМИРОВАНИЯ И ОБРАБОТКИ СИГНАЛОВ С ЧАСТОТНОЙ МАНИПУЛЯЦИЕЙ ДЛЯ ДЕКАМЕТРОВОЙ РАДИОЛИНИИ	127
<i>Литвяков В. С., Доронина А. Д., Панкратов Д. Ю.</i> МОДЕЛИРОВАНИЕ ВОСХОДЯЩЕЙ ЛИНИИ СВЯЗИ СИСТЕМЫ WSMA	135
«Сетевые технологии и системы телекоммуникаций»	
<i>Андреев С. А., Казанцев С. Ю., Свистунова А. И., Титовец П. А.</i> ПЕРСПЕКТИВЫ РАЗВИТИЯ СИСТЕМ ЛАЗЕРНОЙ СВЯЗИ В АТМОСФЕРЕ И ГИДРОСФЕРЕ	141
<i>Кузьмин М.С., Насарани А. П., Машковцева Л. С., Подлесных С. В.</i> ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК ЛАЗЕРНЫХ МОДУЛЕЙ, ВХОДЯЩИХ В МНОГОФУНКЦИОНАЛЬНУЮ ЛАЗЕРНУЮ УСТАНОВКУ (МЛУ-3)	153
<i>Тимощук Ю. С., Маклачкова В. В.</i> АНАЛИЗ РИСКОВ БЕСПРОВОДНОЙ ОПЛАТЫ С ИСПОЛЬЗОВАНИЕМ NFC	160
<i>Ассума Т. Л., Степанов С. Н.</i> АНАЛИЗ СЦЕНАРИЕВ РАСПРЕДЕЛЕНИЯ РЕСУРСА В СЛАЙСАХ ПРИ ГРУППОВОМ ПОСТУПЛЕНИИ ЗАЯВОК	165
<i>Титов А. В., Григорьева Е. Д.</i> РАЗРАБОТКА ЛАБОРАТОРНОГО ПРАКТИКУМА ПО РАЗДЕЛУ «СПЕКТРАЛЬНЫЙ АНАЛИЗ СИГНАЛОВ: ПЕРИОДИЧЕСКИХ, КОНЕЧНОЙ ДЛИТЕЛЬНОСТИ И ДИСКРЕТНЫХ»	172
<i>Мясищева Н. Р., Санников В. Г.</i> СИНТЕЗ НОВЫХ СИГНАЛОВ ДЛЯ ФИЛЬТРА БАТТЕРВОРТА ПО КРИТЕРИЮ МАКСИМУМА В ЦЕНТРЕ ЕГО ОТКЛИКА	179
<i>Лобов С. С., Шаврин С. С.</i> ОПТИМИЗАЦИЯ ПРОПУСКНОЙ СПОСОБНОСТИ СПУТНИКОГО КАНАЛА СВЯЗИ ЗА СЧЕТ ПРИМЕНЕНИЯ АДАПТИВНЫХ АЛГОРИТМОВ ПОМЕХОУСТОЙЧИВОГО КОДИРОВАНИЯ	183

«Информационные технологии, кибернетика и информационная безопасность»**ЗАЩИТА АУТЕНТИФИКАЦИОННЫХ ДАННЫХ САЙТОВ
И WEB-ПРИЛОЖЕНИЙ**

Крейнделин Виталий Борисович,
профессор кафедры ИБ, д.т.н., профессор, МТУСИ, Москва, Россия,
v.b.kreindelin@mtuci.ru

Легков Никита Алексеевич,
магистрант МТУСИ, Москва, Россия,
legkov98@bk.ru

Аннотация

В данной статье дан ознакомительный обзор методов аутентификации в web-приложениях и на web-сайтах. Актуальность данной проблемы связана с увеличением интереса злоумышленников к росту персональной и коммерческой информации, хранящейся на web-ресурсах. Методы были описаны и проанализированы с целью выявления их сильных и слабых сторон. Сейчас невозможно представить какой-либо бизнес-процесс без использования web-сервисов, поэтому проблема аутентификации будет оставаться актуальной. Также была рассмотрена двухфакторная аутентификация, которая обеспечивает более надежную защиту конфиденциальных данных.

Ключевые слова: аутентификация, пароль, защита, конфиденциальность, идентификация, надёжность.

Введение

В наши дни сложилась такая ситуация, при которой нельзя представить информационную систему, в которой бы не обрабатывались данные, содержащие коммерческую или государственную тайну. Интерес третьих лиц, которые не имеют легитимного права на обработку этих данных, возрастает со значительным темпом. В связи с этим, перед разработчиками web-ресурса стоит следующая техническая задача: обеспечить защиту конфиденциальной информации от хищения третьими лицами, последствия которого могут выразиться в изменении, удалении или подмены содержимого ресурса. Для обеспечения должного уровня конфиденциальности и целостности информации существуют методы аутентификации, которые заключается в том, чтобы пользователь смог подтвердить то, что у него есть право доступа к необходимым данным. Чтобы доказать свою подлинность, пользователь должен предоставить некий идентификатор аутентификации, например: логин, пароль, токен, отпечаток пальца, IP адрес, данные геолокации и т.д.

Обеспечение аутентификации на web-ресурсах

Обеспечение безопасности данных информационных web-ресурсов и web-приложений должно осуществляться путем реализации комплексной системы защиты информации и правильно заданной администратором политикой информационной безопасности. В совокупности необходимо использовать законодательные, административно-организационные и программно-технические меры [1,5].

К законодательным мерам можно отнести требования, законы, и нормативные акты, направленные на то, чтобы ограничить злоумышленника от попыток похищения аутентификационных данных. Политика безопасности включает в себя совокупность правил, направленных на защиту информации пользователей web-ресурсов. Администратору web-ресурса необходимо так настроить политику безопасности, чтобы доступ к конфиденциальным данным был у тех пользователей, кто имеет на это законное право.

К программно-техническим мерам можно отнести следующие современные компоненты, которые, как минимум, должны использовать создатели web-ресурса:

- аутентификация и идентификация;
- аудит журнала безопасности и контроль доступа;
- протоколирование и контроль;
- криптография;
- экранирование.

Рассмотрим следующие методы, которые в настоящее время используются для аутентификации пользователей на web-сайтах и в web-приложениях.

Метод парольной аутентификации

Основным традиционным методом аутентификации является метод парольной аутентификации [3,10]. Несмотря на то, что метод является не особо безопасным, он в ту же очередь является очень удобным. Метод заключается в том, что пользователь передает на сервер web- ресурса логин и пароль для успешной аутентификации.

Данный метод имеет множество уязвимостей, поскольку во многих случаях пароль можно подобрать методом логического подбора в силу того, что пользователи придумывают слишком простые и однотипные пароли, либо записывают на бумажном или usb-носителе.

В настоящее время метод является одним из часто используемых в силу своей дешевизны и простоты развертки на стороне ресурса. Он используется как для аутентификации в веб-приложениях, так и на веб-сайтах [2]. Лучше всего применять данный метод в совокупности с другими дополнительными способами защиты, к примеру:

- ответ на контрольный вопрос;
- СМС-код;
- PKI-сертификаты;
- биометрическая аутентификация.

Метод аутентификации по сертификатам

Метод аутентификации по сертификатам представляет собой метод аутентификации на web-сайте по специальным сертификатам, которые могут быть как самоподписанными, так и заверенные специальными удостоверяющими центрами. Сертификат представляет собой некоторый набор атрибутов, позволяющий идентифицировать и авторизовать владельца.

Разработчики своих web-ресурсов в основном реализуют процедуру проверки пользователя на основе сертификата стандарта X.509 [5]. Аутентификация в этом случае происходит в момент соединения с сервером. После успешного запроса пользователя, web-приложение может начать операцию аутентификации на основании следующих сведений, хранящихся в сертификате:

- уникальное имя владельца;
- серийный номер сертификата;
- Хеш-сумма сертификата.

Данный метод является более надежным, чем метод аутентификации по паролям. Это обусловлено тем, что в процессе аутентификации создается электронная цифровая подпись. Методы аутентификации по сертификатам получили широкое распространение и используются в настоящее время. Они обеспечивают надежную аутентификацию и хороший уровень криптозащиты [9,11].

Метод аутентификации одноразовым паролем

Одним из методов проверки подлинности пользователя является способ аутентификации по одноразовым паролям. Данный метод хорошо себя показывает в симбиозе с другими методами аутентификации для реализации двухфакторной аутентификации (2FA). Двухфакторная аутентификация представляет собой дополнительную гарантию защиты того, что к конфиденциальным данным доступ получит только законный пользователь. Основное назначение методов вторичной аутентификации - обеспечить дополнительную защиту аутентификационных данных, поскольку данные в таких методах являются динамичными и одноразовыми, что затрудняет задачу злоумышленника [4].

Наиболее популярными источниками для реализации метода одноразовых паролей являются аппаратные решения, например, аппаратное устройство RSA SecurID, представленный на рисунке 1. Такая технология позволяет создавать одноразовый уникальный код каждые 30 секунд, и позволяет определить права доступа пользователей к информации, прежде чем они начнут выполнять над ней какие-либо операции [7].



Рис 1. Аппаратное устройство RSA SecurID

Данный метод имеет довольно устойчивую защиту, поскольку время жизни одноразовых паролей ограничено, и они могут использоваться только один раз. Даже если злоумышленник сможет перехватить одноразовый пароль, по которому пользователь авторизовался в системе, сделать он с ним ничего не сможет.

Метод широко распространен в онлайн-платежах через банковскую систему. Процесс использования метода очень прост. Пользователю высылается пароль на мобильный номер для подтверждения операции оплаты. После истечения 60 секунд одноразовый пароль уже не имеет силу, и даже если злоумышленнику удастся его перехватить, никакого применения он не сможет ему найти.

Двухфакторная аутентификация

Двухфакторная или многофакторная аутентификация является дополнительной гарантией безопасности конфиденциальной информации, хранящейся на web-ресурсах. Это метод аутентификации пользователя в каком-либо сервисе, при котором система запрашивает у клиента несколько атрибутов для авторизации, тем самым обеспечивая двухслойную, то есть более надежную защиту ресурса. Преимущество метода заключается в том, что даже если злоумышленник сможет получить доступ к учетной записи пользователя, перехватить или узнать одноразовый пароль будет очень трудно. Поскольку одноразовые пароли имеют ограниченный срок действия, то это во много раз затрудняет задачу преступнику.

В настоящее время существует два широко используемых метода для получения одноразового пароля:

- метод на основе SMS (Short Message Service);
- на основе TOTP.

Метод на основе SMS, несмотря на свою простоту, является далеко не всегда надежным. Он имеет ряд проблем с безопасностью, а также приводит к долгому ожиданию отклика от сервера и необходимости в наличии мобильной связи в зоне использования. В 2016 году Национальный институт стандартов и технологий NIST (The National Institute of Standards and Technology) рекомендовал отказаться от этого метода аутентификации в новых системах [3,4].

Известен TOTP (Time-based One-time Password Algorithm) – алгоритм создания одноразовых паролей для защищенной аутентификации, при котором сервер удостоверяется в подлинности клиента. При внедрении двухфакторной аутентификации пользователю необходимо отсканировать QR-код при помощи специального ПО на смартфоне, который в дальнейшем постоянно генерирует новый одноразовый пароль. Метод является более надежным по сравнению с предыдущим, поскольку пароль генерируется на стороне клиента на его смартфоне, к которому он имеет непосредственный доступ. Это объясняется отсутствием фактора передачи пароля от сервера к клиенту для уменьшения шанса его перехвата.

Когда пользователь включает двухфакторную аутентификацию, происходит следующее:

- внутренний сервер создает секретный ключ для этого конкретного пользователя;
- на следующем сервер передает этот секретный ключ телефонному приложению пользователя;
- телефонное приложение запускает счетчик;
- мобильное приложение создаёт одноразовый пароль, используя связку секретного ключа и счётчика;
- телефонное приложение изменяет счетчик через определенный интервал и восстанавливает одноразовый пароль, делая его динамическим.

Актуальность метода двухфакторной аутентификации возраст с каждым днем. Почти каждый разработчик веб-ресурса реализуют ее как дополнительный гарант безопасности своего продукта. В силу своей простоты развертывания, технология 2FA должна развертываться в каждом приложении.

Биометрическая аутентификация

Одним из наиболее современных и интересных методов аутентификации является биометрическая аутентификация [1]. В наши дни используются следующие биометрические данные:

- отпечаток пальца;
- голос;
- лицо.

Аутентификацию по отпечатку пальца можно разделить на несколько типов. В емкостной аутентификации по отпечатку пальца измеряется уровень электрического сигнала, исходящий от наших пальцев. Полученные данные анализируются и вычисляется разница между приподнятой поверхностью отпечатка и его впадиной, после чего формируется отпечаток и сравнивается с исходным.

Одним из широко распространённых методов биометрии является метод сканирования геометрии лица, являющийся простым по своей природе, но довольно затратным в аппаратной обработке изображения. В современных устройствах он называется Face ID. Главное преимущество данного метода биометрии – в бесконтактности и возможности сканирования лица с большого расстояния.

Существует очень удобный в использовании метод биометрии, для которого нужно всего лишь наличие микрофона и звуковой схемы на устройстве пользователя. Аутентификация по голосу быстро развивается, находя применение в коммерческой среде и в бытовых нуждах, но имеет и ряд недостатков, один из которых – низкая точность метода. На успешный результат аутентификации влияют побочные шумы и состояние здоровья человека.

Методы биометрической аутентификации находят все большую область применения. Они применяются в смартфонах, для идентификации личности человека и предоставления доступа и управления системой, используются в специальных камерах на режимных объектах и т.д. Можно только предположить, что развитие метода не будет останавли-

ливаться и он будет набирать высокие темпы развития. Метод имеет высокую надёжность, ведь органы человека индивидуальны.

Но какой бы не была превосходной система, рост интереса злоумышленников к обходу такого вида аутентификации будет расти, и уже есть немало случаев, когда биометрическая аутентификация давала ложное срабатывание, например, при использовании нейросетевых систем, которую создают точную копию как лица, так и отпечатка пальца [1].

Аутентификация по контрольным вопросам

Метод аутентификации с контрольными вопросами представляет собой взаимодействие клиента с web-сайтом, при котором пользователю необходимо ответить на контрольные вопросы, ответы на которые, в теории, знает только он.

Существует два типа контрольных вопросов:

- вопросы, которые задаются на момент регистрации пользователя (ответы же вносятся в базу данных на этом же этапе);
- вопросы, связанные с деятельностью пользователя на web-ресурсе.

При применении данного метода следует учесть, чтобы все задаваемые вопросы были универсальными, поскольку они должны быть ориентированы на всех пользователей в системе [6]. Метод не нашел широкое распространения и практически не используется в наши дни. Метод легко даёт ложное срабатывание. Достаточно знать незначительную информацию о человеке, и можно получить доступ к конфиденциальной информации жертвы.

Заключение

Какие бы способы аутентификации разработчики не использовали в своих решениях, это не исключает возможность перехвата аутентификационных данных в момент их передачи и последующего их шифрования на сервер аутентификации. Для этого аутентификационные данные нужно обрабатывать так, чтобы снизить отрицательные последствия их перехвата. С данной задачей уверенно справляется протокол «https» на базе SSL/TLS-шифрования, но только при передаче данных с уже установленным безопасным соединением [8].

Различные способы аутентификации и авторизации позволяют реализовать высокий уровень защиты конфиденциальных данных. В зависимости от выполняемой задачи, разработчик web-ресурса вправе сам выбрать метод, обеспечивающий наилучшую защиту пользовательских данных.

Литература

1. Разбор сертификата X.509. [Электронный ресурс]. Режим доступа: <https://habr.com/ru/post/194664/>
2. Дзьобан П.И., Хализев В.Н. Защита сайтов и web-приложений // Сборник научных статей студентов Института информационных технологий и безопасности. Краснодар. Вып.3. 2012. С. 102-106.
3. Григорьев М.Ю. Методы обеспечения безопасности использования веб-приложений. [Электронный ресурс]. - Режим доступа: <http://diss.seluk.ru/pr-bezopasnost/1018786-1-magiste>.
4. Открытый проект о безопасности web-приложений – The Open Web Application Security Project (OWASP). [Электронный ресурс]. Режим доступа: <http://owasp.org>.
5. Травкин Ю.В. Персональные данные: Научное издание. М.: Амандавик, 2007. 430 с.
6. Kaspersky Lab. The evolution of phishing attacks: 2011-2013. [Электронный ресурс]. Режим доступа: http://media.kaspersky.com/pdf/Kaspersky_Lab_KSN_report_The_Evolution_of_Phishing_Attacks_2011-2013.pdf.
7. Герасименко В.А. Защита информации в автоматизированных системах обработки данных: в 2-х кн.: Кн. 1. М.: Энергоатомиздат, 1994. 400 с.
8. Зегжда Д.П., Ивашко А.М. Основы безопасности информационных систем. М.: Горячая линия – Телеком, 2000. 452 с.
9. Крейнделин В.Б., Авидэба А.Д. Шифрование Wi-Fi Protected Access. В сборнике: Технологии информационного общества. XI Международная отраслевая научно-техническая конференция: сборник трудов. 2017. С. 294.
10. Бельфер Р.А., Глинская Е.В., Орлов В.Г. Сравнительный анализ шифрования и аутентификации в сотовых сетях связи GSM, GPRS, UMTS И LTE // Фундаментальные проблемы радиоэлектронного приборостроения. 2018. Т. 18. № 4. С. 1099-1104
11. Крейнделин В.Б., Корса И.Д. Оценка эффективности и анализ существующих недостатков межсетевых экранов CISCO ASA. В сборнике: Технологии информационного общества. XI Международная отраслевая научно-техническая конференция: сборник трудов. 2017. С. 258.

PROTECTION OF AUTHENTICATION DATA OF SITES AND WEB

Vitaly B. Kreyndelin,

Professor of the Department of IS, Doctor of Technical Sciences, MTUCI, Moscow, Russia,

v.b.kreindelin@mtuci.ru

Nikita A. Legkov,

Graduate MTUCI, Moscow, Russia,

legkov98@vk.ru

Abstract

This article provides an introductory overview of authentication methods in web applications and websites. The urgency of this problem is associated with an increase in the interest of cybercriminals in the growth of personal and commercial information stored on web resources. The methods have been described and analyzed in order to identify their strengths and weaknesses. Now it is impossible to imagine any business process without using web services, so the problem of authentication will remain relevant. We also discussed two-factor authentication, which provides more reliable protection of confidential data.

Keywords: *authentication, password, protection, confidentiality, identification, reliability.*

МНОГОКЛАССОВАЯ КЛАССИФИКАЦИЯ СЕТЕВЫХ АТАК МЕТОДАМИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА

Кононов Руслан Валерьевич,
магистрант МГУСИ, Москва, Россия,
dremming@mail.ru

Шелухин Олег Иванович,
заведующий кафедрой ИБ МГУСИ, д.т.н., профессор, МГУСИ, Москва, Россия,
sheluhin@mail.ru

Аннотация

В статье рассматривается использование алгоритма интеллектуального анализа в задачах обнаружения и многоклассовой классификации аномалий. Актуальность данной проблемы вызвана развитием информационных и коммуникационных технологий, влекущих за собой рост количества уязвимостей. Приведено описание процесса создания искусственной нейронной сети прямого распространения - многослойный перцептрон. Произведена оценка полученных результатов несколькими алгоритмами. Показано, что созданное программное обеспечение обладает высокой эффективностью в решении поставленных задач.

Ключевые слова: защита информации, информационная безопасность, интеллектуальный анализ, информационные системы, уязвимости, атаки, нейронные сети.

Введение

Сегодня информационные системы повсеместно внедряются на предприятия, что служит увеличению количества возможных уязвимостей. Для противодействия каждому типу уязвимости, обнаруженному в информационной системе на предприятии, требуется выполнение ряда действий со стороны сотрудников службы безопасности. Поэтому, правильная классификация угроз является важным этапом в обеспечении информационной безопасности, в том числе при разработке систем обнаружения вторжений. Наиболее эффективные системы обнаружения вторжений представляют собой сложные аппаратно-программные комплексы. Например, для решения задач обнаружения и последующей классификации аномалий часто используются такие технологии, как накопление статистических данных и *Deep Packet Inspection*, что требует больших вычислительных мощностей [1].

Также, для обнаружения и классификации аномалий используются алгоритмы интеллектуального анализа [2, 18-29]. Данные алгоритмы могут самообучаться и противостоять новому типу угроз при сравнительно небольших требованиях к вычислительным мощностям.

Целью статьи является исследования эффективности использования методов интеллектуального анализа в задачах обнаружения и классификации аномалий.

Описание разрабатываемого программного обеспечения

Для решения задач обнаружения и многоклассовой классификации аномалий используется размеченный набор данных *CICIDS 2017*, который содержит информацию о 14 различных типах трафика [3]. Каждая запись об атаке характеризуется следующими атрибутами: 60 параметров и одна метка трафика. В выбранном наборе данных атаки делятся на шесть основных профилей, которые были сформированы на основе новейшего списка наиболее распространенных семейств атак:

- *Brute Force Attack* – это один из наиболее популярных типов атак, который может быть использован не только для взлома пароля, но также для обнаружения скрытого материала в веб-приложениях;
- *Heartbleed Attack* – это атака, использующая уязвимость, которая позволяет читать оперативную память небольшими фрагментами, размер которых составляет не более 64 КБ. Проблема заключается в том, что уязвимость двухсторонняя, а значит, пользователь может получить данные с уязвимого сервера, а сервер злоумышленника также может получить данные с оперативной памяти пользователя;
- *Botnet* – группа устройств, подключенных к сети Интернет, может быть использована для выполнения вредоносных задач, таких как кража данных, передача вредоносных программ, спам или запуск атак, предоставляющих злоумышленнику доступ к устройству;
- *DoS* атака – данный тип атак осуществляется следующим образом: несколько компьютеров отправляют огромное количество запросов в сеть-жертву, перегружая ее ресурсы и нарушая ее работу. Сеть, использованная для организации *DoS* атаки, как правило, представляет собой некоторое количество зараженных компьютеров;
- *Web Attack* – количество атак, произведенных таким способом, растет с каждым днем. Именно поэтому, организации и частные лица все сильнее задумываются о полноценной защите своих сетей. Существует огромное количество

ство программ и исполняемых кодов, созданных для сканирования компьютера с целью обнаружения уязвимостей в системе безопасности, позволяющих проникнуть в сеть жертвы.

- *Infiltration Attack* – это атака, которая осуществляется изнутри сети-жертвы. Данный тип атак использует уязвимое программное обеспечение, например, *Adobe Acrobat Reader*. После успешного проникновения в сеть, вирус будет запущен на компьютере жертвы и сможет проводить различные атаки на сеть жертвы.

На рисунке 1 представлено соотношение классов трафика в рассматриваемом наборе данных. Для обучения и тестирования ИНС в работе был использован заранее оптимизированный набор данных. Набор готов к использованию в задачах обучения и тестирования нейронных сетей. В качестве основной архитектуры для решения задач обнаружения и многоклассовой классификации аномалий был выбран многослойный перцептрон. Данный тип архитектуры характерен тем, что каждый нейрон предыдущего слоя связан с каждым нейроном следующего слоя [4].

Полный цикл разработки нейронной сети начинается с определения схемы алгоритма. На вход нейронной сети подается набор данных *CICIDS 2017*. Учитывая, что данный набор содержит текстовую информацию, необходимо предварительно обработать его с помощью кодировщика *LabelEncoder* [5]. Далее, используется алгоритм кросс-валидации для проверки работоспособности модели. Следующим шагом работы алгоритма является оценка качества построенной ИНС путем расчета числовых метрик.

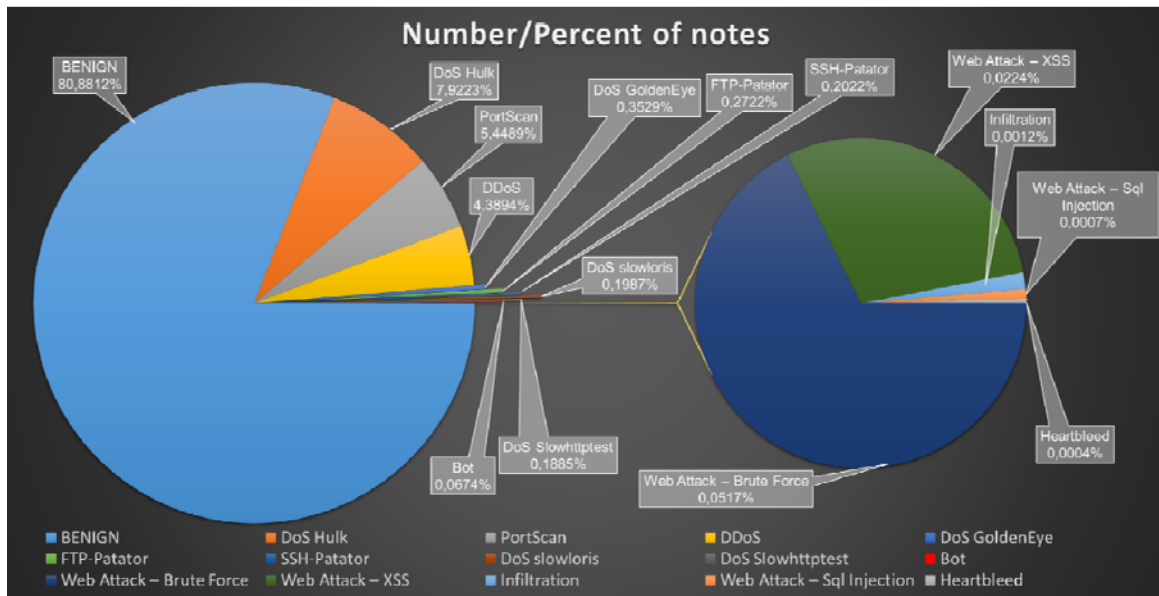


Рис 1. Соотношение классов атак в наборе данных *CICIDS 2017*

На рисунке 2 представлена блок-схема алгоритма обнаружения и многоклассовой классификации аномалий.

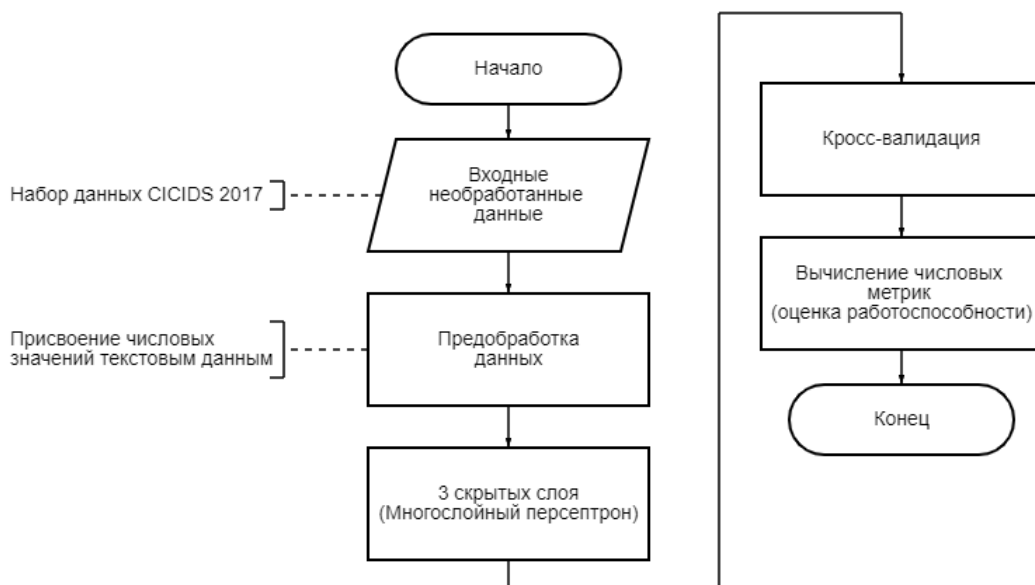


Рис 2. Блок-схема алгоритма выявления сетевых атак

Программная реализация осуществлялась средствами языка программирования Python и дополнительных библиотек, таких как *Scikit Learn*, *LabelEncoder*, *Pandas*, *Numpy*, *Copy*, *PyTorch*, *Logging*.

Scikit Learn – это один из наиболее широко используемых пакетов Python для *Data Science* и *Machine Learning* [6]. Данная библиотека использовалась для разбиения набора данных на две части:

- *train* – данная выборка использовалась для обучения искусственной нейронной сети. Размер выборки составляет 80% от общего значения;
- *test* – данная выборка нужна для конечной оценки работоспособности нейронной сети.

Для решения проблемы невозможности распознавания текстовой информации алгоритмами интеллектуального анализа, было принято решение об использовании библиотеки *LabelEncoder*. С помощью данной библиотеки был создан кодировщик, способный преобразовывать текстовые значения в числовые. В наборе данных *CICIDS 2017* присутствует текстовый параметр ‘*Label*’, который идентифицирует тип трафика. Таким образом, каждому типу трафика было присвоено собственное числовое значение.

Pandas — это специальная библиотека, которая позволяет проводить различные действия с временными рядами и числовыми таблицами [7]. Данная библиотека была использована в задачах чтения и объединения нескольких *csv* файлов набора данных *CICIDS 2017*.

Numpy – это специальный модуль для Python, позволяющий использовать математические и числовые операции в виде пре-скомпилированных функций, которые объединяются в пакеты более высокого уровня [8]. Данная библиотека была использована для простейших операций, таких как формирование случайного значения и обозначение простейших констант.

Модуль *Copy* позволяет осуществлять два типа операций копирования: глубокое копирование и поверхностное копирование [9]. Данный модуль был использован в задачах глубокого копирования, которое применялось в случае, если только что обученная модель показывала лучший результат в сравнении с предыдущей моделью. Использование модуля *Copy* позволило получить наиболее правильно обученную искусственную нейронную сеть.

PyTorch – это библиотека, которая облегчает создание моделей глубокого обучения [10]. При создании искусственной нейронной сети модуль *PyTorch* был использован в следующих задачах:

- использование функционала параллельных вычислений;
- использование таких утилит, как: *DataLoader* - для загрузки набора данных в ИНС, *RandomSampler* для получения содержания выборок в случайном порядке и *SequentialSampler* для получения содержания выборок в последовательном порядке [10,11];
- использование модуля *torch.nn* для получения доступа к удобному инструменту создания полносвязных нейронных сетей;
- использование метода «отсева» для решения проблемы переобучения ИНС;
- оптимизация нейронной сети с помощью изменения параметров скорости обучения.

Модуль *Logging* — это специальный инструмент, используемый в языке Python и предназначенный для логгирования хода работы программы [12]. Данный модуль был использован для ведения журнала о ходе обучения искусственной нейронной сети.

После этапа предобработки данных используется полносвязный слой нейронов, имеющий функцию активации *ReLu*. В качестве метода оптимизации используется алгоритм *AdamW*. Внешний цикл обучения нейронной сети состоит из 6 повторений. Внутренний тренировочный цикл проходит через все тренировочные данные в партиях, размер которых равен 64. Коэффициент скорости обучения нейронной сети был выбран равным 0,00005.

На рисунке 3 представлена структура ИНС с 71 признаком входных данных, 3 скрытыми слоями и 15 результирующими выходами.

В процессе обучения и тестирования производилась настройка модели нейронной сети путем изменения количества скрытых слоев размера *batch-size* и количества нейронов на каждом из слоев. Результаты данных тестов представлены в таблице 1.

Для тестирования работоспособности нейронной сети использовались четыре показателя: *F1 Score*, *Precision*, *Recall* [13]. Исходя из полученных данных, можно сделать вывод о том, что наилучшая конфигурация нейронной сети состоит из 5 эпох обучения и 3 скрытых слоев.

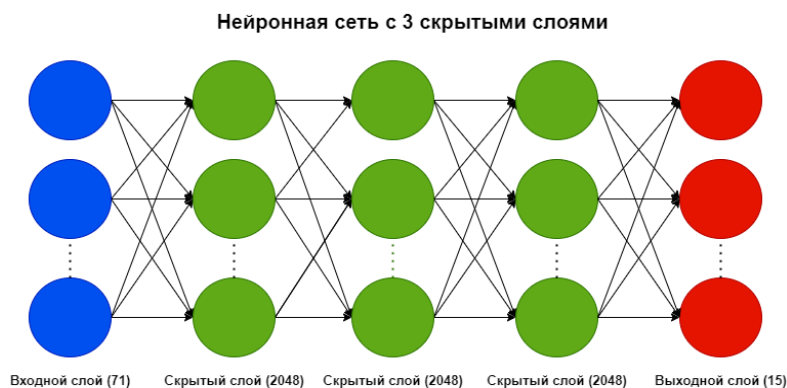


Рис 3. Структура ИНС с 3 скрытыми слоями

Таблица 1

Результаты тестирования различных конфигураций ИНС

Количество эпох	Количество скрытых слоев	Accuracy	Precision	Recall	F1 score
2	2	0.9188	0.9217	0.9043	0.9114
3	2	0.9532	0.9252	0.9654	0.9364
4	2	0.9076	0.9171	0.8932	0.9079
5	2	0.9261	0.9113	0.9258	0.9149
6	2	0.9153	0.9216	0.9211	0.9112
7	2	0.9330	0.9148	0.9332	0.9165
8	2	0.9353	0.9295	0.9143	0.9229
3	3	0.9204	0.9214	0.9112	0.9215
4	3	0.9264	0.9062	0.9304	0.9077
5	3	0.9396	0.9311	0.9521	0.9428

Метрика *Accuracy*(1.1) необходима для определения процентного соотношения правильных ответов, которые дала ИНС [14]. Данная метрика высчитывается с помощью отношения суммы *True Positive* и *True Negative* к сумме всех значений матрицы ошибок.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Метрика *Precision*(1.2) представляет собой отношение числа верно предсказанных ответов с меткой *Positive*, вне зависимости от достоверности предсказанного значения [15].

$$precision = \frac{TP}{TP + FP} \tag{2}$$

Recall(1.3) рассчитывается как отношение числа *Positive* выборок, корректно классифицированных как *Positive*, к общему количеству *Positive* ответов [16]. *Recall* измеряет способность модели обнаруживать выборки, относящиеся к классу *Positive*. Чем выше *Recall*, тем больше *Positive* ответов было найдено.

$$recall = \frac{TP}{TP + FN} \tag{3}$$

Метрика *F1 score*(1.4) необходима для нахождения наилучшего сочетания *Recall* и *Precision* [17].

$$F = 2 \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Исходя из результатов и учета степени важности метрики *Recall*, наилучшую эффективность показала нейронная сеть в следующей конфигурации: 3 эпохи обучения и 2 скрытых слоя.

Рисунок 4 иллюстрирует полученные результаты тестирования различных конфигураций ИНС.

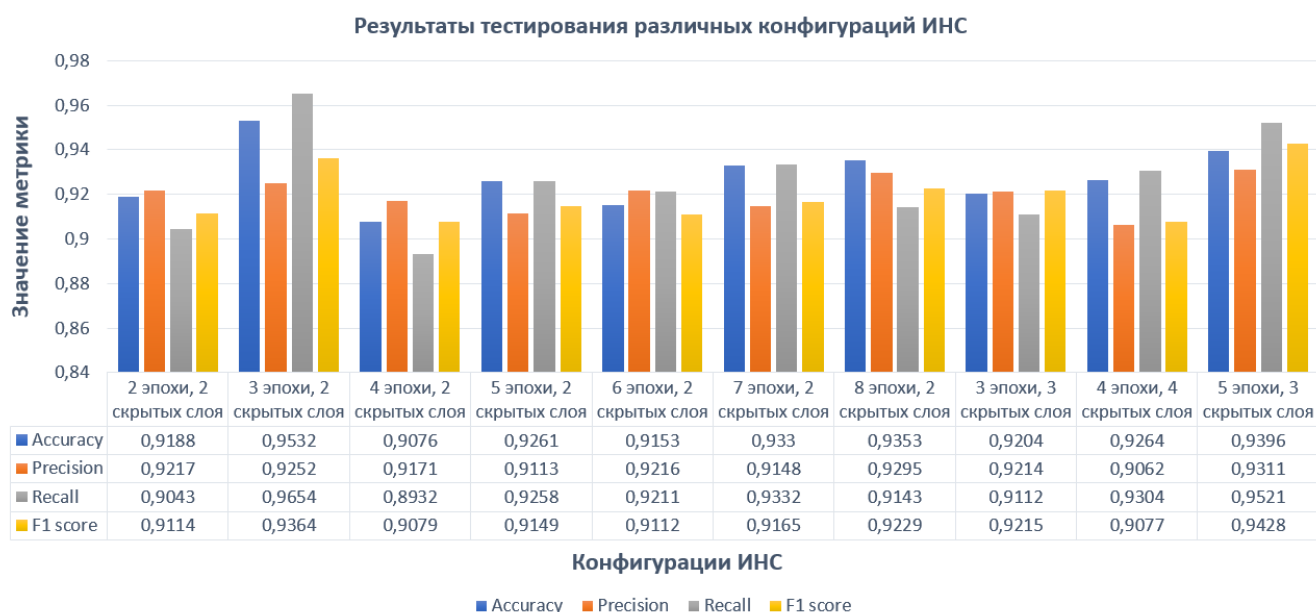


Рис 4. Результаты тестирования различных конфигураций ИНС

Выводы

В результате исследований было установлено, что искусственная нейронная сеть прямого распространения многослойный перцептрон показала высокую точность в рассматриваемой задаче обнаружения и многоклассовой классификации аномалий. Полнота классификации трафика составила 96,54%. Полученные результаты свидетельствуют о высокой надежности разработанного алгоритма.

Литература

1. Шелухин О.И., Ерохин С.Д., Полковников М.В. Технологии машинного обучения в сетевой безопасности. / Под ред. О. И. Шелухина. М.: Горячая линия – Телеком, 2021. 360 с.
2. Sheluhin O.I., Ivannikova V.P. Comparative analysis of informative features quantity and composition selection methods for the computer attacks classification using the UNSW-NB15 dataset // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14, № 10. С. 53-60.
3. Todor Balabanov, Ilyan Zankinski, Kolyu Kolev, "Multilayer Perceptron Training Randomized by Second Instance of Multilayer Perceptron", 2018, [Электронный ресурс]. Режим доступа: https://www.researchgate.net/publication/329680882_Multilayer_Perceptron_Training_Randomized_by_Second_Instance_of_Multilayer_Perceptron, свободный – (16.03.2022).
4. Confusion Matrix, Accuracy, Precision, Recall F1 Score [Электронный ресурс] // medium.com // URL: <https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63c>, свободный – (16.03.2022).
5. Sklearn.preprocessing.LabelEncoder [Электронный ресурс] URL: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>, свободный – (16.03.2022).
6. Что такое Scikit Learn - гайд по популярной библиотеке Python для начинающих [Электронный ресурс] URL: <https://datastart.ru/blog/read/chto-takoe-scikit-learn-gayd-po-populyarnoy-biblioteke-python-dlya-nachinayuschih>, свободный – (16.03.2022).
7. Введение в pandas: анализ данных на Python [Электронный ресурс] URL: <https://khashtamov.com/ru/pandas-introduction/>, свободный – (16.03.2022).
8. NumPy в Python [Электронный ресурс] URL: <https://habr.com/ru/post/352678/>, свободный – (16.03.2022).
9. Модуль copy – поверхностное и глубокое копирование объектов [Электронный ресурс] URL: <https://pythonworld.ru/moduli/modul-copy.html>, свободный – (16.03.2022).
10. A Comprehensive Guide to the DataLoader Class and Abstractions in PyTorch [Электронный ресурс] URL: <https://blog.paperspace.com/data-loaders-abstractions-pytorch/>, свободный – (16.03.2022).
11. PyTorch – ваш новый фреймворк глубокого обучения [Электронный ресурс] URL: <https://habr.com/ru/post/334380/>, свободный – (16.03.2022).
12. Модуль Logging в Python [Электронный ресурс] URL: <https://dev-gang.ru/article/modul-logging-v-python-sugk5e4d8u/>, свободный – (16.03.2022).
13. Шелухин О.И., Симонян А.Г., Ванюшина А.В. Влияние структуры обучающей выборки на эффективность классификации приложений трафика методами машинного обучения // T-Comm: Телекоммуникации и транспорт. 2017. Т. 11. № 2. С. 25-31.
14. Основные метрики задач классификации в машинном обучении [Электронный ресурс] URL: <https://webiomed.ru/blog/osnovnye-metriki-zadach-klassifikatsii-v-mashinnom-obuchenii/>, свободный – (16.03.2022).
15. Multi-Class Metrics Made Simple, Part I: Precision and Recall [Электронный ресурс] URL: <https://towardsdatascience.com/multi-class-metrics-made-simple-part-i-precision-and-recall-9250280bddc2>, свободный – (16.03.2022).
16. Precision and recall [Электронный ресурс] URL: https://en.wikipedia.org/wiki/Precision_and_recall, свободный – (16.03.2022).
17. Micro, Macro & Weighted Averages of F1 Score, Clearly Explained [Электронный ресурс] URL: <https://towardsdatascience.com/micro-macro-weighted-averages-of-f1-score-clearly-explained-b603420b292f>, свободный – (16.03.2022).
18. Шелухин О.И., Шариков А.Ю. Имитация поведения компьютерной системы с помощью искусственных нейронных сетей // T-Comm: Телекоммуникации и транспорт. 2021. Т. 15. № 5. С. 29-37.
19. Шелухин О.И., Раковский Д.И. Выбор метрических атрибутов редких аномальных событий компьютерной системы методами интеллектуального анализа данных // T-Comm: Телекоммуникации и транспорт. 2021. Т. 15. № 6. С. 40-47.
20. Шелухин О.И., Раковский Д.И. Бинарная классификация многоатрибутных размеченных аномальных событий компьютерных систем с помощью алгоритма SVDD // Научные технологии в космических исследованиях Земли. 2021. Т. 13. № 2. С. 74-84.
21. Шелухин О.И., Шариков А.Ю. Имитационное моделирование поведения компьютерной системы с помощью искусственных нейронных сетей // REDS: Телекоммуникационные устройства и системы. 2021. Т. 11. № 1. С. 54-60.
22. Шелухин О.И., Осин А.В., Костин Д.В. Диагностика "здоровья" компьютерной сети на основе секвенциального анализа последовательностных паттернов // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 2. С. 9-16.

23. Шелухин О.И., Осин А.В., Костин Д.В. Мониторинг и диагностика аномальных состояний компьютерной сети на основе изучения "исторических данных" // Т-Сomm: Телекоммуникации и транспорт. 2020. Т. 14. № 4. С. 23-30.
24. Шелухин О.И., Пугачев Д.А. Разработка искусственной иммунной системы на основе отрицательного отбора с применением нейросетевых детекторов для обнаружения компьютерных атак // REDS: Телекоммуникационные устройства и системы. 2020. Т. 10. № 1. С. 3-8.
25. Шелухин О.И., Костин Д.В., Резник И.Ю. Мониторинг и структура аномальных паттернов системных журналов компьютерных систем // REDS: Телекоммуникационные устройства и системы. 2020. Т. 10. № 2. С. 3-8.
26. Иванникова В.П., Шелухин О.И. Бинарная классификация компьютерных атак на примере базы данных UNSW-NB15 // Телекоммуникации и информационные технологии. 2020. Т. 7. № 1. С. 10-18.
27. Шелухин О.И., Барков В.В., Полковников М.В. Сравнительный анализ алгоритмов оценки количества и структуры атрибутов в задачах классификации мобильных приложений // Научные технологии в космических исследованиях Земли. 2019. Т. 11. № 2. С. 90-100.
28. Иевлев О.П., Шелухин О.И., Большаков А.С. Создание и техническая модернизация лабораторий кафедры "информационная безопасность" МТУСИ // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2019. Т. 8. № 1. С. 52-57.
29. Груздев С.П., Шелухин О.И. Бинарная классификация компьютерных атак на информационные ресурсы при помощи нечёткой логики // Телекоммуникации и информационные технологии. 2019. Т. 6. № 2. С. 115-122.

MULTICLASS CLASSIFICATION OF NETWORK ATTACKS BY METHODS OF INTELLECTUAL ANALYSIS

Ruslan V. Kononov,

Graduate MTUCI, Moscow, Russia,

dremming@mail.ru

Oleg I. Sheluhin

Head of the Department of IS of MTUCI, Doctor of Technical Sciences,

Professor, MTUCI, Moscow, Russia,

sheluhin@mail.ru

Abstract

The article discusses the use of the algorithm of intelligent analysis in the problems of detection and multiclass classification of anomalies. The urgency of this problem is caused by the development of information and communication technologies, which entail an increase in the number of vulnerabilities. The process of creating an artificial neural network of direct propagation - a multilayer perceptron is described. The results obtained were evaluated by several algorithms. It is shown that the created software has high efficiency in solving the tasks.

Keywords: *information security, information security, intelligent analysis, information systems, vulnerabilities, attacks, neural networks.*

АНАЛИЗ ГЛИКЕМИЧЕСКИХ ОТКЛОНЕНИЙ ПО АНАМНЕЗУ И АНАЛИЗУ КРОВИ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

*Кривенцова Александра Андреевна,
магистрант МГУСИ, Москва, Россия,
sport.dda@mail.ru*

*Трунов Артем Сергеевич,
старший преподаватель кафедры ИСУиА, МГУСИ, Москва, Россия,
Greek17@ya.ru*

Аннотация

В статье рассматриваются особенности анализа гликемических отклонений по анамнезу и анализу крови с помощью машинного обучения. Анализируются понятие и принцип работы методов машинного обучения, приводятся типы задач машинного обучения по выделяемому целевому признаку. Выявляются возможности использования машинного обучения в медицине в целом и в диабетологии в частности, рассматриваются специфика выделения признаков и методы машинного обучения для проведения анализа гликемических отклонений по анамнезу и анализу крови.

Ключевые слова: анализ крови, анамнез, гликемические отклонения, диабет, машинное обучение

Введение

Сахарный диабет является хроническим прогрессирующим заболеванием эндокринной системы, характеризующимся гипергликемией (повышенным уровнем содержания глюкозы в крови), возникающей вследствие нарушения секреции инсулина – гормона поджелудочной железы [1]. Организация эффективной медицинской помощи даёт возможность обеспечить проведение основных необходимых вмешательств и дальнейшее постоянное наблюдение за пациентами, что позволяет больным диабетом прожить долгую и относительно здоровую жизнь. Повысить качество жизни пациентов, больных диабетом, возможно при помощи своевременной диагностики заболевания и коррекции лечения, для чего необходимо выявление гликемических отклонений по анамнезу и анализу крови [2]. Перспективным направлением профилактики, диагностики и лечения диабета является внедрение методов машинного обучения (далее – МО), позволяющих с высокой точностью анализировать данные и составлять требуемые прогнозы, что делает актуальным исследование возможностей их применения в диабетологии.

Целью работы является изучение особенностей анализа гликемических отклонений по анамнезу и анализу крови с помощью МО. Для её достижения были использованы методы анализа и синтеза научных публикаций и литературных источников по рассматриваемой теме.

МО представляет собой отрасль информатики и статистики, разновидность искусственного интеллекта и способ создания программных продуктов на базе анализа данных [3]. В основе МО лежат компьютерные алгоритмы, в которых для решения задач вместо программирования формальных правил используются исходные данные, подвергающиеся различным математическим преобразованиям в рамках поставленной цели. Итоговый алгоритм часто использует неочевидную логику, позволяющую решать поставленные задачи с высокой степенью эффективности.

Точность алгоритмов МО определяется показателями специфичности и чувствительности и прямо пропорциональна их величине. Определение точности модели проиллюстрировано на рисунке 1.



Рис. 1. Определение точности модели МО

В зависимости от способа, которым задаётся целевой признак, выделяют следующие типы задач МО [4]:

1. Обучение «с учителем». Применяется при необходимости построения прогнозной модели по исходной выборке и значениям целевого признака, который указан для всех её элементов. По типу целевого признака выделяют задачи классификации, в которых целевым признаком является метка класса, принадлежащая множеству нечисловой природы, и регрессионного анализа, в которых прогнозируемый признак может принимать вещественные значения.

2. Обучение «без учителя». В этом случае целевой признак отсутствует. Эта группа включает задачи снижения размерности, требующие формирования системы признаков меньшей размерности без потери существенной информации, и кластерный анализ, для которого требуется разбиение множества наблюдений на однородные группы.

3. Полуавтоматическое обучение. Используется в случае наличия обычно небольшого объёма обучающей информации в виде набора значений целевого признака только для части выборки.

Способности алгоритмов МО к прогнозированию обусловлены их способностью к обобщению и выделению скрытых зависимостей, существующих между входными и выходными данными [5]. Это делает МО мощным инструментом для управления здоровьем, диагностики и лечения отдельных заболеваний. В настоящий момент применение МО в медицине в целом, и в диабетологии в частности, возможно для решения следующих задач [6]:

- агрегирование электронной медицинской документации и обеспечение более широкого доступа к ней;
- облегчение отчётности благодаря автоматическому нахождению необходимых данных;
- сопоставление предыдущих и текущих исследований, осуществляемое с высокой скоростью и точностью;
- быстрое выявление патологий посредством оперативной сортировки результатов анализов и анамнеза;
- внутренняя система проверки медицинских заключений;
- оперативная маршрутизация пациентов;
- контроль качества работы лаборантов и отслеживание связи между отдельными специалистами;
- анализ данных по специфическим параметрам.

Перед построением алгоритма МО, позволяющего проводить анализ гликемических отклонений по анамнезу и анализу крови, необходимо выбрать из базы данных пациентов наиболее информативные признаки [7]. Для автоматической выборки подмножества наиболее релевантных признаков используются алгоритмы отбора признаков, наиболее удобным из которых для анализа данных анамнеза и анализа крови является метод последовательного обратного отбора, относящийся к семейству «жадных» алгоритмов поиска. В результате применения данного метода, в качестве наиболее значимых признаков выделяются клинические признаки пациентов, такие как: возраст, пол, индекс массы тела, информация о течении болезни, осложнениях и сопутствующих заболеваниях, известные лабораторные параметры и прочее. Также выделяются признаки минимального, среднего и максимального уровня глюкозы пациента, градиентные характеристики, индексы на основе метрик DTW подобия, квадратичные, линейные и кубические коэффициенты тренда и параметры варибельности показателей гликемии [8].

Для проведения анализа гликемических отклонений по анамнезу и анализу крови могут использоваться следующие методы МО:

1. Метод опорных векторов (SVM) - гибкий и мощный класс алгоритмов обучения с учителем для задач классификации и регрессии, пример работы, которого представлен на рисунке 2 [9]. Его возможности расширяются при комбинации с ядрами, позволяющей проецировать данные в пространство, имеющее большую размерность. Повысить точность SVM можно при помощи извлечения признаков, становящихся входными данными для классификатора.

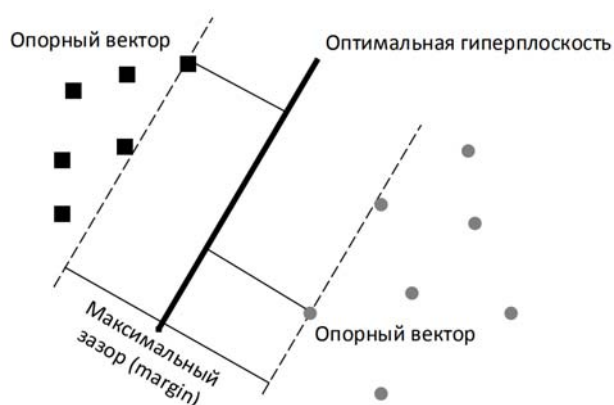


Рис. 2. Пример работы алгоритма SVM [10]

2. Алгоритм k-ближайших соседей (KNN). Этот метод интеллектуального анализа данных классифицирует неизвестный образец, используя известную информацию о классах его соседей. Пример такой классификации отражён на рисунке 3. Решающую роль в точности классификации методом KNN играют функция расстояния и выбор значения параметра k.

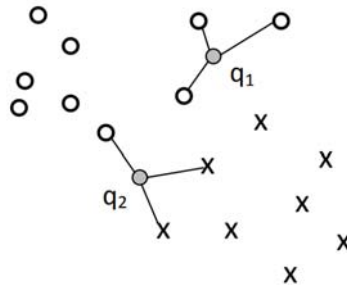


Рис. 3. Пример классификации методом KNN на основе трёх соседей [10]

3. Сверточные нейронные сети. Сеть классической архитектуры является модель VGG, представленная на рисунке 4. Она используется для распознавания признаков на результатах анализа, целью которого является повышение глубины подобных сетей. VGG отличается простотой при наличии полностью связанного уровня и уровня подвыборки.

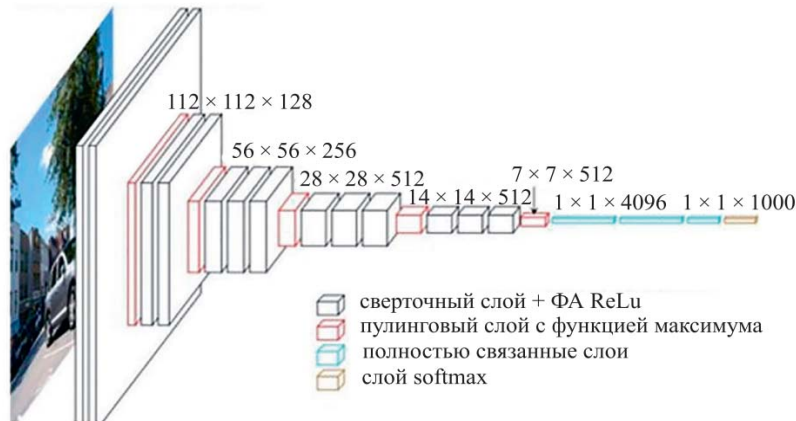


Рис. 4. Архитектура нейронной сети модели VGG

4. Ансамблевые методы. Использование гибридных нейронных сетей позволяет получить наибольшую эффективность анализа, поскольку отдельные компоненты модели могут охватывать разные спектры признаков. К примеру, в случае применения ансамбля глубокой и широкой нейронных сетей (схема на рис. 5), глубокая часть сети обрабатывает регулируемые признаки, а широкая – статические [11]. Эффективность обработки повышается в случае совмещения различных моделей, стремящихся к получению несхожих решений.

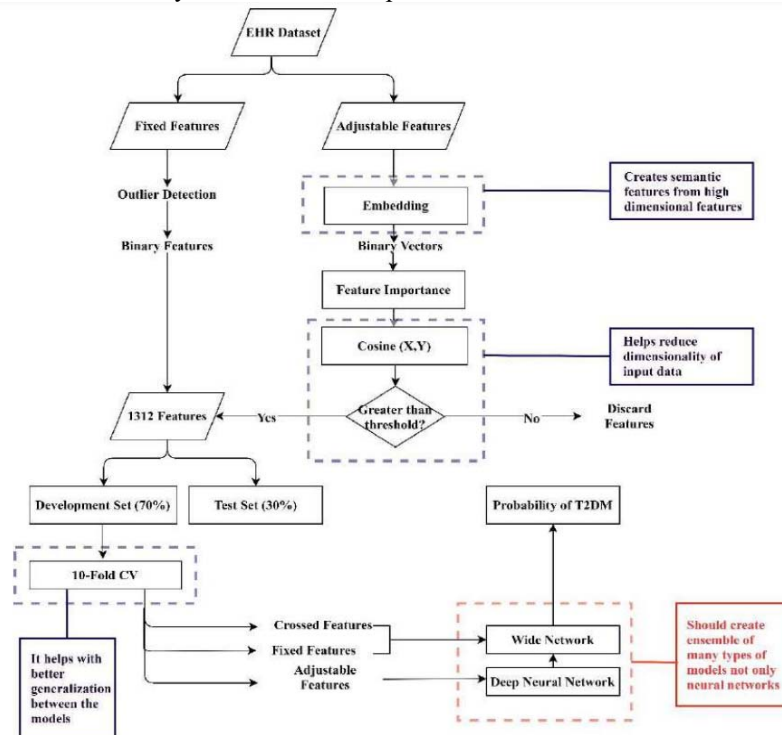


Рис. 5. Блок-схема ансамбля глубокой и широкой нейронных сетей

Согласно исследованиям, точность прогнозов алгоритмов МО по обнаружению сахарного диабета может достигать до 92 %, что значительно выше по сравнению с традиционными методами [12].

Среди используемых алгоритмов МО наибольшую эффективность показали ансамблевые методы, наиболее популярными из которых являются Random Forest и XGBoost, что обусловлено высокой точностью прогнозирования. Добиться повышения точности этих алгоритмов на 12-19 % возможно за счёт устранения их переобучения, что отражено на рисунке 6 [13].

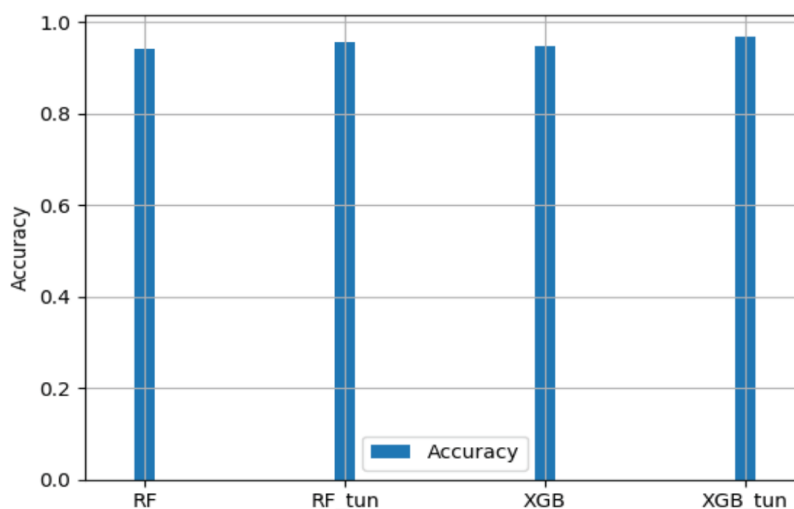


Рис. 6. Точность классификации при использовании алгоритмов Random Forest и XGBoost до и после настройки моделей

Выводы

Машинное обучение, основанное на данных анамнеза и анализа крови, является перспективным подходом для изучения механизмов развития диабета, оптимизации методов его профилактики, диагностики и лечения. МО становится незаменимым инструментом при анализе больших данных в области гематологии и диабетологии, а алгоритмы анализа динамики гликемии, разработанные с помощью МО, используются в автоматизированных системах для введения инсулина. Различные методы машинного обучения могут широко применяться для анализа больших электронных баз данных, регистров сахарного диабета, разработки новых технологий диагностики заболевания и прогноза осложнений, дистанционного мониторинга и скрининга, создания систем поддержки принятия решений и в исследованиях, проводимых в условиях реальной клинической практики, что соответствует глобальным трендам современной медицины.

Литература

1. Глобальный доклад по диабету [Global report on diabetes]. – Женева: Всемирная организация здравоохранения, 2018. 88 с.
2. Кротова О.С. Применение ансамблевых методов машинного обучения для диагностики сахарного диабета // Информация и образование: границы коммуникаций. 2018. № 10 (18). С. 87-89.
3. Лучинин А.С. Искусственный интеллект в гематологии // Клиническая онкогематология. 2022. № 15 (1). С. 16-27.
4. Климонтов В.В., Бериков В.Б., Сайк О.В. Искусственный интеллект в диабетологии // Сахарный диабет. 2021. Т. 24, № 2. С. 156-166.
5. Крутько В.Н., Донцов В.И., Митрохин О.В., Матвеев А.А., Ермакова Н.А., Потемкина Н.С. Искусственный интеллект для здоровьесбережения и развития личностного потенциала (обзор) // Труды Института системного анализа Российской академии наук. 2020. Т. 70, № 3. С. 86-100.
6. Мецьякова А.М., Акопян Э.А., Слинин А.С. Искусственный интеллект в медицинской визуализации. Основные задачи и сценарии развития // Журнал телемедицины и электронного здравоохранения. 2018. № 3 (8). С. 98-102.
7. Кротова О.С., Хворова Л.А., Пиянзин А.И. Математические модели и алгоритмы машинного обучения в диагностике осложнений сахарного диабета первого типа // Известия Алтайского государственного университета. 2021. № 1 (117). С. 97-101.
8. Бериков В.Б., Козинец Р.М., Семенова Ю.Ф., Климонтов В.В. Использование машинного обучения для прогнозирования ночной гипогликемии у пациентов с сахарным диабетом 1 типа // Сахарный диабет-2021: от мониторинга к управлению: материалы IV Российской мультидисциплинарной конференции с международным участием. Новосибирск, 2021. С. 23-26.
9. Елагина Е.А., Маргун А.А. Исследование методов машинного обучения в задаче идентификации клеток крови // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21, № 6. С. 903-911.

10. Петрин Д.А., Белов Ю.С. Технологии работы с редуцированными обучающими данными в задачах классификации // E-Scio. 2019. № 11 (38). С. 566-576.
11. Маллика Г.Ч., Альсадун А., Фам Т.Х.Ф., Абдулла С.Х., Май Х.Е., Прасад П.В.Ч., Нгуен Ч.К.В. Новая интеллектуальная система для обнаружения сахарного диабета 2-го типа с модифицированной функцией потерь и регуляризацией // Труды ИСП РАН. 2021. Т 33. № 2. С. 93-114.
12. Гусев А.В., Новицкий Р.Э., Ившин А.А., Алексеев А.А. Машинное обучение на лабораторных данных для прогнозирования заболеваний // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. 2021. № 14 (4). С. 581-592.
13. Краснова И.А. Анализ влияния параметров алгоритмов Machine Learning на результаты классификации трафика в режиме реального времени // T-Comm: Телекоммуникации и транспорт. 2021. Т. 15, № 9. С. 24-35.

ANALYSIS OF GLYCEMIC VARIATIONS BY HISTORY AND BLOOD TEST USING MACHINE LEARNING

Aleksandra A. Krivencova
Graduate MTUCI, Moscow, Russia,
Nebokrasnoe7@gmail.com

Artem S. Trunov
Senior Lecturer of the Department of IC&AS, MTUCI, Moscow, Russia,
Greek17@ya.ru

Abstract

The article discusses the features of the analysis of glycemic abnormalities by history and blood test using machine learning. The concept and principle of operation of machine learning methods are analyzed, types of machine learning tasks are given according to the selected target feature. The possibilities of using machine learning in medicine in general and diabetology in particular are revealed, the specifics of feature extraction and machine learning methods for analyzing glycemic deviations from anamnesis and blood tests are considered.

Keywords: *blood test, medical history, glycemic abnormalities, diabetes, machine learning.*

РАЗРАБОТКА МНОГОФУНКЦИОНАЛЬНОЙ СИСТЕМЫ ОЦЕНКИ СИТУАЦИИ

Фурсаев Кирилл Вячеславович,
студент МТУСИ, Москва, Россия
fursaev.kirill@gmail.com

Савельев Никита Денисович,
студент МТУСИ, Москва, Россия
nikita.savelev@lm.net.ru

Белов Никита Вадимович,
ассистент кафедры ИСЦиА, Москва, Россия
n.v.belov@mtuci.ru

Аннотация

Статья посвящена разработке многофункциональной интеллектуальной системы оценки ситуации на основе Raspberry Pi с применением технологий интеллектуальной обработки данных. Для решения проблемы удаленного контроля и управления системой, на примере системы контроля доступа, планируется интегрировать алгоритмы компьютерного зрения, облачные технологии, а также создание доступного интерфейса посредством Telegram-бота.

Ключевые слова: умный дом, автоматизация, распознавание лиц, Raspberry Pi, Python, OpenCV, Google drive.

Введение

Умный дом – идея, берущая свое начало с появлением диммирования освещения в шестидесятых годах, становится по-настоящему “умным” в наше время, благодаря технологиям интеллектуальной обработки данных [1]. В действительности “умный дом” – это не включение света по голосовой команде, и не регулирование температуры с телефона. Это совокупность устройств, собирающих информацию, процессов интеллектуальной обработки данных и построенных поведенческих стратегий в сфере автоматизации [2].

В данной статье производится реализация системы оценки ситуации для “умного дома”. Интегрирование технологий различного уровня и прикладной части позволило создать надежную работоспособную систему.

Одной из систем, “Умного дома” является система контроля доступа [3]. В данном проекте реализуется идентификация лиц, пытающихся проникнуть на частную территорию. При масштабировании и проецировании на более высокие уровни, таких как “Умный район” или “Умный город”, системы, разработанные в данной статье, помогут в составлении портретов и отслеживании злоумышленников, а также визуальном контроле за интересующими объектами.

Принцип работы системы заключается в следующем: в режиме активного сканирования система “слушает” датчик движения, при срабатывании датчика, с помощью камеры делается фото и начинается запись видео, далее проводится обработка полученной фотографии на предмет распознавания лиц, добавленных в базу данных. В случае, если лиц не было обнаружено или распознавание было неуспешным, система отправляет уведомление в Telegram со снятыми файлами. Все снятые файлы автоматически выгружаются в облачное хранилище Google Drive.

В данном программном комплексе был решен вопрос удаленного управления системой, то есть в любой момент времени владелец имеет возможность запросить фотографию или видеозапись, включить или выключить режим активного сканирования через Telegram-интерфейс, реализованный посредством чат-бота.

Структура проекта представлена на рисунке 1.



Рис. 1. Структура проекта

В статье рассматриваются основные технические характеристики многофункциональной системы оценки ситуации на основе микрокомпьютера *Raspberry Pi 4B*: пользовательский интерфейс, созданный средствами чат-бота *Telegram*; алгоритмы компьютерного зрения, разработанные с применением библиотеки *OpenCV*; автоматизированная выгрузка файлов в облачный сервис *Google Drive*. Код программного решения реализован на языке *Python 3.9*.

Основные функциональные блоки

В процессе разработки проекта были выделены следующие крупные проектные блоки:

1. Реализация аппаратной части проекта;
2. Создание пользовательского интерфейса для управления системой;
3. Реализация алгоритмов компьютерного зрения;
4. Реализация алгоритмов загрузки данных в облачное хранилище.

Реализация аппаратной части проекта

Логическим ядром системы был выбран одноплатный микрокомпьютер *Raspberry Pi 4*. Выбор конкретно этой модели обусловлен техническими характеристиками – это четырехядерный *ARM* процессор с тактовой частотой 1.5ГГц и 4 гигабайта оперативной памяти. Наличие на борту интерфейса ввода-вывода (*GPIO*) позволяет подключать внешние устройства различной номенклатуры непосредственно к контроллеру [4].

На практике было установлено, что при текущей нагрузке системы, ЦП не нагружается больше чем на 30%, расход ОЗУ в номинальном режиме работы 1 гигабайт, включая занимаемый операционной системой объем. На рисунке 2 показано основное используемое оборудование.

Съемка осуществляется с помощью внешней *web*-камеры *Logitech C270*. Подключение камеры осуществляется напрямую к контроллеру через *USB* интерфейс. Камера позволяет осуществлять фотосъемку в максимальном разрешении 1280x720, но в нашей работе использовалось разрешение 640x426 в целях снижения размера файлов и ускорения их передачи в дальнейшие сервисы

Для снятия информации о движении используется датчик *HC-SR501-RIP*. Это недорогой, надежный и простой в использовании датчик, который удовлетворяет потребностям системы.

Спецификация проекта



Рис. 2. Используемое оборудование

Принцип работы с датчиком заключается в следующем: при срабатывании, датчик отправляет положительное логическое значение на порт ввода контроллера.

Общая схема подключения оборудования изображена на рисунке 3.

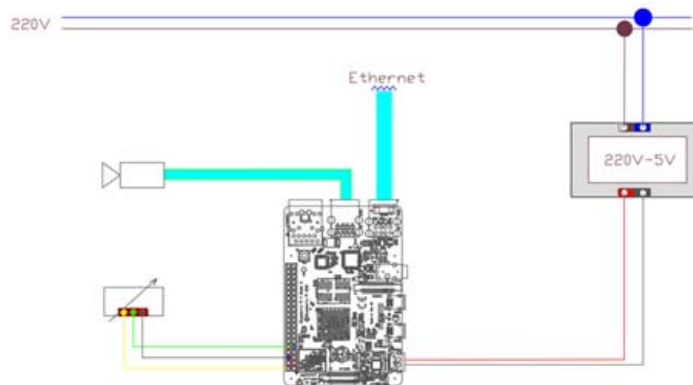


Рис. 3. Схема подключения оборудования

Создание пользовательского интерфейса для управления системой

Одной из простых и удобных форм взаимодействия с различного рода цифровыми системами являются чат-боты [4]. Это простые программы, осуществляющие обработку алгоритмов и сценариев внутри какого-либо сервиса, мессенджера или социальной сети. Возможность создания персональных чат-ботов уже давно предоставляют все крупные интернет-порталы и сервисы, например Вконтакте, *Telegram*, *Instagram*. В данной системе в качестве взаимодействия с человеком был выбран мессенджер *Telegram* [6], что позволило не писать собственную веб-страницу, прописывать логику взаимодействия между контроллером и сервером посредством *http* запросов, и искать сервер для хостинга страницы. Самым главным параметром при создании системы был понятный и «дружелюбный» визуальный интерфейс. Чтобы не приходилось каждый раз прописывать команды, в этом помогла возможность создания встроенной клавиатуры в *Telegram*.

У пользователя всегда есть возможность сделать фотографию, записать видео и включить режим активного сканирования. При запросе фото или видео через чат-бота, файлы отправляются напрямую в *Telegram*. При угрозе проникновения все файлы загружаются в облако и пользователю приходит ссылка на созданную в хранилище папку. Такая реализация обусловлена тем, что пользователь через чат-бота запрашивает одно или максимум несколько фото, которые скорее всего не будут нести какой-либо полезной информации, и сохранять их в облако, тем самым уменьшая его полезный объем, не представляется разумным. В то же время, при угрозе проникновения, система будет делать фото и видео постоянно, учитывая время заложенной задержки, и уже эти файлы, обладающие информационной ценностью, будут храниться в облаке. Также были предусмотрены функции рассылки сообщений с предупреждениями о проникновении, выключения уведомлений и остановки сканирования. По ходу взаимодействия с пользователем чат-бот запоминает сделанный выбор и предлагает пользователю только актуальную клавиатуру. То есть, если был включен режим активного сканирования, то в следующей клавиатуре будет кнопка остановки режима сканирования. Рисунок 4 иллюстрирует пример работы чат-бота.

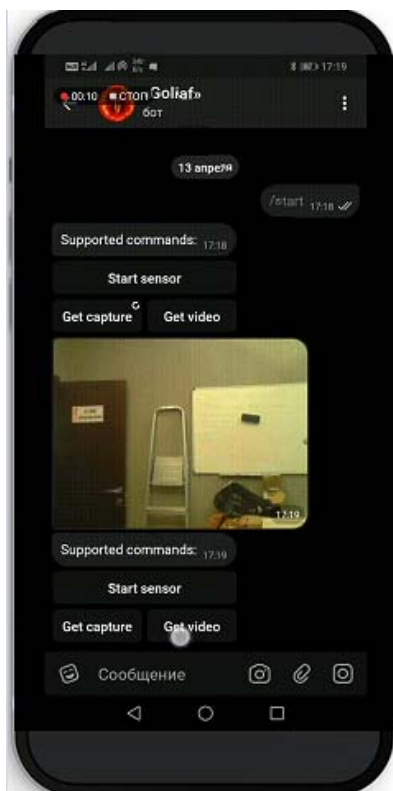


Рис. 4. Функционирование чат-бота

На программном уровне режим активного сканирования представляет собой функцию, запускаемую в новом потоке, которая «слушает» датчик движения с небольшой задержкой. Таким образом, чат-бот интерфейс при включенном режиме сканирования становится многопоточной программой.

Поскольку *Telegram* это многопользовательский сервис, и доступ к любому чат-боту можно получить, зная его *ID* и вбив его в поисковик, было крайне важно устранить угрозу безопасности системе и конфиденциальным данным. Чтобы устранить данную угрозу была прописана фильтрация пользователей. Бот будет отвечать только авторизованным пользователям – хозяевам. Фильтрация проходит по *ID* пользователя *Telegram*. С незарегистрированными пользователями общение не начнется – бот напишет сообщение, уведомляющее злоумышленника о закрытом доступе.

Реализация алгоритмов компьютерного зрения

Для реализации системы распознавания лиц была выбрана *python*-библиотека *OpenCV* [7]. Выбор данной библиотеки обуславливается несколькими факторами: наличие подробной документации, активное сообщество разработчиков, множество встроенных функций и большое количество примеров кода с самым разным функционалом, в т.ч. возможность сохранения изображений встроенными методами.

Одной из подзадач, решенных в процессе создания системы, был сбор фотографий владельцев (авторизованных пользователей) для обучения встроенного в библиотеку *OpenCV* классификатора. С этой целью был написан отдельный скрипт, хранящийся в каталоге проекта. При его запуске пользователю предлагается ввести с клавиатуры свое имя. В случае если пользователя с введенным именем не создано, в директории проекта создается папка с именем пользователя в качестве названия. Затем производится захват видеопотока с камеры, его конвертация в градации серого и вывод на экран в окне приложения. В этом видеопотоке при помощи каскадов признаков Хаара [8] происходит обнаружение фронтальной плоскости лица. Если таковая обнаружена, на экране вывода видеоизображения область лица обводится рамкой синего цвета. При нажатии пользователем клавиши пробела происходит сохранение области, в которой находится фронтальная плоскость лица в *.jpg* файл разрешения $100*100$. Здесь есть риски получить снимок без распознанного лица, если пользователь слишком сильно повернул голову. В дальнейшем планируется интегрировать механизм защиты от неподходящих снимков. При необходимости изображение растягивается до требуемого разрешения.

Блок-схема на рисунке 5 иллюстрирует скрипт сбора данных.

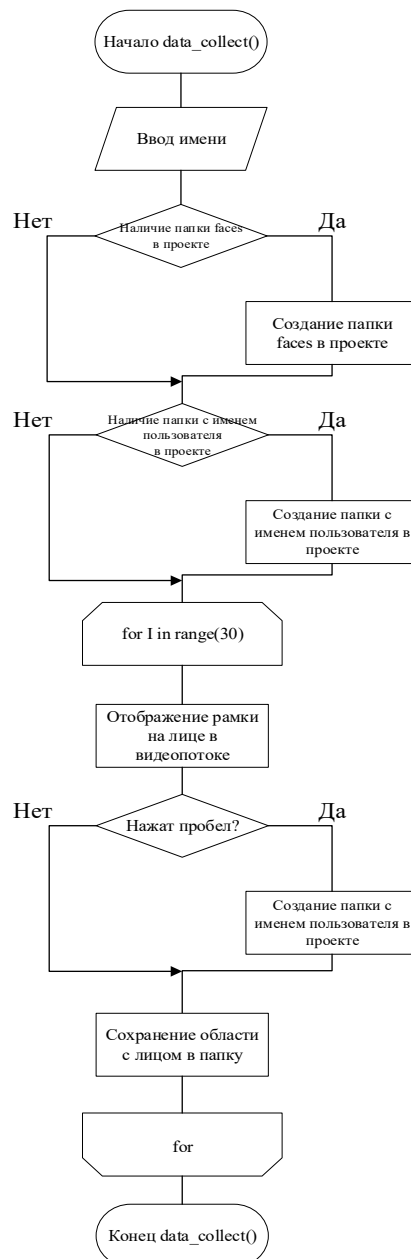


Рис. 5. Блок-схема скрипта сбора данных

Для определения неавторизованного пользователя была реализована функция *detect_face()* в файле *main.py* проекта. Она возвращает логическое значение *True* в случае доступа на территорию неизвестного лица и *False* - в случае доступа авторизованного пользователя. При вызове функции происходит захват видеопотока с камеры и каждый кадр конвертируется в градации серого. Затем покадрово происходит поиск лиц при помощи метода *detectMultiScale* объекта классификатора из библиотеки *openCV*. Для каждого распознанного лица (лица являются кортежами из четырех значений: координат *x* и *y* угла прямоугольника, в котором расположено лицо на изображении, а также ширины и высоты этого прямоугольника) производится определение вероятности его принадлежности авторизованному пользователю.

Если хотя бы одно из лиц на изображении принадлежит авторизованному пользователю с вероятностью более 50%, метод возвращает логическое значение *True*, и метод *sensor_job*, из которого вызывается *detect_face()*, не инициирует оповещение пользователя. В противном случае метод возвращает *False*, и метод *sensor_job()* инициирует оповещение пользователя, в случае если прошло хотя бы 30 секунд со времени последнего оповещения.

Дополнительно учитывается отключение уведомлений (реализовано при помощи чат-бот интерфейса в *Telegram*). Если пользователь пожелал отключить уведомления, то и уведомление о незаконном проникновении он не получит. Таким образом, можно избежать срабатывания системы при приходе гостей, доставке товаров курьером до двери и настройке системы сторонним специалистом, что в целом обеспечивает гибкость пользовательской логики системы.

Реализация алгоритмов загрузки данных в облачное хранилище

Для хранения изображений, сделанных системой при подозрении на проникновение, был использован продукт *Drive* от *Google*. Взаимодействие с облачным хранилищем производилось при помощи *Google Drive API* [9], предоставляющего набор методов, реализующих *http*-запросы для отправки, скачивания, изменения и чтения файлов с самого *Google Drive*. Взаимодействие пользователя с *Google Drive* через *API* показано на рисунке 6.

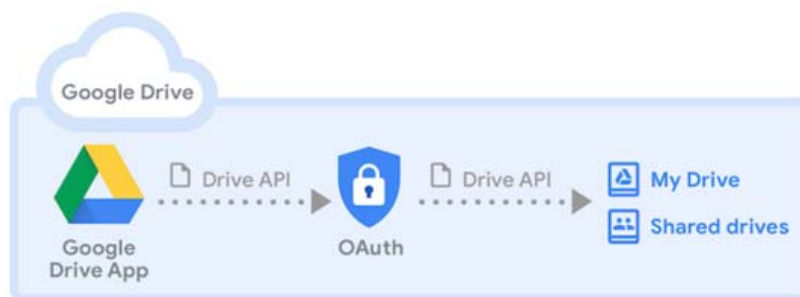


Рис. 6. Взаимодействие с *Google Drive*

Для авторизации *Python*-приложения, использующего *API*, необходимо было создать приложение в *Google console* [10] и включить в его состав *Google drive api*. После этого следовало сгенерировать токен для взаимодействия, который и использовался в приложении.

В коде приложения происходила инициализация в *Google Drive* при первом запуске файла, затем загрузка файлов по необходимости.

Заключение

В результате разработки было интегрировано множество современных технологий и сервисов: от аналитики данных и алгоритмов компьютерного зрения до программирования одноплатных компьютеров и чат-бот интерфейсов.

Система имеет потенциал дальнейшего масштабирования и развития. Так, в план разработки полноценной системы “Умного” дома входит: добавление контроллера промышленного уровня, переход на стандарт беспроводной организации сети *Zigbee*, повышение сложности и точности аналитики данных и перенос вычисления в облачный сервис, добавление нового оборудования и расширение функционала системы.

Несмотря на отсутствие возможности управления доступом, система представляет собой готовое решение в области СКУД, готовое к применению в сфере автоматизации и “Умного” дома.

Литература

1. Как начать проект Интернета вещей? [Электронный ресурс]. Режим доступа: <https://www.osp.ru/os/2015/02/13046278> (Дата обращения: 10.05.2021).
2. Орлов В.Г., Тюмин С.Г. Стандарты беспроводной связи для системы умный дом. Телекоммуникации и информационные технологии. 2020. Т. 7. № 2. С. 20-28.

3. ГОСТ Р 51241-2008 «Средства и системы контроля и управления доступом. Классификация. Общие технические требования. Методы испытаний.» [Электронный ресурс]. Режим доступа: https://expert-01.com/assets/images/uslugi/kontrol_dostupa/GOST%20R%2051241-2008.pdf (Дата обращения: 10.05.2021).
4. Raspberry Pi Tech Specs [Электронный ресурс]. Режим доступа: <https://www.raspberrypi.org/products/raspberry-pi-4-model-b/specifications/> (Дата обращения: 10.05.2021).
5. Ураев Д.А. Классификация и методы создания чат-бот приложений // Наука, техника и образование. 2019. №9(62). С. 1-4.
6. Telegram APIs [Электронный ресурс]. Режим доступа: <https://core.telegram.org/> (Дата обращения: 10.05.2021).
7. OpenCV [Электронный ресурс]. Режим доступа: <https://opencv.org/> (Дата обращения: 10.05.2021).
8. Хорошева А.А., Федько М.Д., Юсупов А.Р. Применение каскада Хаара для распознавания птиц // Молодой исследователь Дона. 2020. №2(23). С. 2-3.
9. Google Drive API [Электронный ресурс]. Режим доступа: <https://developers.google.com/drive/api/v3/about-sdk> (Дата обращения: 10.05.2021).
10. Google Console [Электронный ресурс]. Режим доступа: <https://console.cloud.google.com/?hl=ru> (Дата обращения: 10.05.2021).

DEVELOPMENT OF MULTI-PURPOSE SITUATION ASSESSMENT SYSTEM

Kirill V. Fursaev,
Student MTUCI, Moscow, Russia,
Kirill.fursaev@gmail.com

Nikita D. Saveliev,
Student MTUCI, Moscow, Russia,
nikita.saveliev@lm.net.ru

Nikita V. Belov,
Assistant of the Department of ISUiA, MTUCI, Moscow, Russia,
n.v.belov@mtuci.ru

Abstract

This article is devoted to the development of a multi-purpose intelligent situation assessment system, based on Raspberry Pi and intelligent data processing technologies. To solve the problem of distant system control it's planned to integrate computer vision algorithms and to create user interface with the use of Telegram chat bot tools.

Keywords: Smart home, automatization, face recognition, Raspberry Pi, Python, OpenCV, Google drive.

АНАЛИЗ ПОДХОДА К РАЗРАБОТКЕ ПРИЛОЖЕНИЙ С «ЧИСТОЙ» АРХИТЕКТУРОЙ

Губин Алексей Сергеевич
студент МТУСИ, Москва, Россия,
alecmei.gubin@yandex.ru

Тугова Наталья Владимировна
заведующий кафедрой БИ, к.т.н., доцент, МТУСИ, Москва, Россия
e-natasha@mail.ru

Аннотация

В работе проанализирована так называемая «чистая» архитектура приложений. На начальных этапах разработки приложений может показаться, что можно упустить из виду недочеты в архитектуре приложений в пользу скорости разработки. Но такой подход может обернуться большими проблемами при увеличении масштаба системы. Чтобы не допустить этого, надо изначально придерживаться стандартов и применять лучшие практики. Показано как проектировать расширяемую архитектуру на примере Flutter-приложения.

Ключевые слова: чистая архитектура, Flutter, BLoC, архитектура управления состоянием, SOLID, DRY, KISS, YAGNI, проектирование архитектуры

Введение

Разработка программного обеспечения сопряжена с решением целого ряда задач, одна из таких задач – проектирование архитектуры. Архитектура – это структура компонентов, их взаимосвязи, принципы и направления развития, определяющие их разработку и эволюцию [1, 9-14]. Компонент (в информационных системах) – программный модуль, реализующий некоторую функцию или набор функций, который решает определенные подзадачи в рамках общих задач системы [2]. Некоторые разработчики пренебрегают проектированием архитектуры и со временем начинают замечать в своей работе, такое явление как «синдром следующего утра» [3] – когда то, что ранее функционировало исправно перестает работать из-за изменений в другой части системы. Со временем неправильно спроектированная архитектура начинает создавать проблемы для команды разработчиков. Внесение любых изменений требует все больших трудозатрат, которые растут с большой скоростью. Это ведет к необходимости расширения штата программистов и, соответственно, к увеличению затрат на разработку программного продукта. Со временем затраты могут превысить доходы, и дальнейшая разработка продукта станет невыгодной для компании, ее придется приостановить.

ИТ-компании должны следить за качеством своих программных продуктов и вовремя принимать меры для предотвращения увеличения расходов на разработку.

Преимущества хорошо продуманной архитектуры

Несмотря на то, что разработчики не занимаются расчетом доходов и расходов компании, они, устремившись на работу, берут ответственность за техническую составляющую системы. Экономия времени в краткосрочной перспективе ценой «чистоты» программного кода может привести к ряду проблем для разработчика. Требования к программному продукту со стороны заказчика могут меняться, а внесение изменений будет затруднено накопленными проблемами в программном коде, что может привести к значительному увеличению объема работ.

С другой стороны, сама компания заинтересована в продуманной архитектуре системы. Чтобы программный продукт был прибыльным, время на его разработку и поддержку должно быть минимальным. Компания может ощущать последствия накопления архитектурных проблем очень быстро, это может выражаться в таком показателе, как время на подготовку нового обновления. Когда над одним проектом трудится множество разработчиков, они могут делиться на команды, занимающиеся отдельными компонентами, но бывает так, что изменения пересекаются и необходимо разрешать конфликты в программном коде. Для разрешения таких конфликтов и сборки стабильной версии проекта в компаниях выделяют один день в неделю, остальные четыре дня (при пятидневной рабочей неделе) остаются на разработку. Но когда проект расширяется, а архитектура не удовлетворяет растущим требованиям бизнеса, время на разрешение таких конфликтов увеличивается и вместо одного дня, это может занимать более двух дней. Фактически, при этом отдельно взятый разработчик занят написанием кода менее трех дней в неделю, вместо заложенных четырех. Программисты просто перестают быть эффективными, и основным занятием становится разрешение конфликтов, вместо разработки нового функционала.

Какую архитектуру можно считать хорошей?

Хорошая архитектура может показаться недостижимой, однако стоит понимать, что все принципы разработки программного обеспечения являются скорее рекомендациями, чем правилами. Хорошая архитектура – такая архитектура, которая легко поддается изменениям, требует малого количества трудозатрат и сохраняет этот показатель на относительно невысоком уровне в течение эксплуатации системы [3]. Если трудозатраты на изменение системы увеличиваются с каждым новым обновлением, то такая система имеет плохую архитектуру.

Для того чтобы архитектура оставалась хорошей, разработчики сформулировали принципы написания программного кода. Рассмотрим самый популярный набор принципов – *SOLID*, акроним от пяти следующих принципов [3]:

- *S* – принцип единственной ответственности (*single responsibility principle*);
- *O* – принцип открытости/закрытости (*open/closed principle*);
- *L* – принцип подстановки Барбары Лисков (*Liskov substitution principle*);
- *I* – принцип разделения интерфейсов (*Interface segregation principle*);
- *D* – принцип инверсии зависимостей (*Dependency inversion principle*).

Каждый из этих принципов накладывает определенные ограничения на работу программиста, но в совокупности помогает экономить время на разработку.

Помимо классических принципов, выделяют еще несколько популярных [4]:

- *KISS* – делай короче и проще (*keep it short and simple*).

Принцип заключается в том, что большинство систем работают лучше, если они остаются простыми. Не следует использовать более сложные средства, если в этом нет необходимости.

- *DRY* – не повторяйся (*don't repeat yourself*).

Принцип гласит, что повторение кода необходимо снижать к минимуму. Изменение единственного элемента системы не должно вести к необходимости изменения других.

- *YAGNI* – вам это не понадобится (*You aren't gonna need it*).

Принцип декларирует, что необходимо отказываться от излишней функциональности, так как это занимает избыточно много времени и накладывает ограничения на то, что будет сделано в будущем.

Интересно, что во многих названиях принципов встречаются акронимы. Эти принципы довольно очевидны, однако их знание является обязательным для разработчика. Так же есть принципы, описывающие как должны взаимодействовать компоненты системы. Некоторые из них являются модификациями принципов *SOLID*. Под компонентами имеется в виду некоторые части большой системы, объединенные одной идеей. Это такие принципы как [3]:

- *REP* (*Reuse equivalence principle*) – принцип эквивалентности повторного использования.

Принцип гласит, что повторно используемые компоненты должны иметь версию выпуска и гарантировать оповещение об их изменении в будущем.

- *CRP* (*common reuse principle*) – принцип совместного использования.

Принцип утверждает, что связанные по своей идее классы должны объединяться в компоненты и изменяться одновременно. Если классы изменяются по разным причинам, то их необходимо разделить.

- *CCP* (*common closure principle*) – принцип согласованного изменения.

Принцип гласит, что если классы настолько тесно связаны, что они изменяются вместе и по одним и тем же причинам, то их стоит поместить в один компонент.

Ранее был упомянут термин «синдром следующего утра», но как предотвратить его появление? Такое явление происходит вследствие нарушения принципа ацикличности зависимостей (*acyclic dependencies principle*).

На рисунке 1 приведен пример циклического ориентированного графа. Если в графе можно из отдельно взятого компонента пройти по направлению стрелок зависимостей и вернуться в этот же компонент, то такой граф называют циклическим.

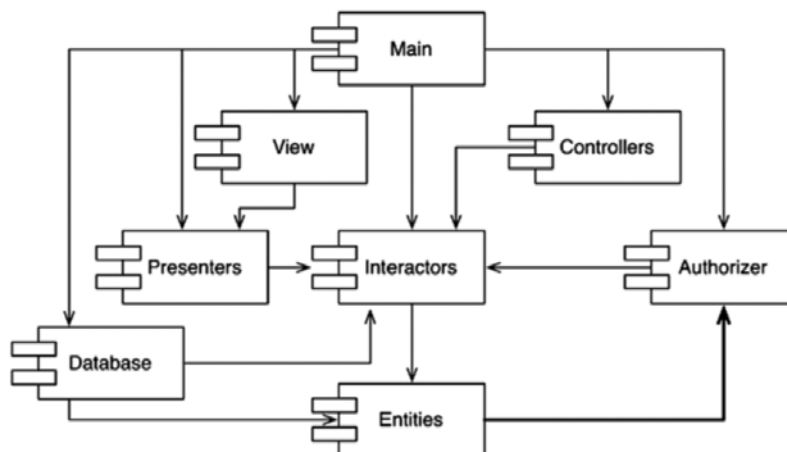


Рис. 1. Ориентированный граф с циклами [3]

Цикличность усложняет выпуск новых версий и тестирование системы. Так, разработчики компонента *Interactors* для выпуска новой версии должны проверить совместимость с компонентом *Entities*, но *Entities* зависит от *Authorizer*, который, в свою очередь, зависит от *Interactors*. Взаимодействие этих трех компонентов образует цикл и, как следствие, тесную связь. Значит, ошибки в таких компонентах будут возникать особенно часто. Программисты, занимающиеся разработкой компонента *Authorizer*, при публикации новых версий могут внести такие изменения, что команда разработчиков компонента *Interactors* может выявить неправильное поведение программы, несмотря на то что *Authorizer* зависит от него, а *Interactors* зависит от *Authorizer* лишь косвенно, через компонент *Entities*.

Чтобы разорвать такой цикл, Роберт Мартин, автор книги «Чистая архитектура» [3], предлагает применить принцип инверсии зависимостей и создать компонент, от которого будут зависеть компоненты *Entities* и *Authorizer* (рис. 2).

Таким образом, разрывается циклическая зависимость, созданием компонента *Permission*, от которого зависят *Authorizer* и *Entities*. Создан «устойчивый» компонент. Такие «устойчивые» компоненты сложны для изменения, а самые неизменяемые компоненты в системе – абстрактные и безопасные.

Все представленные принципы грамотно сформулированы и обоснованы, однако следовать им необязательно, это лишь рекомендации. При разработке необходимо избегать избыточного проектирования, чтобы не создавать сущности, которые никогда не пригодятся и, возможно наложат ограничение на разработку в будущем. Одна из целей архитектора программного обеспечения – как можно дольше оставлять варианты для изменения системы.

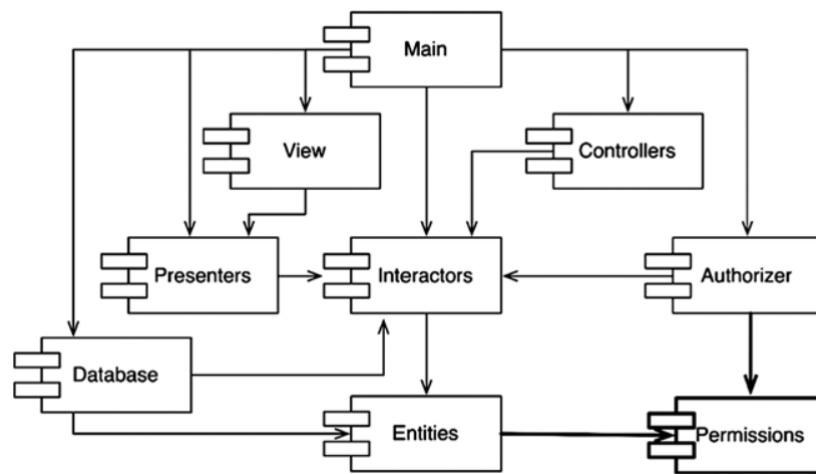


Рис. 2. Граф с разорванной циклической зависимостью [3]

Понятие «чистой» архитектуры

В основе «чистой» архитектуры лежат слои, зависимости которых направлены в центр (рис. 3) [3].



Рис. 3. Обобщенная структура приложения с «чистой» архитектурой

На рисунке стрелками показано направление зависимости слоев. Всего продемонстрировано 4 слоя, однако Роберт Мартин утверждает, что число слоев может различаться, и зависит от требований, предъявляемых системе. Помимо этого, в правом нижнем углу показано, как слои взаимодействуют между собой на примере контроллера и презентатора, где контроллер – это класс, который оперирует объектами и событиями, а презентатор – обертка для пользовательского интерфейса, куда поставляются все необходимые данные для его отрисовки. С помощью интерфейсов «порт вывода варианта использования» и «порт ввода варианта использования» контроллер становится независимым от презентатора. Поток управления при этом направлен от контроллера к презентатору, однако контроллер не имеет зависимости от презентатора.

Программное обеспечение не должно зависеть от деталей, например базы данных, фреймворка, варианта отображения данных. То есть даже если архитектору изначально известно, какие технологии будут использоваться, он должен сделать вид, что не знает об этом и спроектировать архитектуру независимую от деталей. Также на картинке изображены концентрические круги, где написаны названия слоев. Чем ближе слои к центру, тем выше уровень абстракций в компонентах.

На слое сущностей описываются критические бизнес-правила, которые будут изменяться редко, они независимы от всех остальных слоев. Под критическими бизнес-правилами подразумеваются такие правила, которые будут существовать всегда, вне зависимости от того, как обрабатывается информация: с помощью компьютера или вручную. На следующем слое находятся варианты использования. Варианты использования описывают бизнес правила на уровне приложения, т. е. какие действия может совершать пользователь и каким образом это происходит.

За вариантами использования следуют контроллеры, презентаторы, а также шлюзы, в которых описаны конкретные способы получения данных. Внешний круг представляет собой детали, которые изменяются чаще всего: способ ввода и вывода информации, устройство, пользовательский интерфейс.

Архитектура управления состоянием приложения

Помимо архитектуры приложений, есть **архитектура управления состоянием**, или **менеджер состояний** – это способ хранения и изменения состояния представления, отделяющий бизнес-логику от интерфейса [5]. Одной из таких архитектур является *BLoC* (*Business Logic Component – компонент бизнес-логики*). Она пользуется популярностью среди *Flutter*-разработчиков и была разработана в *Google* [5]. *BLoC* позволяет отделить бизнес-логику от представления данных, что облегчает читаемость кода и дает возможность простого изменения логики и пользовательского интерфейса. Архитектура *BLoC* ориентирована на быструю работу приложений. Это значит, что потоки ввода и вывода асинхронны. Класс *BLoC* принимает события (заявки) от интерфейса и обрабатывает их, после чего может публиковать в канал (на который подписан интерфейс) изменения, произошедшие в его состоянии. Интерфейс при получении изменений, обновляет отображение и транслирует актуальные данные [5]. В классе *BLoC* можно реализовать, например, обращение к внутренним слоям «чистой архитектуры». Архитектура управления состоянием по отношению к «чистой» архитектуре приложений – это такая же деталь, которую можно заменить.

Пример разработки приложения на Flutter с «чистой архитектурой»

В рассмотренном ниже примере будет реализована «чистая» архитектура и менеджер состояний *BLoC*. Разработка велась на *Flutter* – фреймворке для разработки кроссплатформенных приложений на языке программирования *Dart*.

Приложение позволит просматривать снимки *NASA*, которые были созданы с помощью их марсоходов: *Spirit*, *Opportunity*, *Curiosity*, *Perseverance*. Диаграмма вариантов использования приложения приведена на рисунке 4.

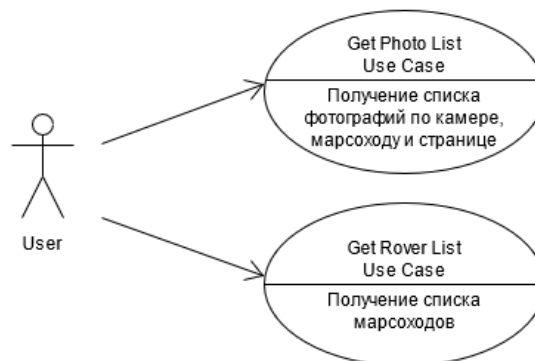


Рис. 4. Диаграмма вариантов использования приложения

У пользователя есть возможность получения списка марсоходов и выбора фотографии, снятой заданным марсоходом на определенную дату. Для этого *NASA* предоставляет открытый интерфейс *REST API*.

На рисунке 5 приведена обобщенная архитектура разработанного приложения и направление зависимостей.

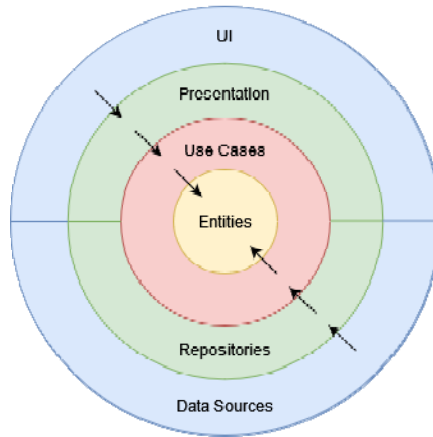


Рис. 5. Обобщенная архитектура разработанного приложения

Стрелки на схеме указывают направление зависимостей, которые направлены вовнутрь, в сторону абстракций. Имеем четыре слоя: *Entities* – сущности, используемые в данной системе, *Use Cases* – варианты использования нашего приложения, *Presentation* – представления-изоляторы бизнес-логики от интерфейса пользователя, *Repositories* – репозитории, с помощью которых организуется способ получения данных и их обработка, *UI* – пользовательский интерфейс, *Data Sources* – источники данных (в данном случае *API NASA*).

На следующем этапе зарисуем *UML*-диаграмму, отображающую зависимости между классами и компонентами системы (рис. 6).

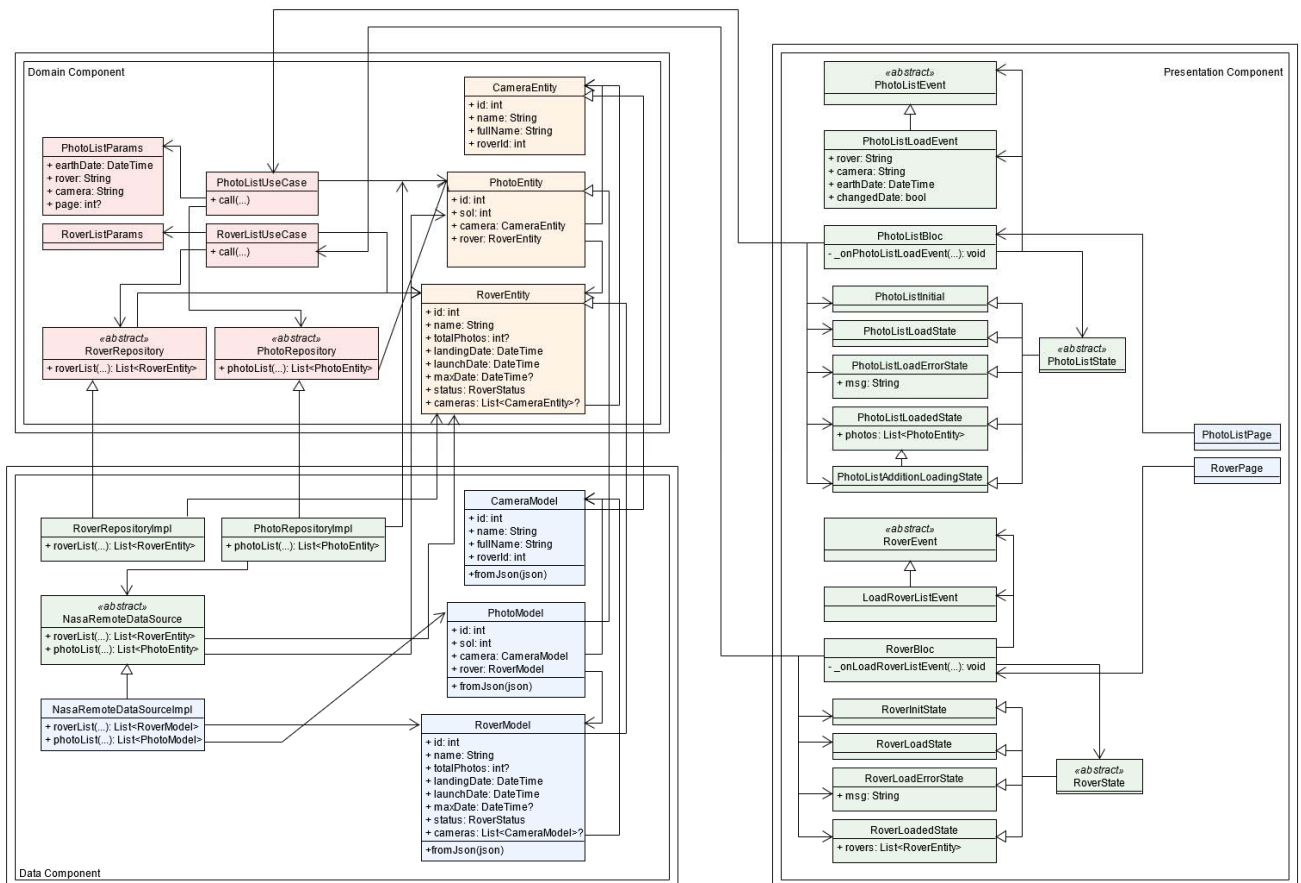


Рис. 6. *UML*-диаграмма приложения

Отдельные компоненты были выделены с помощью двойных линий. Можно увидеть, что стрелки зависимостей между компонентами направлены в одном направлении.

Почему компоненты в точности не совпадают со слоями на концентрических кругах рисунка 5? Классы, которые имеют крепкие связи, были выделены в одни компоненты, но при этом, каждый компонент делится на слои. Компонент *Data*, работающий с данными, имеет два слоя: *Repository* и *Data Sources*. Компонент *Presentation*, отвечающий за

пользовательский интерфейс, содержит в себе слои *Presentation* и *UI*. Компонент *Domain*, включающий в себя бизнес-логику, состоит из слоев *Entity* и *Use Cases*.

Особый интерес на этой схеме вызывает отношение *RoverRepository*, *PhotoRepository* и *RoverRepositoryImpl*, *PhotoRepositoryImpl* соответственно. Здесь применен принцип инверсии зависимостей, чтобы избежать зависимости компонента *Domain* от компонента *Data* и не создавать цикла в зависимостях. Это некий «мост», соединяющий два компонента.

Также интересны классы *Entities* и *Models*, являющиеся объектами передачи данных в системе. Это следствие реализации шаблона проектирования *Repository* (рис. 7) [4].



Рис. 7. Шаблон проектирования *Repository*

В первоначальном виде шаблона проектирования *Repository* класс *Model* должен преобразовываться в *Entity* вспомогательным классом – его называют *Mapper*. Однако в данном случае класс *Model* почти не отличается от *Entity*, и было принято решение наследовать *Model* от *Entity*.

Так же на *UML*-диаграмме можно увидеть реализацию архитектурного паттерна *BLoC*. Схематичное изображение паттерна изображено на рисунке 8.

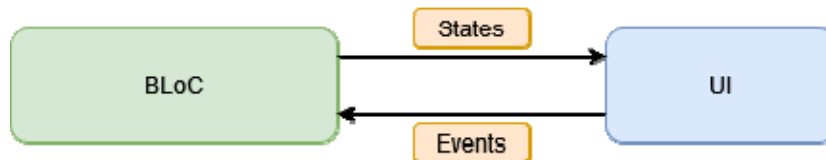


Рис. 8. Архитектурный шаблон проектирования *BLoC*

Класс *BLoC* имеет входной и выходной потоки. На вход поставляются события, которые направляются в обработчики. Обработчики, в свою очередь, могут менять состояние *BLoC* и публиковать изменения в поток с состояниями приложения, *UI* подписывается на изменение состояний и запускает процесс повторной отрисовки с учетом нового состояния.

Этот архитектурный шаблон проектирования был реализован на языке *Dart*, он представляет собой абстрактный класс, от которого будут реализованы все *BLoC*-классы, программный код которого приведен в листинге 1.

Листинг 1 – Абстрактный класс *BlocBase*
`import 'dart:async';`

```
//базовый класс для всех bloc
abstract class BlocBase<State, Event> {
  //текущее состояние bloc'a
  State _state;

  //getter, состояния
  get state => _state;
  //конструктор класса, принимает начальное состояние bloc'a
  BlocBase(this._state);

  //контроллер потока для событий
  final _eventController = StreamController<Event>.broadcast();

  //воронка для публикации события
  Sink<Event> get inEvent => _eventController.sink;

  //поток для отслеживания событий
  Stream<Event> get _outEvent => _eventController.stream;

  //контроллер потока состояний
  final _stateController = StreamController<State>.broadcast();
```

```

//воронка для публикации новых состояний блока
Sink<State> get _inState => _stateController.sink;

//поток для отслеживания новых состояний
Stream<State> get outState => _stateController.stream;

//функция для добавления слушателя событий
void on<E extends Event>(Function(E) handler) {
  _outEvent.listen((event) {
    if (event is E) {
      handler(event);
    }
  });
}

//функция для добавления нового события в bloc
void emit(State state) {
  _state = state;
  _inState.add(state);
}

//функция, вызываемая при уничтожении виджета, закрывает потоки
void dispose() {
  _eventController.close();
  _stateController.close();
}

```

Помимо этого, для фреймворка *Flutter* был создан виджет предоставления доступа к классу *BLoC* в дереве виджетов – *BlocProvider* [6], программный код которого приведен в листинге 2. Виджет – это основной строительный блок пользовательского интерфейса приложений [7].

Листинг 2 – Виджет для предоставления доступа к классу *BLoC*

```

import 'package:flutter/material.dart';
import 'package:nasa_space_data/common/bloc_base.dart';

//виджет для поставки класса bloc между вложенными виджетами
class BlocProvider<T extends BlocBase> extends StatefulWidget {
  //дочерний элемент
  final Widget child;

  //класс для доставки
  late final T bloc;

  BlocProvider({Key? key, required this.child, bloc})
    : bloc = bloc(), super(key: key);

  @override
  State<BlocProvider> createState() => _BlocProviderState();

  //функция для получения блока в дереве виджетов
  static T of<T extends BlocBase>(BuildContext context) {
    BlocProvider<T> provider =
      context.findAncestorWidgetOfExactType<BlocProvider<T>>()!;
    return provider.bloc;
  }
}

//состояние виджета
class _BlocProviderState extends State<BlocProvider<BlocBase>> {
  @override
  Widget build(BuildContext context) {
    return widget.child;
  }
}

```

```

@override
void dispose() {
  //вызов закрытия потоков
  widget.bloc.dispose();
  super.dispose();
}
}

```

Дерево директорий приложения продемонстрировано на рисунке 11.

Такая информационная система легко расширяема. Заменить компоненты, находящиеся на внешних слоях, не составляет сложности.

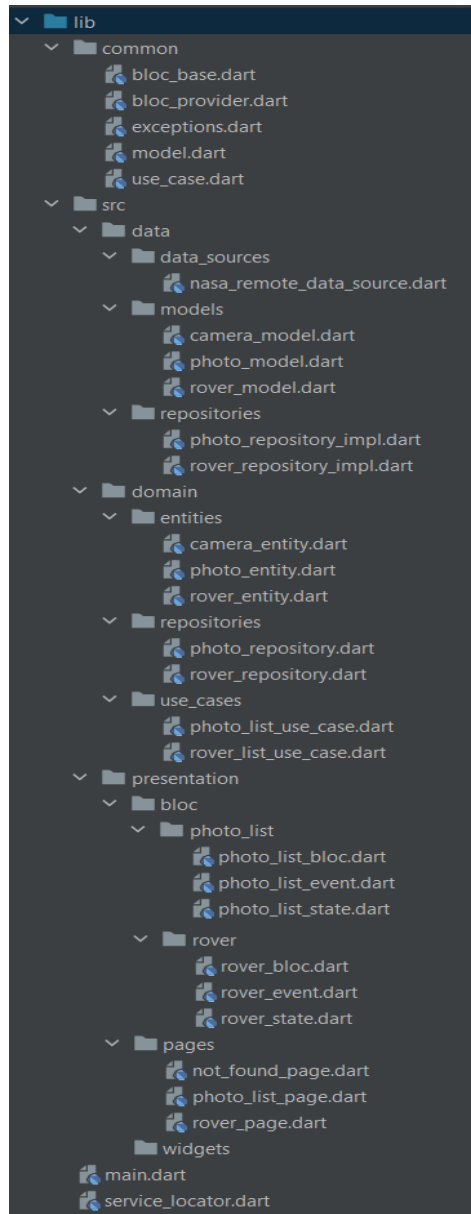


Рис. 11. Дерево директорий разработанного приложения

К примеру, если *NASA* захочет поменять способ получения данных с *REST API* на *Long Poll API*, то изменения не затронут слой бизнес-логики, они будут изолированы в компоненте, в котором находятся, так как являются лишь реализацией интерфейсов из более абстрактных слоев.

Интерфейс программы продемонстрирован на рисунке 12.

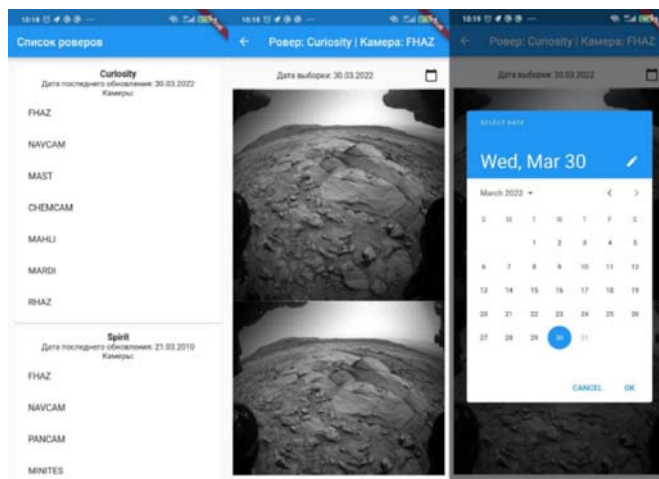


Рис. 12. Демонстрация возможностей приложения

Полный исходный код размещен на ресурсе *github* [8].

Заключение

Проектирование и поддержание «чистоты» архитектуры важно как для заказчиков, так и для разработчиков. Хорошая архитектура в долгосрочной перспективе экономит большое количество времени и денег. Для того чтобы архитектура была «хорошей», необходимо придерживаться как подходов к построению «чистой архитектуры», так и общепринятых принципов написания программного кода. В этом случае система становится легко расширяемой на внешних слоях. Например, любые изменения в требованиях к интерфейсу системы не затронут слой бизнес-логики, и потребуют минимальных усилий разработчиков.

Литература

1. *Воронцов Ю. А.* Распределенные информационные системы: для направления 09.03.03. учебно-методическое пособие для бакалавров. М.: МГУСИ, 2016. 16 с.
2. *Советов Б. Я.* Архитектура информационных систем. – 2012. – 288 с.
3. *Мартин Р.* Чистая архитектура. Искусство разработки программного обеспечения. СПб.: Питер, 2022. 352 с.
4. *Millett S.* Professional ASP. NET Design Patterns. – John Wiley & Sons, 2010. – 1638 с.
5. *Павленко Андрей.* Выбираем архитектуру для Flutter-приложения. [Электронный ресурс] – режим доступа: <https://otus.ru/nest/post/1918/> (Дата обращения: 05.04.2022).
6. *Магин Ярослав.* Разделение бизнес-логики и UI во Flutter с помощью BLoC-архитектуры. [Электронный ресурс] – режим доступа: <https://inostudio.com/blog/articles-develop/razdelenie-biznes-logiki-i-ui-vo-flutter-s-pomoshchyu-bloc-arkhitektury/> (Дата обращения: 05.04.2022).
7. *Алеев А.* Быстрый старт Flutter-разработчика //Издательские решения». 2019. 246 с.
8. Исходный код разработанного приложения [Электронный ресурс]. Режим доступа: https://github.com/WebMad/nasa_space_data (Дата обращения: 05.04.2022).
9. *Тутов А.В., Тугова Н.В., Ворожцов А.С., Андреев И.А.* Многокритериальная оптимизация размещения виртуальных машин по физическим серверам в облачных центрах обработки данных // Т-Comm: Телекоммуникации и транспорт. 2021. Т. 15. № 1. С. 28-34.
10. *Kirov D.E., Toutova N.V., Vorozhtsov A.S., Andreev I.A.* Feature selection for predicting live migration characteristics of virtual machines // Т-Comm. 2021. Т. 15. № 7. С. 62-70.
11. *Дубельщиков А.А., Тугова Н.В.* Навыки яндекса.алиса: от идеи до реализации // Телекоммуникации и информационные технологии. 2020. Т. 7. № 2. С. 92-97.
12. *Тутов А.В., Тугова Н.В., Ворожцов А.С.* Моделирование процессов распределения ресурсов в облачных центрах обработки данных // Т-Comm: Телекоммуникации и транспорт. 2017. Т. 11. № 4. С. 76-80.
13. *Ворожцов А.С., Тугова Н.В.* Методика использования пакета программ *cloudsim* в учебном процессе // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2017. Т. 6. № 1. С. 13-15.
14. *Ворожцов А.С., Тугова Н.В.* Преподавание комбинированных методов многокритериальной оптимизации в дисциплине теория принятия решений // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2017. Т. 6. № 4. С. 4-8.

ANALYSIS OF THE APPROACH TO DEVELOPING APPLICATIONS WITH A "CLEAN" ARCHITECTURE

Alexey S. Gubin

Student of MTUCI, Moscow, Russia

alecmei.gubin@yandex.ru

Natalia V. Toutova

Head of the Department of BI, Ph. D., Associate Professor, MTUCI, Moscow, Russia

e-natasha@mail.ru

Abstract

At the initial stages of development, it may seem that you can overlook the shortcomings in the application architecture in favor of the speed of development. But such an approach can turn into big problems when the scale of the system increases. To prevent this, it is necessary to initially adhere to standards and apply best practices. The paper analyzes the so-called "clean" application architecture. It shows how to design an extensible architecture using the example of a Flutter application.

Keywords: *clean architecture, Flutter, BLoC, state management architecture, SOLID, DRY, KISS, YAGNI, architecture design*

НОВАЯ СОЦИАЛЬНАЯ СЕТЬ ДЛЯ ЛЮДЕЙ С МЕНТАЛЬНОЙ ИНВАЛИДНОСТЬЮ

*Романова Зоя Ивановна,
студент МТУСИ, Москва, Россия,
romanova.zoya.2002@mail.ru*

*Кудряшова Анастасия Юрьевна,
Старший преподаватель кафедры БИ, МТУСИ, Москва, Россия,
a.i.kudriashova@mtuci.ru*

Аннотация

В работе рассмотрена история создания коммуникатора, предназначенного для детей и молодых людей с ментальной инвалидностью, который, в последствие, трансформировался в особую социальную сеть. Разобраны предпосылки и причины появления программы. Описаны основные этапы разработки, тестирования, продвижения и использования коммуникатора на пути от идеи до реального воплощения. Развитие коммуникатора рассматривается на стыке двух основных направлений работы: программирования и коррекционной педагогики, что представляет особый интерес.

***Ключевые слова:** социальная сеть, web-приложение, создание web-приложения, люди с ограниченными возможностями здоровья, социальный проект, создание мобильного приложения.*

Введение

Число инвалидов в мире постоянно растёт. По данным федерального реестра инвалидов на февраль 2022 года в России 489789 молодых людей инвалидов (18-30 лет) и 730472 ребёнка-инвалида (до 18 лет), что на 27893 человека больше, чем за аналогичный период 2021 года. Значительная часть из них составляет людей с ментальными нарушениями.

В современном обществе декларируются равные права обычных людей и людей с ограниченными возможностями здоровья, однако многие привычные для первых вещи, недоступны для вторых. Это касается и общения в сети. Люди с ментальной инвалидностью обычно не могут читать и писать, многие не могут говорить и имеют множество сопутствующих заболеваний, таких как нарушения зрения, моторики и другие. В числе их, зачастую, социальные сироты, от которых отказались родители, а значит проживают они в сиротских учреждениях, где не имеют собственного телефона, либо имеют телефон без сим-карты. Обычные мессенджеры и социальные сети сложны из-за переизбытка информации, имеют мелкие элементы интерфейса. У них основная часть информации текстовая, а для многих сервисов требуется номер телефона для регистрации. По наблюдениям АНО «Ресурсный центр «Вера. Надежда. Любовь» лишь наиболее «сильные» подопечные могут с помощью наставников зарегистрироваться в социальной сети «В Контакте», однако пользуются они очень ограниченной частью функционала: могут опубликовать фотографию, поставить «лайк» или послушать музыку, но не более того.

В качестве альтернативы специалистами АНО «Ресурсный центр «Вера. Надежда. Любовь» для таких людей был разработан особый коммуникатор ЭМО. С самого начала своего существования интерфейс и функционал ЭМО тестировался детьми с особенностями развития, при этом собиралась и учитывалась обратная связь от педагогов-дефектологов. В данной статье рассмотрена история создания и тестирования особого коммуникатора, который на сегодняшний день не имеет аналогов в России [1].

История создания: от идеи до первой версии

В 2010-х годах в рамках государственной программы «Информационное общество» активно создавались новые информационные ресурсы. Это коснулось и организаций, работающих в сфере социальной защиты. Создавались новые информационные системы, однако большинство из них были скорее для сотрудников, и включали базы достигнутых результатов и аналитику по ним. В 2018 году была задумана программа для самих благополучателей: детей и молодых людей с ментальными нарушениями. Уникальность её - в возможности именно дистанционного общения. В 2018 году уже существовали коммуникаторы для данной категории детей, но их, можно было использовать только как дополнительный помощник при очном общении ребёнка и взрослого. Другая проблема, которую специалисты надеялись решить с помощью программы – наладить дружеские связи между самими детьми. В сиротских учреждениях для детей с особенностями психики явно наблюдалась проблема – дети соревновались за внимание взрослого, но практически не взаимодействовали между собой. Предполагалось, что возможность написать письмо однокласснику и получить ответ подтолкнет эту пару детей продолжить общение очно.

Разработка началась с исследования. Было создано несколько интерактивных интерфейсов на *HTML*, *CSS* и *JavaScript*, похожих друг на друга, с одинаковым функционалом. Они имели две страницы: выбор собеседника и чат. Собеседник выбирался по фото, а общение в чате происходило с помощью специальных картинок-пиктограмм. Интерфейсы были предложены детям и молодым людям с ментальными нарушениями в возрасте от 10 до 19 лет, воспитанникам ЦССВ «Вера. Надежда. Любовь» (ныне ЦСПР «Семь-Я»). Дети были поделены на группы и у каждой был свой вариант интерфейса. Исследование выявило наиболее оптимальное расположение элементов на экране, которое используется и сейчас. Программа получила название ЭМО, сокращённо от «эмодзи», либо «эмоции». Слово понравилось лёгкостью произношения, что полезно людям с нарушениями речи.

Первая версия ЭМО появилась в 2019 году в рамках проекта «ЭМО – коммуникативный тренажёр для особых сирот». В ней было около 20 пользователей. Прорисовка интерфейса лишь отдаленно напоминала нынешнюю. Это было веб-приложение, которое имело 4 станицы: авторизация; пользователи, с которыми есть диалог; пользователи, с которыми диалога нет и сама страница диалога. Доступны были только личные чаты (рисунок 1). Регистрации как таковой не было - аккаунты добавлялись администратором. На тот момент считалось, что так безопаснее, так как была потребность оградить подопечных от посторонних людей из интернета. Была только версия для браузера компьютера (на тот момент её считали приоритетной), при этом 100% пользователей не имели смартфонов. Реализация была на *PHP* с использованием внутренней *CMS* дружественной организации [2].

Это был именно тренажер, который так и позиционировался. Предполагалось, что дети рано или поздно перейдут из него в обычные мессенджеры и социальные сети. Однако к концу проекта стало очевидно, что некоторые дети не смогут перейти в более сложные программы, скорее всего, никогда, но могут освоить ЭМО.

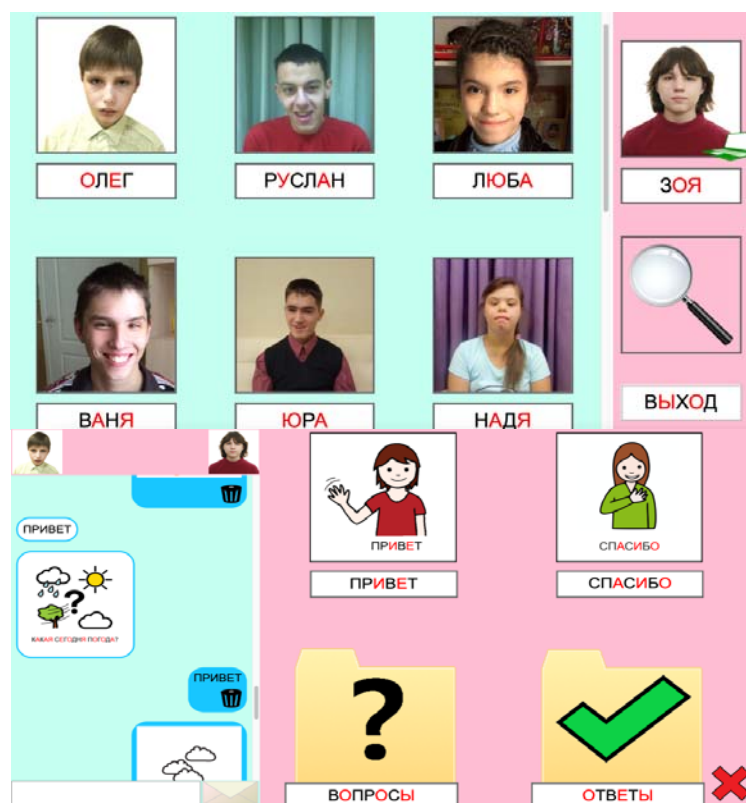


Рис. 1. Страницы выбора собеседника диалога в первой версии

Если эти дети остаются в ЭМО, то было логично оставить там и их более «сильных» одноклассников, чтобы им было с кем общаться, а также добавить их педагогов и других значимых взрослых. Сделать из ЭМО площадку, где люди с ментальными нарушениями могут на равных общаться как между собой, так и с обычными людьми.

ЭМО начал расти, добавлялись новые пользователи, появлялись запросы на новый функционал. Некоторые педагоги хотели использовать ЭМО как инструмент на своих занятиях, и сразу выявились проблемы с производительностью: при трех и более одновременных подключениях происходили сбои. Делать новый функционал при подобной работе было неразумно, а разработчики используемой *CMS* могли посоветовать разве что перестать её использовать, так как ЭМО не был типовым сайтом, под разработку которых она создавалась.

История создания: от первой версии до собственного сервера

Разработка новой версии началась весной 2020 года. Срок был больше, чем на разработку первой, однако ожидалось, что это будет рабочая версия на более длительный период. Из технологий были выбраны: *MongoDB* в качестве базы данных, *Express* в качестве серверной части и *Angular* в качестве клиента. Для начала, это давало возможность

написать абсолютно всё на JavaScript (разве что кроме верстки), а значит, для дальнейшего сопровождения не нужно искать специалистов по другим языкам программирования. *MongoDB* привлекала стоимостью и возможностями масштабирования, а также возможностью подключить *Mongo Atlas* в качестве тестовой базы для разработки, что избавляло разработчиков от необходимости устанавливать базу дополнительно на свои компьютеры и синхронизировать версии в случае необходимости. Помимо этого это наиболее часто используемая база совместно с *Express*. *Express* был выбран как один из популярнейших *Node.js* фреймворков, имеющий много информации в сети, а значит, в случае возникновения каких-то проблем, их будет проще устранить. На *Express* планировалось создать *REST API*, к которому будет подключаться веб-приложение на *Angular*, а в дальнейшем – мобильное приложение, о чём уже начали задумываться [3].

SPA (Single Page Application) архитектура была выбрана поскольку *SEO* не представлялось приоритетной задачей, планировалось продвигать ЭМО через закрытые каналы, имеющие непосредственный выход на целевую аудиторию, но не через сам сайт ЭМО. Идея определенной «закрытости» системы не ушла, хоть и круг пользователей планировалось расширять. При этом преимущества *SPA* были очень кстати: меньше затрат на разработку и высокая скорость работы приложения.

По характеру будущего веб-приложения *SPA* архитектура казалась очень хорошим решением. Для *SPA* требовалось использовать фреймворк (либо библиотеку) на клиенте. Сейчас самая популярная из таких технологий – *React*, но был выбран второй по популярности вариант – *Angular*. Во-первых, в *React* используется специальный *JSX*-синтаксис, это могло бы усложнить перенос вёрстки из первой версии, в то время как к компонентам *Angular* можно подключить шаблон с расширением *.html* и стили *.css*. Во-вторых, *React* – это всё-таки библиотека, которая содержит в себе не так много функционала и требует множество дополнительных пакетов, в то время как *Angular* является полноценным фреймворком со множеством возможностей «из коробки». Например, *Angular* имеет собственный удобный *http*-клиент, в то время как в *React* приходится использовать встроенные в *JavaScript Fetch API*, либо *XmlHttpRequest*, либо дополнительные библиотеки вроде *Axios* для запроса к серверу.

К концу лета 2020 года веб-приложение было готово. Оно полностью повторяло функционал первой версии, однако прорисовка элементов интерфейса была другая, более приятная глазу, чем в первой версии, она используется и сейчас. Далее возник вопрос с хостингом. Сначала казалось, что максимум для пяти одновременно подключенных человек хватит самого обычного хостинга. Были опробованы некоторые варианты и сделан вывод, что попытки разместить на хостинге *node.js* приложение ни к чему хорошему не приводят: для *node.js* в любом случае требуется устанавливать много дополнительных пакетов, а на хостингах имеется большое количество ограничений, обусловленных тем, что размещаемое веб-приложение будет делить ресурсы с другими сайтами на этом же хостинге. Помимо этого, сам *node.js*, а также эти дополнительные пакеты могут конфликтовать с другими сайтами, что вызовет проблемы у всех. Было принято решение арендовать *VPS (virtual private server)*, что по стоимости сравнимо с хостингом, однако при этом предоставляется сервер с *root*-правами, что дает возможность устанавливать туда любое программное обеспечение. Был выбран самый дешевый тариф компании *Formex* (1 ядро, ОЗУ 1 ГБ, SSD Диск 10 ГБ, ОС *Ubuntu 18.04 LTS*, локация в Москве). Рассматривались разные провайдеры и было выбрано несколько с подходящими тарифами и хорошими отзывами, а окончательное решение принималось по простому тесту – одинаковый вопрос в техподдержку каждой компании. Выбрана была та, где дали наиболее понятный и квалифицированный ответ на вопрос за адекватное время.

Было понятно, что хранить на дисковом пространстве *VPS* все пиктограммы, а также загружаемые пользователями файлы нелогично. 10 Гб быстро закончатся, придется переходить на более дорогие тарифы, которые, помимо размера диска увеличивают также и остальные параметры виртуального сервера, но дополнительные мощности пока что не нужны. А когда они потребуются, диска уже будет не хватать – замкнутый круг. Было принято решение хранить файлы в облачном сервисе *AWS S3*. Ведро и загрузка в него были быстро настроены. Для реализации самой загрузки вдобавок к пакету *multer* теперь использовались *multer-s3* и *aws-sdk*.

На *VPS* был установлен *pm2* для настройки переменных окружения и с помощью *pm2* приложение было запущено на *localhost*. После был настроен прокси-сервер *nginx*, который прослушивал порт 80 и переправлял запросы с него на нужный порт *localhost*.

Хотелось красивое и короткое доменное имя и обязательно латиницей – так удобнее подключать корпоративную почту, однако *emo.ru*, *emo.com* и *emo.org* оказались заняты. Из доступных хорошим вариантом показался *emo.su* (какая-то история у него была, но не показалась плохой). К тому же *su* – это зона для РФ и стран постсоветского пространства, а ЭМО как раз планировался как ресурс для русскоговорящих людей и с учётом этого был выбран данный вариант. В начале осени 2020 года была запущена вторая версия ЭМО [4].

Современная версия ЭМО

С одной стороны, больше настолько капитальных перестроек не было, однако ЭМО периода осени 2020 больше напоминал первую версию, чем современную. С запуском новой версии буквально за пару месяцев количество пользователей достигло 100, и многие хотели иметь новый функционал. Новые пользователи очень сильно различались физическими возможностями, и их уровень владения ЭМО был разным. Казалось, каждое добавление новой функции будет усложнять программу, делая ее, в конечном счете, недоступной для наиболее «слабых» ребят – тех самых, для которых изначально программа задумывалась.

В качестве решения были введены отключаемые функции. В современном ЭМО любые функции или элементы интерфейса, которых не было в первой версии (минимальной рабочей), могут быть отключены в настройках пользовате-

ля. Это дает возможность включать наиболее «слабых» ребят и обучать их сначала самым базовым навыкам: увидеть, что пришло письмо, кликнуть, выбрать ответ, отправить. Приобретая базовые навыки, ребенок может переходить к освоению более сложных функций [5]. На данный момент в ЭМО есть 17 отключаемых страниц, функций или элементов интерфейса. Это довольно необычно: в привычных мессенджерах люди скрывают свой сетевой статус от других, здесь же скрывают сетевой статус других от себя. И не только сетевой статус.

Полный список настроек ЭМО представлен в таблице 1.

На рисунках 2, 3 и 4 соответственно представлены страницы выбора: «собеседника», «диалога» и «фотолайки» доступные в современной версии ЭМО.

Таблица 1

Список настроек в современной версии ЭМО

Функция	Комментарий
Онлайны	Отображает зеленый компьютер рядом с фотографией пользователя, который в сети, или зеленый глаз, если пользователь в чате вашем чате. Используется для повышения количества «онлайн-диалогов», когда пользователь написал, и ему сразу же ответили. Отключается как отвлекающий символ
Текстовые подписи	Под каждой пиктограммой есть её тестовый вариант, при клике на который подпись отправляется текстом. Включается умеющим читать
Обозначение, прочитано ли сообщение	Добавляет рядом с каждым сообщением глазик либо перечеркнутый глазик, обозначающий, виде ли собеседник сообщение. Отключается как отвлекающий символ.
Фамилии	Пользователи отображаются с именами и фамилиями, либо только с именами. Не читающим детям фамилии не актуальны.
Голосовые сообщения	Возможность отправить голосовое сообщения включается пользователям, которые умеют говорить.
Именинники	Добавляет подарок рядом с аватаром человека, у которого сегодня день рождения. И фильтр по именинникам в поиске. Включается практически всем, однако некоторым отключается как отвлекающий символ - если ребёнок совсем не может сфокусироваться ни на одном элементе интерфейса и задача сократить их количество до минимума.
Смена цветных и черно-белых пиктограмм	В ЭМО пиктограммы имеют цветные и черно-белые версии. Маленькие дети любят и лучше воспринимают цветные картинки, а подросткам рекомендуют использовать черно-белые - с них легче перейти на текст. Данная настройка разрешает или запрещает пользователю менять версии пиктограмм прямо из чата.
Пиктограммы по умолчанию	Настройка определяет, какие пиктограммы видит пользователь, когда заходит в чат. Если смена отключена, изменить вариант он не сможет.
Отправка предложений	Функция включает конструктор предложений - пользователь из пиктограмм, текста, голосовых сообщений или загруженных файлов собирает предложение и отправляет его одним сообщением. Обычно, отправляя одним сообщением картинку и текст, человек не выбирает, в каком порядке они идут. В ЭМО пользователь может это выбрать, может также чередовать в предложении тест и картинки.
Подсказки к ответам	Функция показывает значок вопроса рядом с некоторыми сообщениями-пиктограммами, по клику на который пользователь видит варианты ответа на сообщение. Эта функция как раз для «слабых» детей, а тем, кто и так знает, как отвечать на какие вопросы, она не нужна.
Озвучка текста	Ещё одна функция для «слабых» детей. Её очень любят не умеющие читать. А тем, кто читает, её отключают, чтобы они тренировались читать.
Подсветка «Простой язык»	Делает весь текст заглавными буквами, гласные красными. Эту функцию включают всем ребятам, начиная от самых «слабых» и отключают только тем, кто умеет читать достаточно хорошо. Не читающие благодаря такому шрифту запоминают, как выглядят некоторые слова. Тем, кто начинает учиться читать, так легче читать. А тем, кто читает хорошо, разумнее переходить на обычный шрифт.
Время отправки сообщения	Включается пользователям, которые способны ориентироваться во времени или обучаются этому.
Функционал «Мероприятия»	Функционал поддерживает инициативу ребят. Они могут, используя простой интерфейс, создать мероприятие (например, поход в музей или мастер-класс), пригласить на него людей и обсудить детали мероприятия в общем чате. Фотографии с прошедших мероприятий отправляются в раздел «Фотолайки», где их может посмотреть и поставить «лайк» любой пользователь. Это очень важный функционал для воспитанников и выпускников учреждений для сирот. В интернатах детей часто не спрашивают, хотят ли они на экскурсию, их просто везут туда. Здесь ребята учатся для начала определять, хотят ли они участвовать, обозначая свое решение с помощью кнопок «хочу» или «не хочу». Затем самим инициировать мероприятие. Но некоторым самым начинающим пользователям для начала лучше освоить базовый функционал, чтобы не путаться в страницах.

Функционал «Тренажеры»	Содержит обучающие тренажеры для различных уровней подготовки. Они могут использоваться педагогами на занятиях и создаются по их запросам. Сейчас в ЭМО 4 тренажера. Также этот раздел содержит «Рейтинг». Баллы в рейтинг даются за любую активность в ЭМО и успешное прохождение тренажера. Стимулирует ребят активнее осваивать функционал.
Функционал «Видео комнаты»	Видеоконференции на несколько человек. Это одно из последних нововведений ЭМО. Сейчас активно тестируется ребятами.
Цветовая тема	Каждый пользователь может выбрать себе два цвета, которые будут использоваться на всех страницах. По умолчанию голубой и розовый, как в первой версии. Возможна установка однотонного оформления
Инверсия цветов	Для любителей темной темы и необычных настроек. Инверсия работает умно: она не инвертирует фотографии и некоторые кнопки, для которых важен цвет (красный крестик, зеленая галочка). Текст при инверсии всегда только белый. Если включен простой язык, все буквы будут заглавные.
Настройки	Изменение личных данных и любых настроек всегда доступно администратору, этот параметр показывает, какие настройки сам пользователь может себе изменять.

Помимо настроек пользователя существуют также настройки библиотек пиктограмм. Отдельные пиктограммы и целые папки могут быть скрыты от «слабых» пользователей, оставляя им самый минимум, вплоть до одной карточки «привет».

Для голосовых сообщений и видеокончат потребовался *SSL*-сертификат, так как современные браузеры в целях безопасности не дают доступ к камере и микрофону сайтам, работающим по протоколу *http*. Был получен бесплатный сертификат от компании *Let's Encrypt*, он удовлетворил потребности. В *nginx* был настроен редирект всех поступающих запросов на 443 порт, с которого запрос уже направлялся на *localhost*.



Рис. 2. Современная страница выбора собеседника

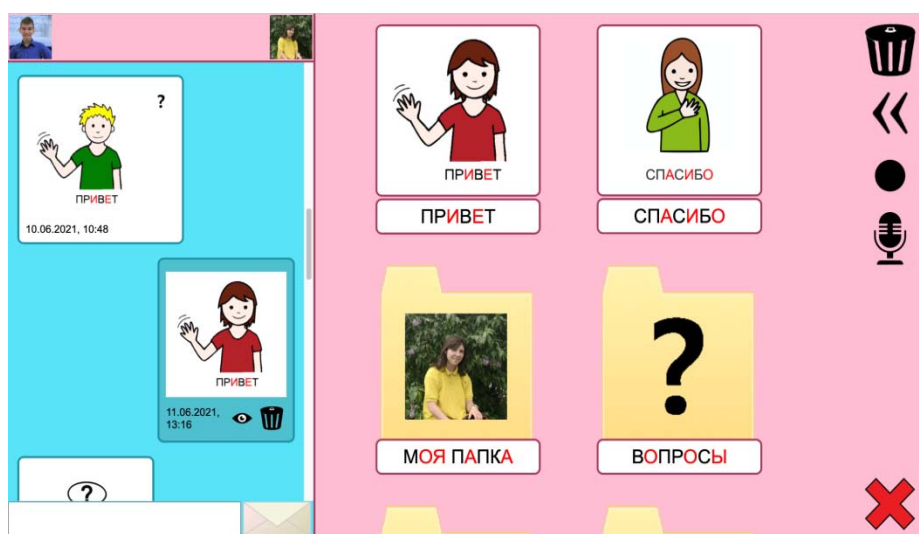


Рис. 3. Современная страница диалога

При отладке видеокончат было обнаружено, что вебсокеты и *peer-to-peer* соединения никак не взаимодействуют с базой или другими частями программы. Было принято решение вынести их в отдельный файл и запустить на отдельной от остальной части порта, чтобы они меньше влияли на остальной функционал, а остальной функционал – меньше влиял на них.

Теперь ЭМО обслуживают два сервера на *Express*.

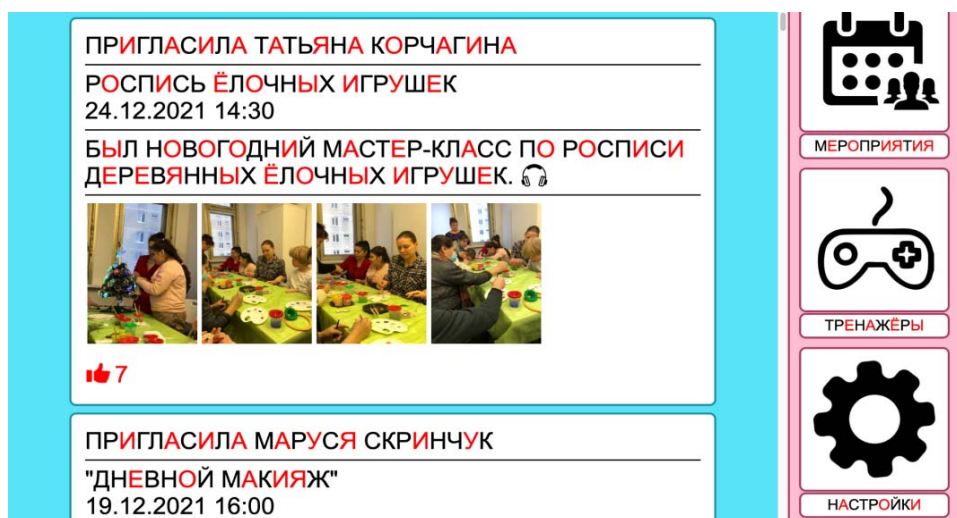


Рис. 4. Страница «Фотолайки»

Администрирование в ЭМО привязано к организациям. Есть несколько админов с полным доступом, которые добавляют в приложение организации и людей в них и видят всю статистику. В организации может быть администратор с доступом только к ней. Он создает и редактирует пользователей только в данной организации и наблюдает статистику только по ней. Также все предлагаемые мероприятия проходят модерацию [6].

Безопасность в ЭМО рассматривается с двух точек зрения: во-первых – недопустимость нежелательного контента, а во-вторых – защита пользовательских данных. Задача исключения недопустимого контента решается за счёт закрытости системы. До зимы 2022 года создать нового пользователя системы мог только администратор. Но сейчас и это стало слишком сложно, так как желающих много, и заполнять форму о каждом пользователе стало нерационально, особенно если это обычный человек (специалист, родитель или волонтер), способный сам вбить свои данные в форму. Тогда были придуманы запросы на регистрацию. Пользователь сам заполняет информацию о себе, и может войти в ЭМО и протестировать какой либо функционал, но писать он там сможет только администраторам системы, при этом сам не будет виден в поиске. Администратор, увидев запрос, проверяет пользователя и, в случае возникновения вопросов, может задать их непосредственно через ЭМО. Когда вопросы снимутся, администратор подтвердит аккаунт, и пользователь получит доступ к полному функционалу в соответствии с назначенной ему ролью в системе, либо его аккаунт будет удален. Безопасность пользовательских данных обеспечивается в ЭМО, прежде всего отсутствием сбора номеров телефонов, адресов электронных почт, и тем более данных банковских карт и другой пользовательской информации. Выбрав любую комбинацию букв в качестве логина, пользователь сразу попадает в систему, и уже через неё общается с администратором и получает подтверждение аккаунта. Для восстановления пароля он может таким же способом обратиться к администрации. Все пароли хранятся в базе системы в зашифрованном виде, прежде всего для защиты аккаунтов администраторов, а другая информация из базы, даже если, несмотря на все меры предосторожности, «утечет», будет мало полезна мошенникам или хакерам.

Мобильное приложение

Если первые 20 пользователей ЭМО не имели смартфонов, то с выходом второй версии и началом активного роста веб-приложения, ситуация в корне изменилась. В ЭМО постоянно добавлялись в качестве пользователей педагоги, семейные дети и их родители, и более «сильные» ребята, имеющие смартфоны [7]. К лету 2021 года количество запросов от пользователей на мобильное приложение приблизительно исчислялось десятками.

Мобильное приложение ЭМО, доступное в настоящее время в *App Store* и *Google Play*, было создано в кратчайшие сроки в августе 2021 года. Ресурсов было крайне мало, поэтому было принято решение использовать кроссплатформенный фреймворк как минимум для первой версии. *Flutter* может считаться очень хорошим решением, но имеет один существенный минус – необходимость знания *Dart*. Специалистов со знанием этого языка программирования, готовых заниматься проектом, не обнаружилось, что неудивительно, учитывая ограниченный бюджет. Поэтому был выбран *React Native*, использующий *JavaScript*. От фреймворков, использующих *WebView*, пришлось отказаться из-за невозможности использовать в них нативные функции операционной системы. Это необходимо для отправки голосовых сообщений, загрузки файлов или фотографий, а также доступа к камере.

Как когда-то задумывалось, мобильное приложение обращается к тому же *API*, что и веб-приложение. Оно также применяет настройки пользователя. Однако не все так хорошо, как хотелось бы: из-за того, что *API* находится в веб-

приложении, новый функционал всегда появляется первым в вебе, а мобильное приложение отстает на одну, а то и несколько версий. В данный момент из перечисленного в таблице функционала, далеко не всё реализовано в мобильном приложении. Приложение также имеет на данный момент немало «косяков» и плохо протестированных мест, так как пока существует малое количество времени. Тем не менее, уже имеется около 100 активных пользователей приложения на *Android*.

Поддержка проекта со стороны государства и общества

Проект «ЭМО – коммуникативный тренажер для особых сирот» был внесен в реестр практик с доказанной эффективностью фонда Тимченко и регулярно обновляется новыми данными. Это повышает статус проекта и дает преимущества при получении нового финансирования. ЭМО и связанные с ним проекты ежегодно получают несколько грантов от государства и коммерческих организаций, деньги уходят прежде всего на работу с самими ребятами – тьюторов, диагностики ребят, сбора данных для анализа достигнутых результатов, а также технической составляющей – разработки нового функционала, технической поддержки имеющегося, оплаты сервера и домена. Были случаи материальной поддержки со стороны крупных российских IT-компаний, которые передавали компьютеры для обучения ребят.

ЭМО был положительно оценен высокопоставленными чиновниками. АНО «Ресурсный центр «Вера. Надежда. Любовь» регулярно посещают представители департамента труда и социальной защиты населения города Москвы, а также мэр Москвы С.С. Собянин. После каждого такого визита в различных СМИ появляется от нескольких десятков до сотни публикаций, упоминающих данный проект.

Налажено тесное сотрудничество с волонтерским корпусом МГППУ. С 2020 создатели ЭМО и использующие его специалисты, несколько раз в год приглашаются на мероприятия от МГППУ для презентации ЭМО. Многие студенты проявляют заинтересованность и приходят в проект в качестве волонтеров [8].

Планы и перспективы

В данный момент в связи с политической ситуацией основные силы брошены на перенос всех данных на российские сервера. Сам перенос уже произведен, однако требуется тестирование, так как некоторые отложенные ранее функции могут теперь работать некорректно.

В настоящее время расписаны дальнейшие планы на ближайший год. Планируется реализация немалого количества нового функционала из очереди запросов, пересмотр дизайна, доработка мобильного приложения до уровня веба, продвижение и тиражирование ЭМО в другие регионы (в данный момент приложение используется только в Москве, в Московской и Челябинской областях). В списке нового функционала присутствуют профили пользователей и сообщества, что превратит ЭМО в полноценную социальную сеть, где люди с ментальными особенностями смогут взаимодействовать как между собой, так и с другими людьми.

В новых реалиях в связи с недоступностью многих зарубежных сервисов, число которых постоянно растет, ЭМО может стать альтернативой для организаций сферы социальной защиты. Когда появятся сообщества и возможность вести блог, данные организации смогут вести его в ЭМО, чтобы быть ближе к своей целевой аудитории. Для самих же русскоязычных людей с ментальными нарушениями альтернатив ЭМО не было и раньше.

Заключение

В данной статье было рассмотрено создание и развитие уникального IT-продукта в условиях ограниченных ресурсов. С самого начала существования ЭМО тестировался целевой аудиторией и за три года существования прошел путь от коммуникативного тренажера до мессенджера, а теперь движется в сторону особой социальной сети. Приложение привлекает своим необычным интерфейсом, напоминая игру. Проект регулярно получает поддержку и положительную обратную связь, как от сторонних людей, так и непосредственно от пользователей системы.

Литература

1. Бобылева И., Заводилкина О., Романова З. и Романова Е. ЭМО – коммуникативный тренажер для особых сирот. Методические рекомендации для специалистов. Москва, 2020.
2. Штягинова Е.А. Альтернативная коммуникация: методический сборник, Новосибирск: 2012. С.8.
3. Холмс С. *Стек MEAN*. Mongo, Express, Angular, Node. СПб.: Питер, 2017. 496 с. (Серия «Библиотека программиста»).
4. «ЭМО» – коммуникативный тренажер для особых сирот [Электронный ресурс]. Режим доступа: <https://deti.timchenkofoundation.org/2020/12/30/jemo-kommunikativnyj-trenazher-dlja-osobyh-sirot/>, свободный – (28.03.2022).
5. ЭМО-мессенджер, цифровые технологии помощи, видеосказки для малышей: итоги Форума социальных инноваций регионов [Электронный ресурс]. Режим доступа: <https://dszn.ru/press-center/news/7213>, свободный – (28.03.2022).

6. Журнал для специалистов сферы семейного устройства и профилактики социального сиротства [Электронный ресурс]. - Режим доступа: https://deti.timchenkofoundation.org/wp-content/uploads/2020/12/sid-mag-4-2020-29.12.pdf?Utm_source=email&utm_medium=email&utm_campaign=sidjournal&utm_content=4-12-2020&utm_term=pdf , свободный – (28.03.2022).

7. Орлов В.Г. Механизмы сетевого доступа и защиты данных в LTE / В сборнике: Технологии информационного общества. М.: Издательский дом Медиа Паблишер, 2018. С. 146-148.

8. Собянин С. С. Центр сопровождаемого проживания. Помогаем молодым людям с ментальными особенностями [Электронный ресурс]. Режим доступа: <https://www.sobyenin.ru/tsentr-soprovozhdaemogo-prozhivaniya>, свободный – (28.03.2022).

NEW SOCIAL NETWORK FOR PEOPLE WITH MENTAL DISABILITIES

Zoya I. Romanova,
Student of MTUCI, Moscow, Russia
romanova.zoya.2002@mail.ru

Anastasiya Yu. Kudryashova,
Senior Lecturer of the Department of BI, MTUCI, Moscow, Russia,
a.i.kudriashova@mtuci.ru

Abstract

The article discusses the history of the creation of a communicator for children and young people with mental disabilities, which later grew into a special social network. The prerequisites and reasons for the emergence of the program are analyzed. The main stages of development, testing, promotion and use of the communicator on the way from the idea to the present day are described. The development of a communicator is considered at the intersection of two main areas of work: programming and correctional pedagogy, which is of particular interest.

Keywords: *social network, web application, creation of a web application, people with disabilities, social project, creation of a mobile application.*

МЕТОДИКА ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ НАБОРА ДАННЫХ ДЛЯ БИНАРНОЙ И МНОГОКЛАССОВОЙ КЛАССИФИКАЦИИ АТАК

*Ландызин Андрей Николаевич,
магистрант МГУСИ, Москва, Россия,
andrej-landyzin@yandex.ru*

*Шелухин Олег Иванович
заведующий кафедрой ИБ МГУСИ, д.т.н., профессор, МГУСИ, Москва, Россия,
sheluhin@mail.ru*

Аннотация

Предложена методика предварительной обработки набора данных о сетевом трафике в формате pcap для обучения и тренировки нейросетевого алгоритма в задачах бинарной и многоклассовой классификации. Показана целесообразность сэмплирования, балансировки и выявления весомых признаков обрабатываемого набора данных. Представлены результаты обучения алгоритма бинарной классификации сетевого трафика на наличие атак на рабочем наборе данных. Показано, что проведенная предварительная обработка набора данных соответствует заявленным целям, а полученные датасеты подготовлены для проведения обучения нейросетевых алгоритмов бинарной и многоклассовой классификации.

***Ключевые слова:** нейронные сети, обработка данных, информационная безопасность, машинное обучение, системы обнаружения вторжений.*

Введение

Для повышения безопасности периметров информационных систем, системы обнаружения и предотвращения вторжений должны постоянно совершенствоваться путем внедрения методов машинного обучения [1, 9-22]. В случае систем обнаружения вторжений, алгоритмы машинного обучения базируются на анализе больших наборов данных для выявления полезной информации с целью обнаружения аномалий сетевого трафика.

Исходные данные, как правило, представляют собой пары «входящие данные – аннотация», собранные из различных источников и используются для обучения модели выполнению узкоспециализированной задачи с высоким уровнем точности. Они могут состоять из «сырых данных» в виде изображений, текстов, содержать аннотации, метки или связи и т. д. Модели машинного обучения анализируют аннотации данных обучения с тем, чтобы затем обрабатывать и размечать их. О наборе данных обычно можно судить по 5 признакам [2]:

1. Валидность – соответствие данных описываемому ими процессу в области форматирования;
2. Точность – соответствие данных описываемому ими процессу в области отражения происходящего процесса;
3. Полнота – наиболее полно характеризует соответствие данных описываемому процессу;
4. Плотность – отсутствие противоречивой информации;
5. Единообразие – соответствие данных различных типов единым принятым структурам.

Термин «подготовка данных» включает в себя сбор данных, последующую их очистку от зашумленных значений, где это возможно технически, а также преобразование категориальных переменных в числовые значения, нормализация переменных и заполнение пропущенных значений [3].

В задаче обучения нейросетевых алгоритмов качество моделей наиболее сильно зависит от качества исходных данных, поставляемых для обучения. В узкоспециализированных задачах, к которым можно отнести безопасность сетевого трафика, сами данные зачастую не являются идеальными. Как правило, в датасетах присутствуют пропуски, допускаются шумы и избыточность признаков [4]. В связи с этим используются различные методы предварительной обработки данных. Например, методы сэмплирования, выявления значимых признаков, генерации искусственных недостающих данных, благодаря которым можно добиться значимого повышения качества классификации.

Целью данной работы является разработка методики предварительной обработки на примере набора данных *CI-CIDS2017* [5], результатом которой является подготовленный датасет, используемый в качестве обучающей выборки при обучении нейросетевых алгоритмов бинарной и многоклассовой классификации.

Описание набора данных

Набор данных *CICIDS2017* является одним из наиболее популярных открытых наборов данных, используемых в качестве обучающей выборки для разработки алгоритмов машинного обучения в системах обнаружения и предотвращения вторжений. Он содержит как доброкачественный трафик, так и записи о распространенных атаках на информационные системы. Набор данных содержит помеченные потоки, которые основаны на метке времени, IP-адресе источника, IP-адресе назначения, порта пункта назначения, используемых протоколов и типов атак. Набор реализован так называемой системой В-профиля, что означает, что этот набор данных был создан на основе действий человека в сети.

Набор данных включает в себя следующие типы атак:

- веб-атаки;
- *brute-force* атаки;
- *DoS/DDoS* атаки;
- *infiltration* атаки;
- *heartbleed* атаки;
- нерегламентированное сканирование;
- атаки с использованием botnet.

По сравнению с другими открытыми наборами данных *CICIDS2017* обладает следующими преимуществами, позволяющими использовать его для обучения нейросетевых алгоритмов:

- Атакуемая сеть представляет собой конфигурацию сети, наиболее приближенную к реальной, используемой в корпоративных компьютерных сетях, что позволяет достаточно полно описывать исходные необработанные данные, на которых будет обучен нейросетевой алгоритм.
- Атаки произведены с разного типа устройств, с использованием разного инструментария для ручного проникновения и проведения атак типа «отказ в обслуживании».
- Набор данных промаркирован метками о типе атак, что позволяет реализовать алгоритмы обучения с учителем.

Вместе с тем, набор данных, собранный в условиях реального функционирования сетевой

инфраструктуры и не подвергнутый предварительной обработке обладает свойствами, способными понизить качество и скорость обучения алгоритма.

Анализ набора данных

Рассматривая набор данных *CICIDS2017* с точки зрения численных характеристик, отметим, что количество уникальных записей составляет 2830743 штуки, каждая из которых определена набором из 85 признаков. Распределение записей по классам атак представлено в таблице 1.

Таблица 1

Распределение количества записей по классам атак

Класс атаки	Кол-во записей
<i>BENIGN</i>	2273097
<i>DoS Hulk</i>	231073
<i>PortScan</i>	158930
<i>DDoS</i>	128027
<i>DoS GoldenEye</i>	10293
<i>FTP-Patator</i>	7938
<i>SSH-Patator</i>	5897
<i>DoS slowloris</i>	5796
<i>DoS Slowhttptest</i>	5499
<i>Bot</i>	1966
<i>Web Attack - Brute Force</i>	1507
<i>Web Attack - XSS</i>	652
<i>Infiltration</i>	36
<i>Web Attack - Sql Injection</i>	21
<i>Heartbleed</i>	11

Исходя из представленных в таблице 1 данных, можно отметить достаточно малое количество записей, относящихся к классам «*Heartbleed*», «*Web Attack – Sql Injection*» и подобные им. Основную долю в 80,3% записей представляют записи о трафике типа *BENIGN*, являющемся нормальным сетевым трафиком.

Кроме того, важное значение имеет большое количество признаков, присущих каждой записи. Известно, что для уменьшения вычислительных затрат, а также для улучшения производительности модели предпочтительно сокращение входных переменных до пограничного количества, при котором не снижается качество модели [6].

Произведем оценку признаков набора данных с использованием *SelectKBest* [7]. Метод *SelectKBest* оценивает признаки с помощью функции (в данном случае $f_classif$ для классификации), а затем «удаляет все, кроме k функций с наивысшей оценкой». Веса, полученные после процедуры оценки признаков, представлены в таблице 2.

Используя полученные данные построим тепловую карту датасета, позволяющую наглядно оценить взаимную корреляцию признаков между собой. Для большей наглядности тепловая карта была разбита на 2 сектора, представленные на рисунках 1 и 2. Выявленные особенности необработанного набора данных отражают необходимость проведения дополнительных манипуляций с датасетом для обеспечения улучшения характеристик входных выборок для алгоритмов бинарной и многоклассовой классификации

Таблица 2

Выявленные веса признаков

Номер метрики	Название метрики	Вес метрики
5	<i>Total Length of Bwd Packets</i>	56657.901620
64	<i>Subflow Bwd Bytes</i>	56657.901620
33	<i>Fwd Header Length</i>	11418.511569
54	<i>Fwd Header Length.1</i>	11418.511569
34	<i>Bwd Header Length</i>	6998.634218
...
19	<i>Fwd IAT Total</i>	3121.005137
75	<i>Idle Max</i>	3096.703192
1	<i>Flow Duration</i>	3085.364157
25	<i>Bwd IAT Mean</i>	2962.936697

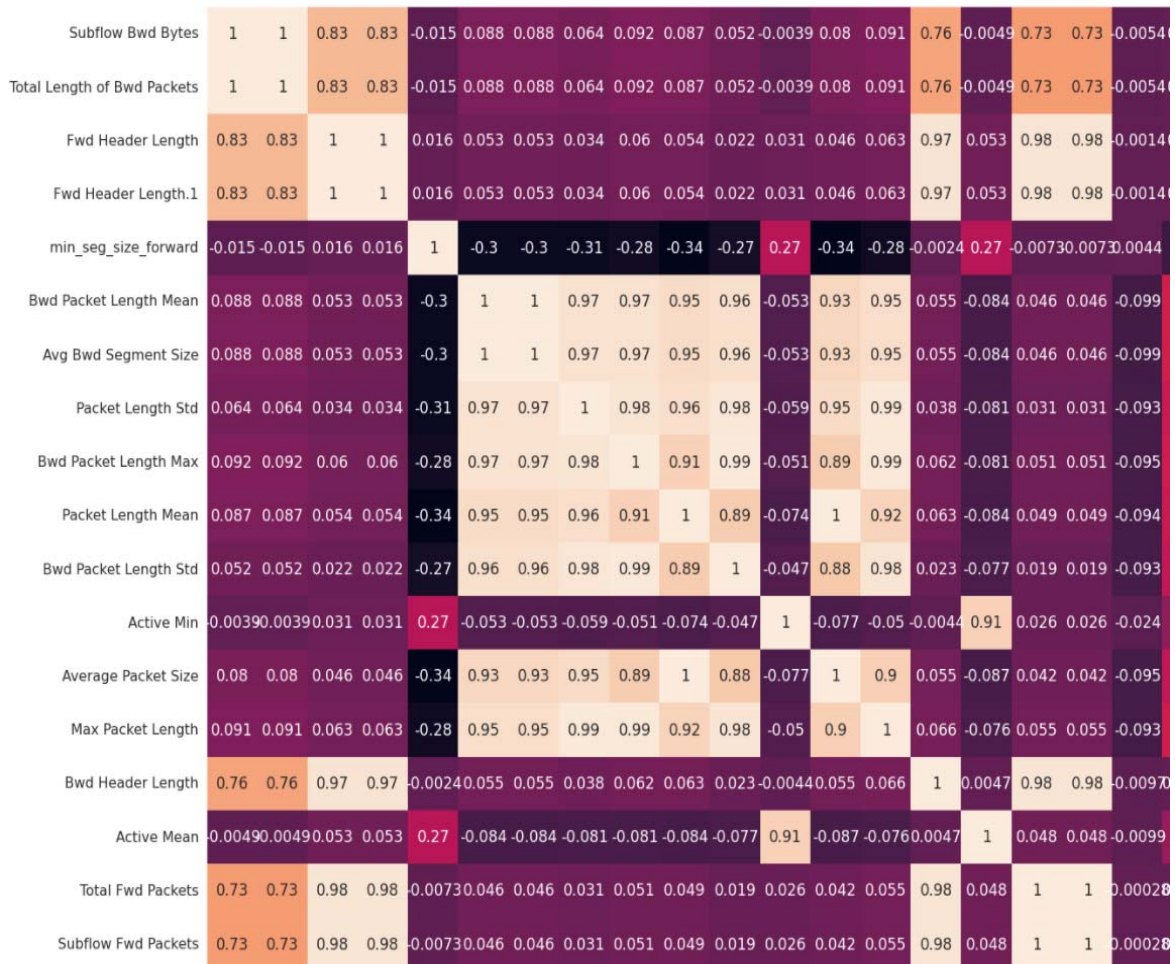


Рис 1. Тепловая карта признаков необработанного датасета *CICIDS2017*. Сектор 1

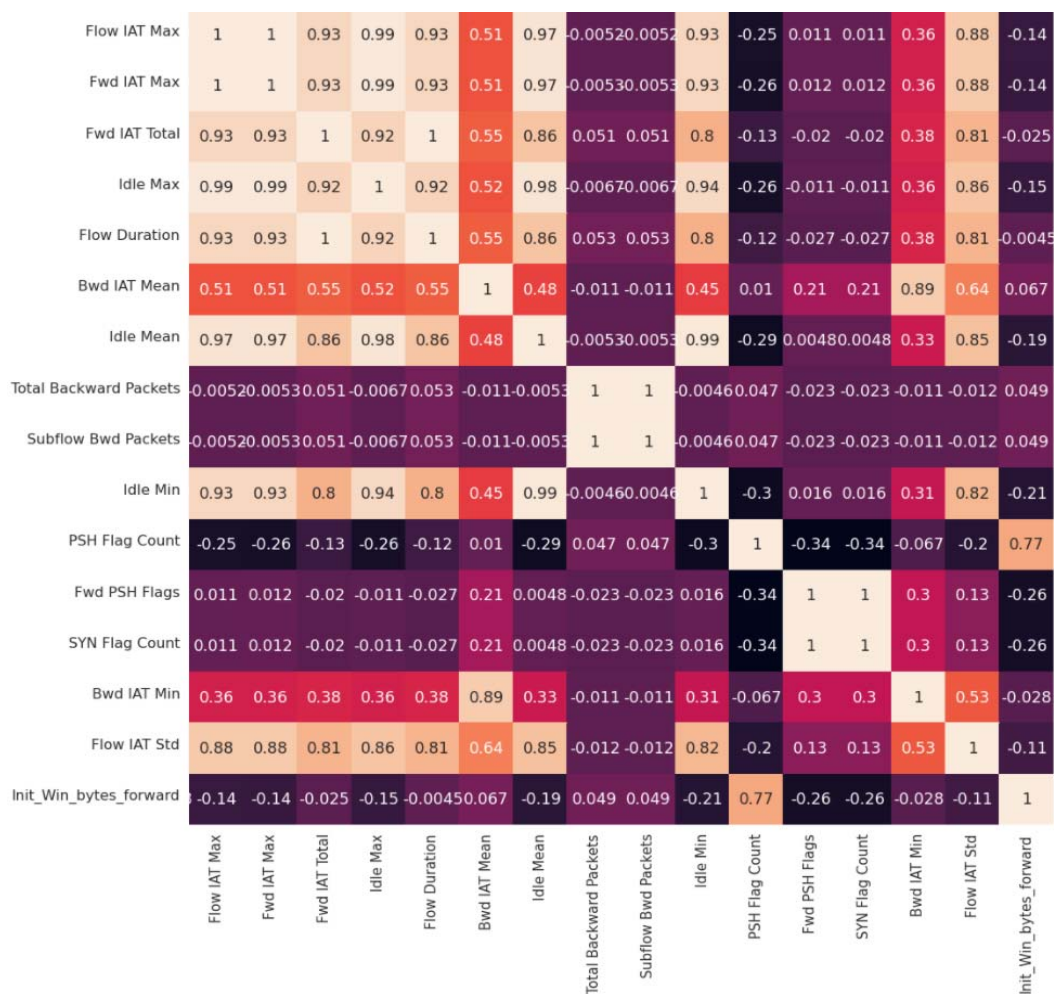


Рис. 2. Тепловая карта признаков необработанного датасета CICIDS2017. Сектор 2

Корреляция наиболее выражена для следующих признаков:

- Total Length of Bwd Packets
- Subflow Bwd Bytes
- Fwd Header Length
- Fwd Header Length.l
- Bwd Header Length
- Total Fwd Packets
- Subflow Fwd Packets
- Bwd Packet Length Mean
- Avg Bwd Segment Size
- Packet Length Std
- Fwd IAT Total
- Idle Max
- Max Flow Duration
- Flow IAT Max
- Idle Mean
- Packet Length Mean
- Idle Min
- Bwd Packet Length Std
- Bwd Packet Length

Обработка набора данных

Изменение количества записей в той или иной категории производится с использованием метода сэмплинга. Для создания баланса количества данных каждого типа определены следующие параметры, представленные в таблицу 3.

Коэффициенты сэмпирования при обработке датасета для бинарного классификатора

Признак	Коэффициент сэмпирования
<i>BENIGN</i>	0.0625
<i>DDoS</i>	0.08
<i>PortScan</i>	0.06
<i>DoS Hulk</i>	0.04

Коэффициент сэмпирования для пакетов, не помеченных как атака, выведен следующим образом:

1. Ожидаемое количество записей для категории *DoS GoldenEye*, представленной в исходном датасете: 10293 шт. Данная категория была выбрана по принципу среднего значения между максимально возможным (*BENIGN*: 2273097 шт.) и минимально возможным (*Heartbleed*: 11 шт.) количеством пакетов. Данный способ позволяет уменьшить количество записей, которые будут отброшены в процессе сэмпирования, что положительно влияет на качество входных данных для нейросетевого алгоритма.

2. Ожидаемое количество записей для каждой из 14 категорий, при которых наименьшим образом подвергнутся удалению существующие или добавлению искусственные данные, составляет: 141139 шт. Исходя из этого, требуемое значение коэффициента сэмпирования оценивается формулой:

$$K = \frac{1}{BENIGN / \sum categories} = \frac{1}{2273097 / 141139} \approx 0.0625 \tag{1}$$

Отбросив данные аналогичным способом из всего датасета и построив для полученного набора данных тепловую карту, представленную на рисунке 2, убедимся, что проводимые манипуляции дали положительный результат. В итоге планируемая к созданию нейросетевая модель бинарной классификации будет в меньшей степени подвержена переобучению, а скорость обучения сократится.

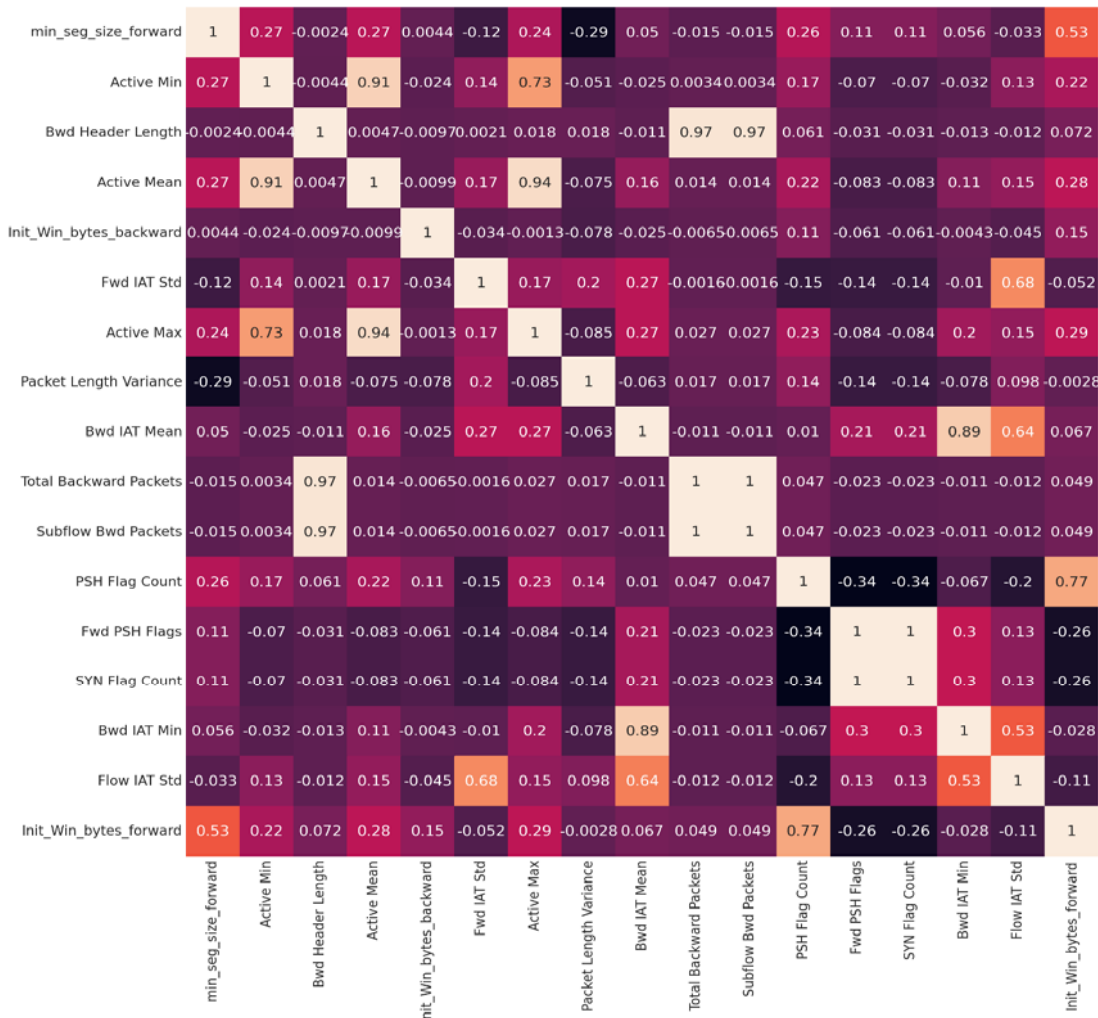


Рис. 2. Тепловая карта признаков предварительно обработанного датасета

Рассмотрим бинарную классификацию как приоритетную, в условиях построения нейросетевого алгоритма, находящегося в составе систем обнаружения и предотвращения вторжений.

Выделим записи типа *BENIGN*, количество которых после семплирования составило 142069 шт., в отдельный датасет, который не будет подвержен искусственной генерации записей для балансировки. Для датасета очищенного от записей об безопасном трафике проведем процедуру генерации методом *SMOTEtomek* [8].

Метод *SMOTEtomek* объединяет два алгоритма – *SMOTE* и *Tomek*. В отличие от случайной передискретизации, которая дублирует только некоторые случайные примеры из класса меньшинства, использование метода *SMOTE* позволяет генерировать реализации на основе расстояния между данными (обычно с использованием евклидова расстояния) и метода ближайших соседей. В результате генерируемые синтетические данные оказываются относительно близкими к пространству признаков в классе меньшинства, и добавляют «новую информацию» к данным, в отличие от случайной передискретизации. В результате метод *SMOTE* позволяет создавать элементы в непосредственной близости от существующих, как это схематически показано на рисунке 3.

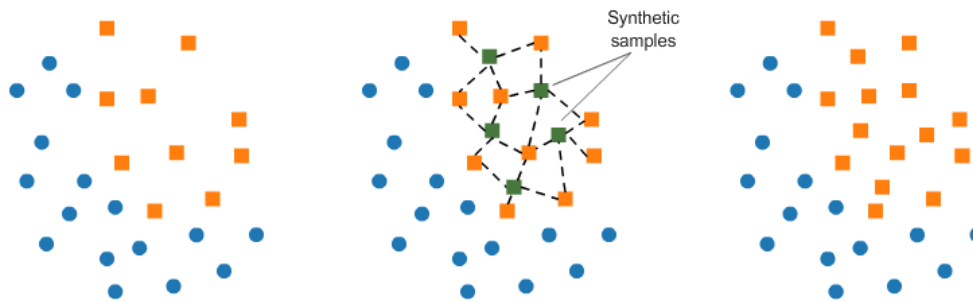


Рис. 3. Графическая интерпретация метода *SMOTE*

Метод *Tomek Links* использует правило для выбора пары наблюдений (допустим, **a** и **b**) вида:

- Ближайший сосед элемента **a** – элемент **b**;
- Ближайший сосед элемента **b** – элемент **a**;
- **a** и **b** относятся к разным классам, то есть, к классу меньшинства и большинства и наоборот соответственно;
- Организация ссылки *Tomek*

Затем, элементы, относящиеся к большинству, удаляются для нормализации датасета. Схематично это представлено на рисунке 4.

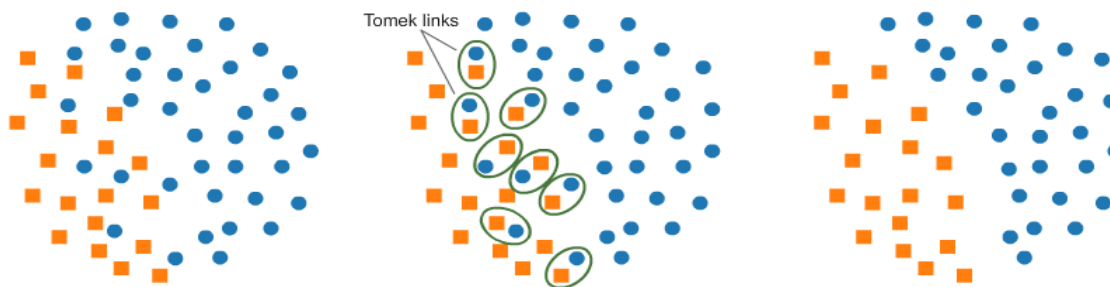


Рис. 4. Графическая интерпретация работы метода *Tomek*

После проведения операции передискретизации и объединения наборов данных получим датасет с количественными характеристиками, сведенными в таблицу 4 и представленными в виде диаграммы на рисунке 5.

Таблица 4

Количественные характеристики полученного набора данных

Тип атаки	Количество записей
Heartbleed	10242
DoS GoldenEye	10241
PortScan	10241
Bot	10240
FTP-Patator	10240
Infiltration	10240
Web Attack SQL Injection	10238
SSH-Patator	10237
DDoS	10224
DoS Hulk	10220
DoS Slowhttptest	10214
DoS Slowloris	10210
Web Attack XSS	9412
Web Attack Brute Force	9409
BENIGN	142069

Распределение количества записей по классам

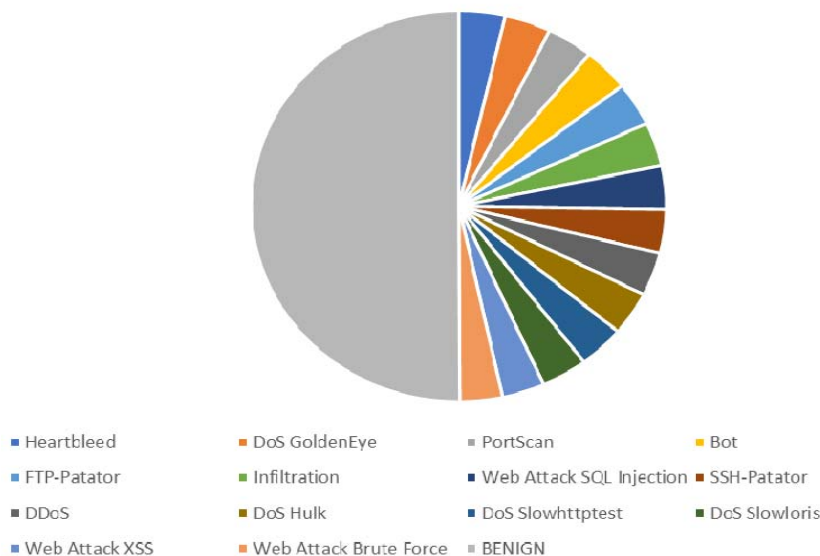


Рис. 5. Диаграмма количественных характеристик полученного набора данных

Рассмотрим обработку набора данных для многоклассового классификатора, где соблюдение баланса между классами записей является приоритетной целью.

На этапе сэмплирования данных, для балансировки количества записей, выделим из необработанного датасета медианный класс, не являющийся классом *BENIGN*. В наборе данных *CICIDS2017* данным классом является *DoS GoldenEye* с количеством записей равным 10293 шт. Проведем процедуру сэмплирования с коэффициентами, представленными в таблице 5. Коэффициент сэмплирования получен с использованием формулы (1).

Таблица 5

Коэффициенты сэмплирования при обработке датасета для многоклассового классификатора

Признак	Коэф. сэмплирования
<i>BENIGN</i>	0.004
<i>DDoS</i>	0.08
<i>PortScan</i>	0.06
<i>DoS Hulk</i>	0.04

Для полученного после передискретизации набора для генерации искусственных записей для остальных классов также применим метод *SMOTEotomek*.

Выходные данные представлены в таблице 6 и на диаграмме (рисунок 6).

Таблица 6

Количественные характеристики полученного набора данных

Тип атаки	Количество записей
<i>Heartbleed</i>	10293
<i>DoS GoldenEye</i>	10286
<i>PortScan</i>	10289
<i>Bot</i>	10284
<i>FTP-Patator</i>	10293
<i>Infiltration</i>	10292
<i>Web Attack SQL Injection</i>	10288
<i>SSH-Patator</i>	10292
<i>DDoS</i>	10242
<i>DoS Hulk</i>	10250
<i>DoS Slowhttptest</i>	10268
<i>DoS Slowloris</i>	10275
<i>Web Attack XSS</i>	8993
<i>Web Attack Brute Force</i>	8999
<i>BENIGN</i>	10293

Распределение количества записей по классам

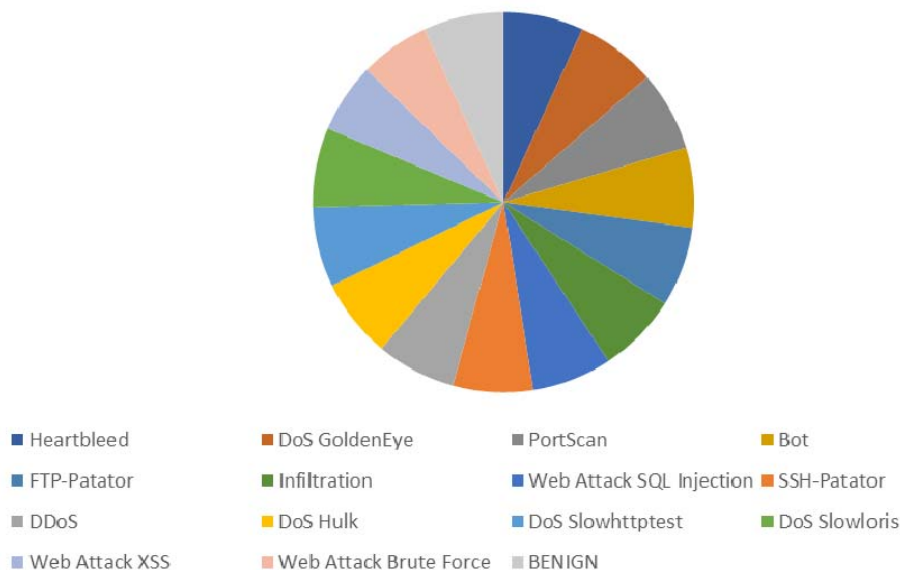


Рис. 6. Диаграмма количественных характеристик полученного набора данных

Анализ результатов и их применения в нейросетевых алгоритмах

Анализируя полученные данные, можно сделать вывод о том, что полученный набор данных для бинарной классификации разделен в соотношении 50/50, а объема входных данных сокращен путем удаления из общего датасета полей с наибольшей корреляцией.

В случае многоклассовой классификации записи распределены в равных пропорциях с сохранением данных из необработанного датасета с целью обеспечения наибольшего соответствия исходным данным. Излишние, наиболее скоррелированные признаки также, удалены.

Применение полученных наборов данных, на тестовых конфигурациях реализованных нейросетевых алгоритмов классификаций подтверждают повышение качества классификации и снижении времени на обучение нейросети.

В ходе исследований реализован алгоритм бинарной классификации, представляющий собой рекуррентную нейронную сеть с 3 скрытыми слоями LSTM, и анализировались метрики:

- *accuracy* – доля правильных ответов алгоритма;
- *precision* – доля истинно положительных классификаций;
- *recall* – доля положительно классифицированных объектов из всех положительно классифицированных объектов;
- *f1* – средне-гармоническое отношение метрик *precision* и *recall*.

Результаты экспериментов сведены в таблицы 7-10, а также графически интерпретированы в диаграммах, представленных на рисунках 7-10.

Таблица 7

Результаты тестирования необработанного датасета для бинарной классификации

	25 эпох	26 эпох	27 эпох	28 эпох	29 эпох	30 эпох	31 эпоха	32 эпохи	33 эпохи
A precision	0,92	0,92	0,91	0,92	0,92	0,91	0,96	0,97	0,87
A recall	0,98	0,96	0,98	0,98	0,96	0,98	0,99	0,99	1,00
A F1	0,95	0,94	0,95	0,95	0,94	0,95	0,98	0,99	0,93
B precision	0,83	0,73	0,86	0,83	0,73	0,86	0,94	0,92	0,96
B recall	0,57	0,61	0,53	0,57	0,61	0,53	0,81	0,84	0,25
B F1	0,67	0,66	0,65	0,67	0,66	0,65	0,87	0,88	0,40

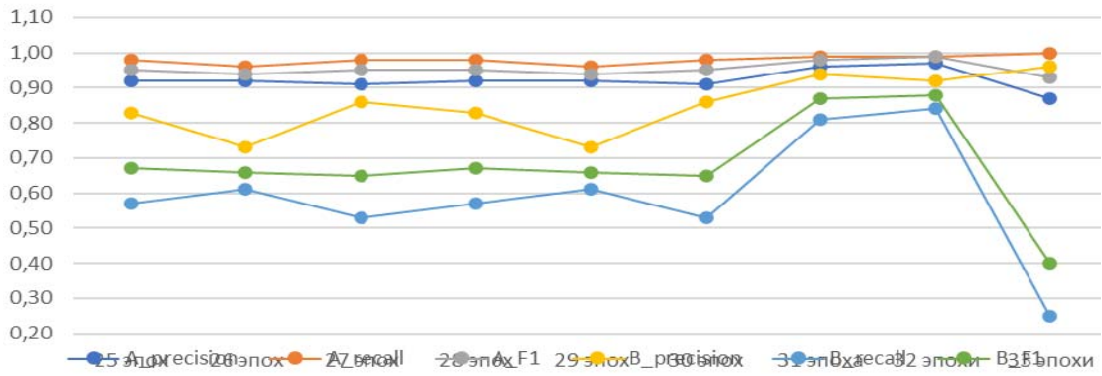


Рис. 7. Результаты тестирования необработанного датасета для бинарной классификации

Таблица 8

Результаты тестирования обработанного датасета для бинарной классификации

	25 эпох	26 эпох	27 эпох	28 эпох	29 эпох	30 эпох	31 эпоха	32 эпохи	33 эпохи
A_precision	0,93	0,94	0,95	0,95	0,97	0,96	0,96	0,97	0,98
A_recall	0,98	0,99	0,98	0,99	0,98	0,99	0,96	0,97	0,97
A_F1	0,95	0,96	0,97	0,97	0,97	0,98	0,96	0,98	0,98
B_precision	0,98	0,99	0,98	0,99	0,98	0,99	0,96	0,97	0,97
B_recall	0,93	0,94	0,96	0,95	0,97	0,97	0,97	0,98	0,99
B_F1	0,95	0,96	0,97	0,97	0,98	0,98	0,97	0,98	0,99

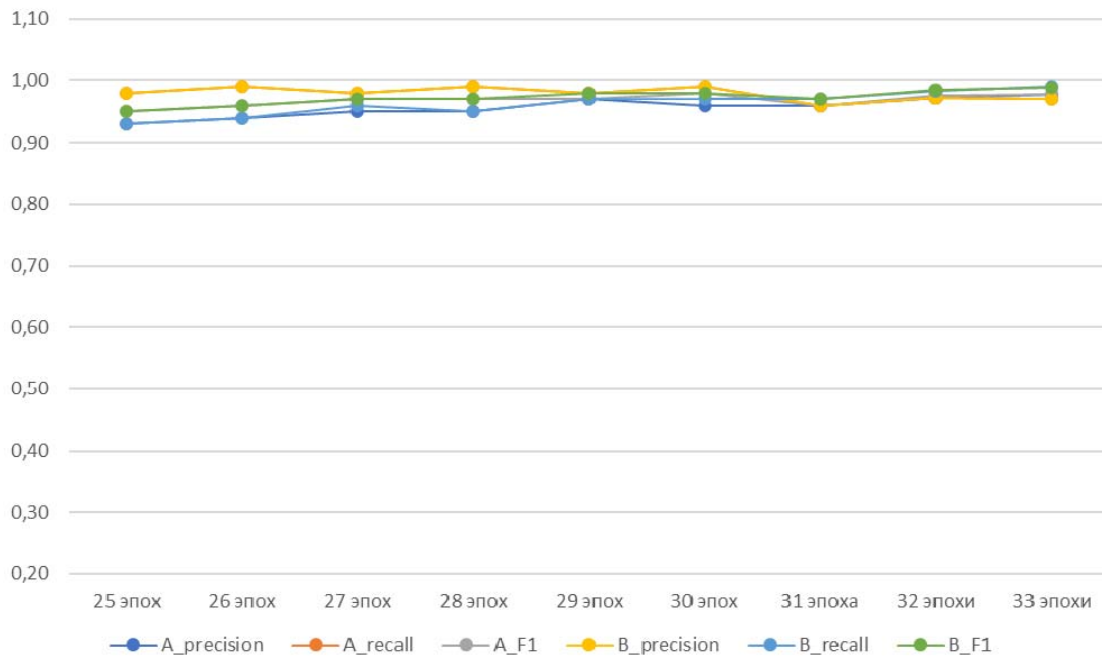


Рис. 8. Результаты тестирования обработанного датасета для бинарной классификации

Таблица 9

Результаты тестирования необработанного датасета для многоклассовой классификации

	25 эпох	26 эпох	27 эпох	28 эпох	29 эпох	30 эпох	31 эпоха	32 эпохи	33 эпохи
Recall	0,976	0,955	0,982	0,976	0,955	0,982	0,989	0,985	0,998
Precision	0,918	0,924	0,912	0,918	0,924	0,912	0,963	0,969	0,870
Accuracy	0,908	0,898	0,907	0,908	0,898	0,907	0,960	0,962	0,874
F1	0,946	0,939	0,946	0,946	0,939	0,946	0,976	0,977	0,929
ROC	0,771	0,781	0,755	0,771	0,781	0,755	0,900	0,914	0,625

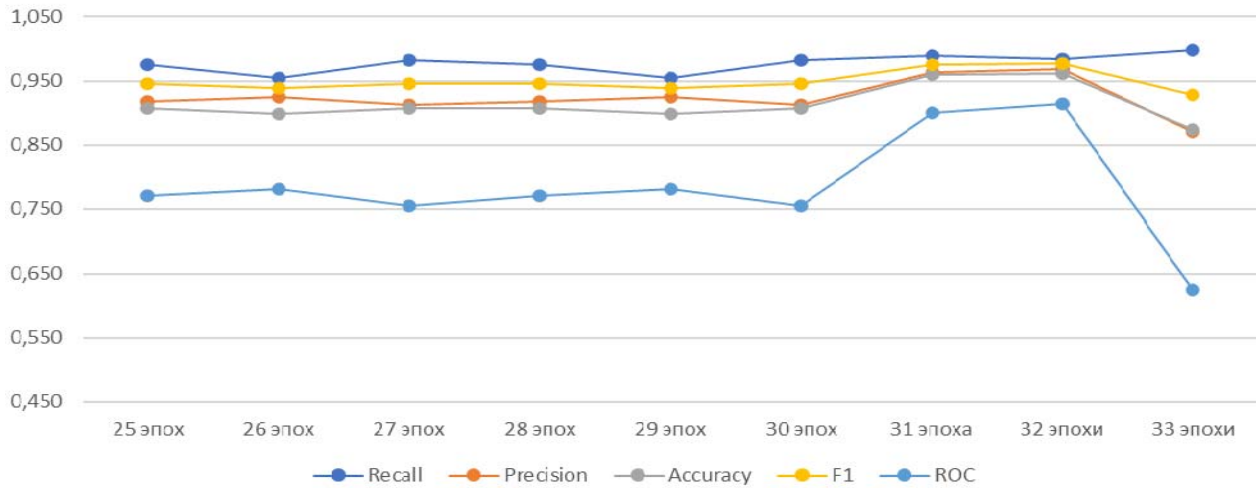


Рис. 9. Результаты тестирования необработанного датасета для многоклассовой классификации

Таблица 10

Результаты тестирования обработанного датасета для многоклассовой классификации

	25 эпох	26 эпох	27 эпох	28 эпох	29 эпох	30 эпох	31 эпоха	32 эпохи	33 эпохи
Recall	0,983	0,985	0,982	0,987	0,983	0,987	0,961	0,972	0,970
Precision	0,926	0,939	0,953	0,952	0,965	0,963	0,964	0,977	0,983
Accuracy	0,954	0,962	0,969	0,970	0,975	0,976	0,963	0,976	0,978
F1	0,954	0,961	0,968	0,969	0,974	0,975	0,962	0,974	0,977
ROC	0,955	0,963	0,969	0,970	0,975	0,976	0,963	0,975	0,977

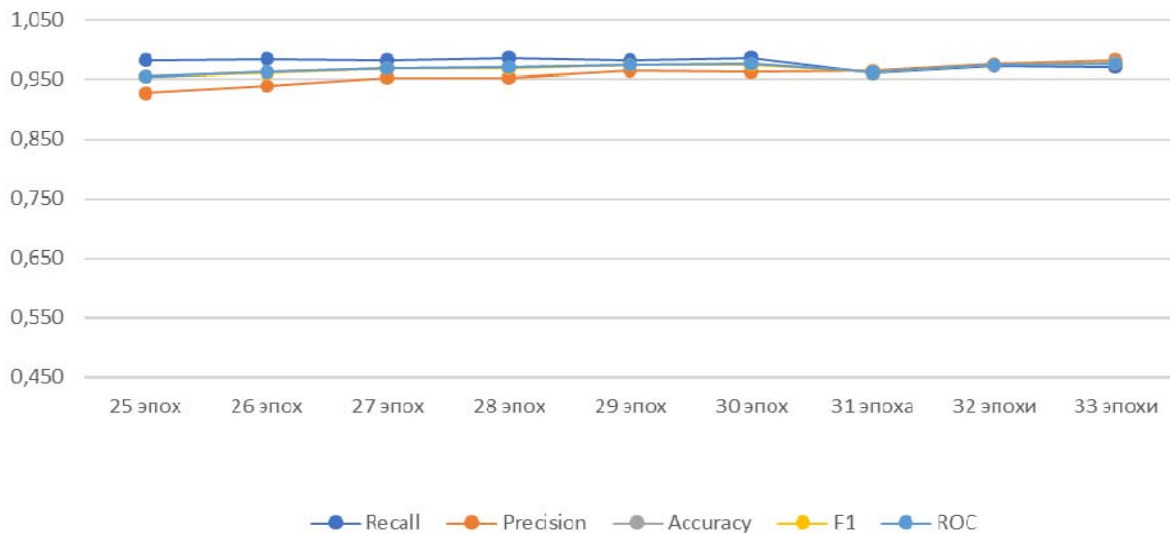


Рис. 10. Результаты тестирования обработанного датасета для многоклассовой классификации

Заключение

Исходя из результатов, представленных в таблицах 7-10 и на диаграммах 7-10, можно сделать вывод, что проведенная с помощью предложенной методики предварительная обработка набора данных обеспечивает достижение заявленных целей исследования. Полученные датасеты подготовлены для проведения обучения нейросетевых алгоритмов бинарной и многоклассовой классификации.

Литература

1. H. Lui and B. Lang. "Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey," [Электронный ресурс]. Режим доступа: <https://www.mdpi.com/2076-3417/9/20/4396>. (Дата обращения 29.03.2022).

2. Data Preparation for Machine Learning [Электронный ресурс]. Режим доступа: <https://www.akkio.com/data-preparation-for-machine-learning> (Дата обращения 29.03.2022).
3. Способы обеспечения качества данных для машинного обучения [Электронный ресурс]. Режим доступа: <https://vc.ru/ml/353279-sposoby-obespecheniya-kachestva-dannyh-dlya-mashinnogo-obucheniya> (Дата обращения 29.03.2022).
4. Data Preparation and Feature Engineering for Machine Learning [Электронный ресурс]. Режим доступа: <https://developers.google.com/machine-learning/data-prep> (Дата обращения 29.03.2022).
5. Intrusion Detection Evaluation Dataset (CIC-IDS2017) [Электронный ресурс]. Режим доступа: <https://www.unb.ca/cic/datasets/ids-2017.html>
6. How to Choose a Feature Selection Method for Machine Learning [Электронный ресурс]. Режим доступа: <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-daa/#:~:text=Feature%20selection%20is%20the%20process,the%20performance%20of%20the%20model>. (Дата обращения 29.03.2022).
7. sklearn.feature_selection.SelectKBest [Электронный ресурс]. Режим доступа: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html (Дата обращения 29.03.2022).
8. SMOTETomek [Электронный ресурс]. Режим доступа: <https://imbalanced-learn.org/stable/references/generated/imblearn.combine.SMOTETomek.html> (Дата обращения 29.03.2022).
9. Шелухин О.И., Симонян А.Г., Ванюшина А.В. Влияние структуры обучающей выборки на эффективность классификации приложений трафика методами машинного обучения // Т-Comm: Телекоммуникации и транспорт. 2017. Т. 11. № 2. С. 25-31.
10. Sheluhin O.I., Ivannikova V.P. Comparative analysis of informative features quantity and composition selection methods for the computer attacks classification using the UNSW-NB15 dataset // Т-Comm: Телекоммуникации и транспорт. 2020. Т. 14, № 10. С. 53-60.
11. Шелухин О.И., Шариков А.Ю. Имитация поведения компьютерной системы с помощью искусственных нейронных сетей // Т-Comm: Телекоммуникации и транспорт. 2021. Т. 15. № 5. С. 29-37.
12. Шелухин О.И., Раковский Д.И. Выбор метрических атрибутов редких аномальных событий компьютерной системы методами интеллектуального анализа данных // Т-Comm: Телекоммуникации и транспорт. 2021. Т. 15. № 6. С. 40-47.
13. Шелухин О.И., Раковский Д.И. Бинарная классификация многоатрибутных размеченных аномальных событий компьютерных систем с помощью алгоритма SVDD // Научные технологии в космических исследованиях Земли. 2021. Т. 13. № 2. С. 74-84.
14. Шелухин О.И., Шариков А.Ю. Имитационное моделирование поведения компьютерной системы с помощью искусственных нейронных сетей // REDS: Телекоммуникационные устройства и системы. 2021. Т. 11. № 1. С. 54-60.
15. Шелухин О.И., Осин А.В., Костин Д.В. Диагностика "здоровья" компьютерной сети на основе секвенциального анализа последовательностных паттернов // Т-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 2. С. 9-16.
16. Шелухин О.И., Осин А.В., Костин Д.В. Мониторинг и диагностика аномальных состояний компьютерной сети на основе изучения "исторических данных" // Т-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 4. С. 23-30.
17. Шелухин О.И., Пугачев Д.А. Разработка искусственной иммунной системы на основе отрицательного отбора с применением нейросетевых детекторов для обнаружения компьютерных атак // REDS: Телекоммуникационные устройства и системы. 2020. Т. 10. № 1. С. 3-8.
18. Шелухин О.И., Костин Д.В., Резник И.Ю. Мониторинг и структура аномальных паттернов системных журналов компьютерных систем // REDS: Телекоммуникационные устройства и системы. 2020. Т. 10. № 2. С. 3-8.
19. Иванникова В.П., Шелухин О.И. Бинарная классификация компьютерных атак на примере базы данных UNSW-NB15 // Телекоммуникации и информационные технологии. 2020. Т. 7. № 1. С. 10-18.
20. Шелухин О.И., Барков В.В., Полковников М.В. Сравнительный анализ алгоритмов оценки количества и структуры атрибутов в задачах классификации мобильных приложений // Научные технологии в космических исследованиях Земли. 2019. Т. 11. № 2. С. 90-100.
21. Иевлев О.П., Шелухин О.И., Большаков А.С. Создание и техническая модернизация лабораторий кафедры "информационная безопасность" МТУСИ // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2019. Т. 8. № 1. С. 52-57.
22. Груздев С.П., Шелухин О.И. Бинарная классификация компьютерных атак на информационные ресурсы при помощи нечёткой логики // Телекоммуникации и информационные технологии. 2019. Т. 6. № 2. С. 115-122.

DATA SET PRE-PROCESSING TECHNIQUE FOR BINARY AND MULTICLASS ATTACK CLASSIFICATION

Andrey N. Landyzin,
undergraduate MTUCI, Moscow, Russia
andrej-landyzin@yandex.ru

Oleg I. Sheluhin
Head of the Department of IS of MTUCI, Doctor of Technical Sciences,
Professor, MTUCI, Moscow, Russia,
sheluhin@mail.ru

Abstract

A technique for preprocessing a set of data on network traffic in the pcap format for learning and training a neural network algorithm in binary and multiclass classification problems is proposed. The expediency of sampling, balancing and identifying significant features of the processed data set is shown. The results of training the algorithm for binary classification of network traffic for the presence of attacks on the processed data set are presented. It is shown that the pre-processing of the data set corresponds to the stated goals, and the resulting datasets are prepared for training neural network algorithms for binary and multiclass classification.

Keywords: *neural networks, data processing, information security, machine learning, intrusion detection systems.*

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ ПО РАСПОЗНАВАНИЮ ЖЕСТОВ

Субботин Антон Андреевич,
магистрант МТУСИ, Москва, Россия,
anton140299@gmail.com

Воронова Лилия Ивановна,
заведующий кафедрой ИСУиА, д.ф.-м.н., профессор, МТУСИ, Москва, Россия,
voronova.lilia@yandex.ru

Аннотация

В статье описывается применение нейронных сетей в области распознавание жестов с помощью фреймворка MediaPipe. Приведены результаты проектирования, реализации и обучения нейронной сети и обогащение базы данных с изображениями жестов с учётом соотношения к их классам. Выполнено тестирование системы распознавания жестов с помощью фреймворка MediaPipe. Результатом исследования является нейросеть, которая способна распознавать с высокой точностью определённые жесты буквенного алфавита.

Ключевые слова: жесты, машинное обучение, нейронная сеть, MediaPipe, Jupyter Notebook, сурдофон, переводчик

Введение

В связи с прогрессом в сфере информационных технологий, появляются всё более совершенные устройства и ПО, которые, в частности дополняют и улучшают жизнь, как обычных людей, так и граждан с ограниченными возможностями. Общение – это процесс обмена информацией и взаимодействия между людьми, основанный на восприятии и понимании друг друга. В России в настоящее время, по данным Минздрава РФ, насчитывается около 200 тысяч инвалидов по слуху и слабослышащих граждан. По оценке Всероссийского общества глухих заметные проблемы со слухом имеются у более чем 13 миллионов жителей России [1]. С учётом этого актуальным является использование машинного обучения в сфере распознавания жестов, для решения проблем значительного числа населения, которое испытывает трудности с восприятием звуковой информации. Помимо этого, технология распознавания жестов способствует самостоятельному развитию человека, так как может активно использоваться в обучении, нивелируя помощь сурдопереводчика [13-20].

Обзор продуктов-аналогов

Существует несколько видов технологий для распознавания жестов, как с использованием компьютерного зрения, так и без него. В состав технического оборудования для распознавания жестов на основе изображения с использованием компьютерного зрения обычно входят: камера, различные маркера, сенсор глубины и стереокамера, рисунок 1. В системах, не использующих технологию компьютерного зрения для распознавания жестов, применяют перчатки и браслеты или аналогичные носимые устройства с установленными на них сенсорами. При этом также требуются подключаемые с помощью проводного соединения акселерометры и гироскопы [3].

Первый сенсор глубины изображения, первоначально предназначенный для консоли Xbox 360 в 2010 г., и позднее для персональных компьютеров под управлением ОС Windows, выпустила компания Microsoft в 2012 г. Существуют также и альтернативные сенсоры глубины: Xtion Pro, Leap Motion, DUO 3D, Intel Perceptual Computing. Распознавание жестов руки связано с двумя проблемами: определение руки и распознавание жеста (рис. 2).

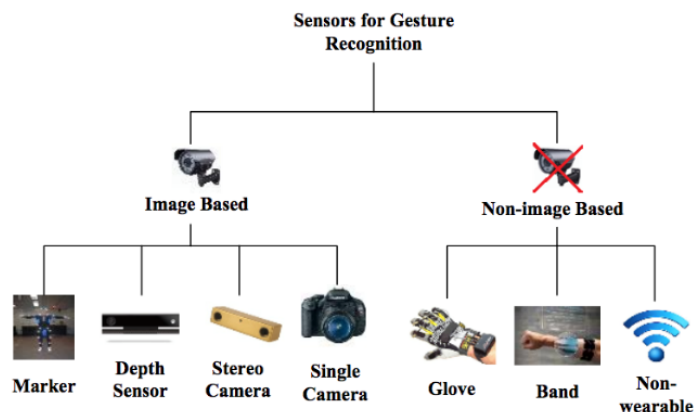


Рис. 1. Оборудование для распознавания жестов

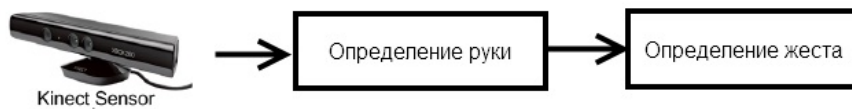


Рис. 2. Простая схема распознавания жеста с помощью *Kinect Sensor*

Методика исследования распознавания жестов в машинном обучении.

Методика, используемая в построении нейронной сети для распознавания жестов, включает в себя фреймворк, разработанный Google, *MediaPipe* [4]. На схеме, представленной на рисунке 3, изображен сценарий нейросети, основанием которого служат библиотеки для предобработки данных, а также *MediaPipe* в *Jupyter Notebook*.

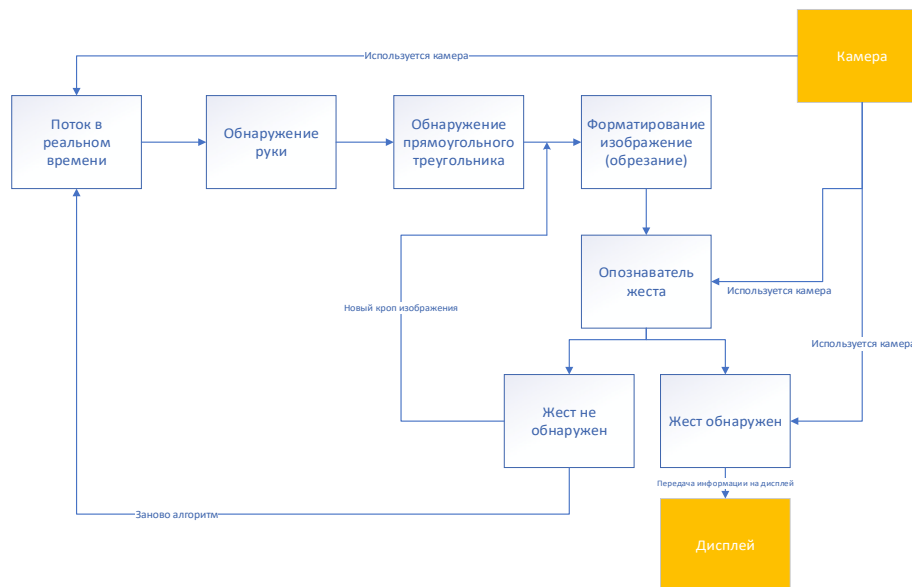


Рис. 3. Сценарий работы алгоритма нейросети на основе *MediaPipe*

Схем работы нейросети с использованием *MediaPipe* [5], включает следующие процедуры:

1. Захват камерой видеопотока в реальном времени.
2. Обнаружение камерой ладони методом построения точек и образование прямоугольного треугольника на ней (рис. 4).

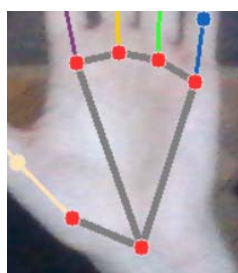


Рис. 4. Прямоугольный треугольник из точек

3. Обрезание нейросетью изображения и вырезание лишних деталей для фокусирования изображения ладони. В роли опознавателя жеста выступает скрипт, который строится исходя из расположения 21 опорной точки на каждой ладони (рис. 5).

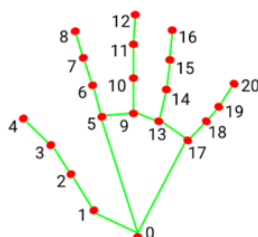


Рис. 5. Построение точек *MediaPipe*

4. В зависимости от расположения этих опорных точек, нейросеть выдаёт определённый ответ: если точки не совпадают с написанным скриптом, то весь алгоритм немедленно повторяется в реальном времени.

5. В случае если точки распознаны, и алгоритм распознавания буквы по опорным точкам определил символ, соответствующий изображению символ выводится на дисплей в виде ответа на определённый жест.

На рисунке 6 приведён пример ответа на жест с определённым расположением точек на ладони, соответствующий символу «А».

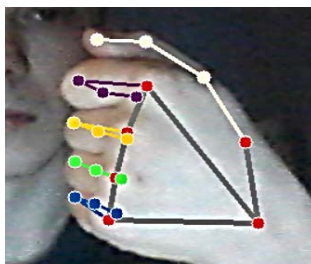


Рис. 6. Пример выполнения точек

Описание выбранной архитектуры нейронной сети

К задачам классификации изображений относятся свёрточные нейронные сети (НС). Данный вид НС принимает начальное изображение и в ответ выводит его класс – в нашем случае символ, соответствующий жесту. На рисунке 7 представлена схема свёрточной нейронной сети [6,7].

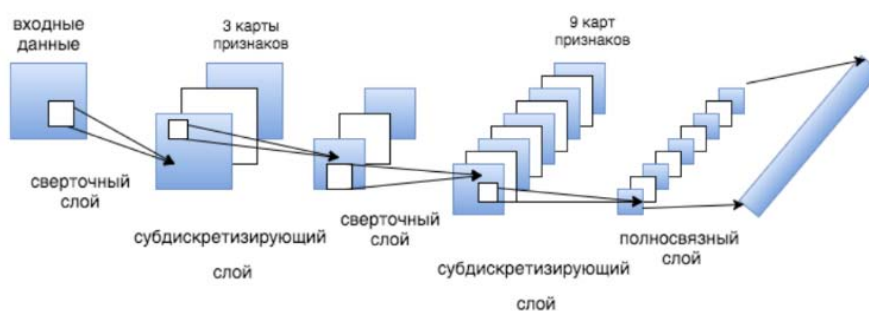


Рис. 7. Свёрточная нейронная сеть

Реализация нейронной сети для распознавания жестов

Для того, чтобы НС работала корректно необходимо загрузить в неё большой объем данных изображений жестов. В роли *dataset* будет выступать набор жестов, который был загружен с сервера «Сурдофон». Первоначально были добавлены фотографии жестов и обозначены их названия в соответствии с алфавитом (рис. 8).



Рис. 8. Формирование дата-сета в «Сурдофон» из жестов РЖЯ

На основании *dataset* была сформирована модель набора параметров *keras_model.h5*.

На рисунке 9 изображено построение точек и линий ладоней, а также их обозначения и названия.

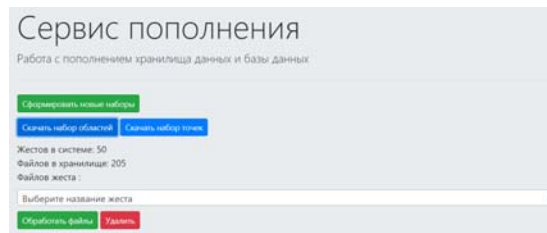


Рис. 9. Скачивание набора данных для обучения нейронной сети

Работа с фреймворком *MediaPipe* начиналась с загрузки библиотеки *MediaPipe* в программном обеспечении *Anaconda*. В следующем шаге выполнялась команда `pip install mediapipe` [8]. После выполненных действий остаётся импортировать библиотеку (рис. 10) в проект и инициализировать модель параметров, которая была сформирована ранее (*keras_model.h5*) и затем модель утилиты рисования *MediaPipe* [9]. Инициализация изображена на рисунке 8.

Утилита для рисования помогает изобразить 21 точку и линии, соединяющие эти точки на вашем изображении или кадре (рис. 11). После обнаружения ладони по всему изображению модель ориентира руки выполняет точную локализацию ключевых точек. Обнаружение происходит с помощью регрессии, то есть прямого прогнозирования координат. Модель усваивает последовательное представление внутренней позы руки и устойчива даже к частично видимым рукам.

```
Ввод [2]: import cv2
import mediapipe as mp
mp_drawing = mp.solutions.drawing_utils
mp_drawing_styles = mp.solutions.drawing_styles
mp_hands = mp.solutions.hands
```

Рис. 10. Импортирование библиотеки *MediaPipe*

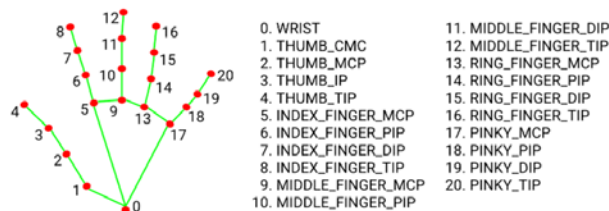


Рис. 11. Расположение точек и линий, которое формирует библиотека *MediaPipe*

Для того, чтобы обучить нейронную сеть опознавать по точкам ладони, необходимо написать скрипт. Часть скрипта уже была представлена разработчиками *MediaPipe* на сайте *mediapipe.dev*. После запуска проекта в *Jupyter Notebook* [10] мы получили результат, который изображён на рисунках 12 и 13.

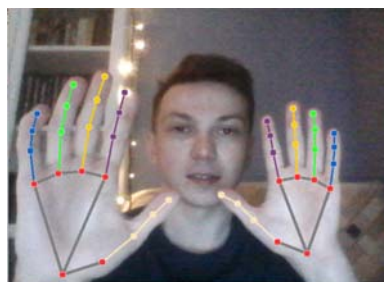


Рис. 12. Результат распознавания ладоней

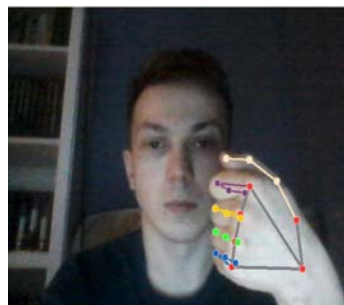


Рис. 13. Результат распознавания закрытой ладони

Из изображений на рисунках 12 и 13 можно заметить, насколько точно алгоритм фреймворка *MediaPipe* строит точки на ладони, несмотря на избыточность или недостаток света. Также стоит учесть, что даже при закрытой ладони алгоритм правильно учитывает положение опорных точек.

Существует генерация 3D модели в веб-версии *MediaPipe* (рис. 14), где показывается 3D схема руки. [11,12].

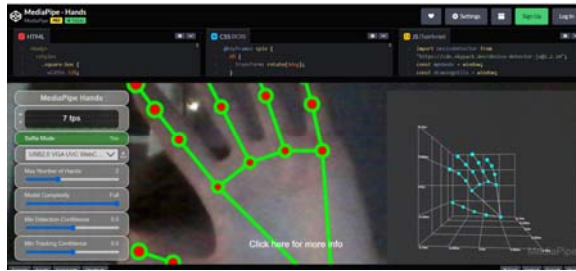


Рис. 14. *MediaPipe Online* с построением 3D модели ладони

Следующим шагом является создать алгоритма, для точного вывода буквы по определённому изображению. Для примера, возьмём такие буквы как А, Л, М, Ж, и проверим, насколько точно, написанная нами НС распознает жесты [12]. Скрипт представлен на рисунок 15. Укажем путь на изображении с жестом, служащим обозначением буквы «Л» (рис. 16).

```

In [1]: from keras.models import load_model
        from PIL import Image, ImageOps
        import numpy as np

        # load the model
        model = load_model('keras_model.h5', compile=False)

In [2]: # create the array of the right shape to feed into the keras model
        # the 'length' or number of images you can put into the array is
        # determined by the first position in the shape tuple, in this case 1.
        data = np.zeros(shape=(1, 224, 224, 3), dtype=np.float32)
        # replace 1011 with the path to your image!
        image = Image.open('L.jpg')
        # resize the image to a 224x224 with the same strategy as in 170:
        # resizing the image to be at least 224x224 and then cropping from the center
        size = (224, 224)
        image = ImageOps.fit(image, size, Image.ANTIALIAS)

        # return the image into a numpy array
        image_array = np.asarray(image)
        # normalize the image
        normalized_image_array = (image_array.astype(np.float32) / 255.0) - 1
        # load the image into the array
        data[0] = normalized_image_array

        # run the inference
        prediction = model.predict(data)
        print(prediction)

[[[0.4969544e-04 9.9865615e-01 1.1719587e-03 2.2188811e-05]]]
    
```

Рис. 15. Пример скрипта с входным изображением жеста «Л»

Нейросеть распознала жест правильно, так как показатель точности = 0.998



Рис. 16. Жест «Л» из словаря РЖЯ

Выполним импорт библиотеки *load_model* (загрузка модели) из *keras.model*. Импортируем библиотеки изображений *Image*, *ImageOps* и библиотеку *numpy*. Запустим написанный нами скрипт. Алгоритм выводит точность совпадения данных букв (А,Л,М,Ж) с их жестом букв из словаря РЖЯ.

Для того, чтобы убедиться в правильности работы нейросети, проведём такие же операции для буквы «А», (рис. 17). Само обозначение буквы «А» – сложнее.



Рис. 17. Жест «А»

Рассмотрим, как нейросеть распознает данный жест. Выставим путь на изображение, с жестом, который обозначает букву «А». Запустим скрипт и получим в ответ точности распознавания жеста, рисунок 18.

```
[9.9639338e-01 1.4965233e-03 1.9741755e-03 1.3589072e-04]
```

А

Рис. 18. Результат точности распознавания жеста «А»

На этот раз, первая точность, которая принадлежит букве А, составляет 0,996. Как можно заметить, остальные точности, примерно, равны единице. Исходя из этого сделать вывод, что в данном случае нейросеть верно распознаёт жест. При рассмотрении других видов жестов, например, буквы «Ж» или буквы «М», нейросеть иногда «путает» жесты, и выдаёт примерно следующие выходные параметры (рисунки 19-21):

```
[[0.41200528 0.0442825 0.53987676 0.00383546]]
```

Рис. 19. Результат точности распознавания жеста «Ж»

При распознавании буквы Ж точность жеста составила 53%. Такое же распознавание получилось для буквы М, так как НС распознала в нём букву Л.

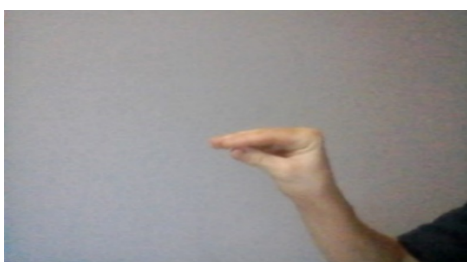


Рис. 20. Жест «Ж»

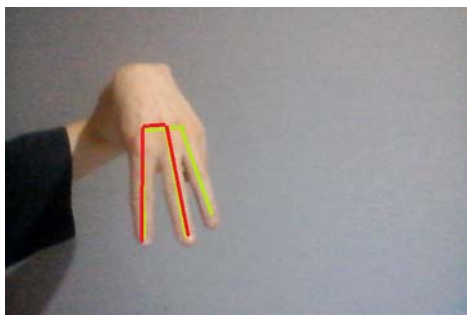


Рис. 21. Жест «М»

Для предотвращения подобных ошибок необходимо предоставлять нейросети больше данных с изображением тех или иных букв.

Выводы

Результатом работы с фреймворком и библиотекой *MediaPipe* является НС, которая способна различать жесты и выдавать достаточно точное соответствие определённого жеста символу русского алфавита. В качестве испытательных буквенных символов при тестировании НС были использованы символы «А», «Л», «М», «Ж», распознавание которых с применением предложенного алгоритма определения сходства жеста русского жестового алфавита с буквами показало высокую точность.

Литература

1. Ибрагимов Р.Ю. «Социальные проблемы лиц с нарушениями слуха» // Психология и педагогика: методика и проблемы практического применения, 2010. №11-1. С. 121-126.
2. Солдафон 2.0, сервер для создания DataSet и изображением жестов. [Электронный ресурс]: <https://sign-language.eu-central-1.elasticbeanstalk.com/> (дата обращения 22.03.2022).
3. Артемов М.Д., Воронова Л.И., Воронов В.И., Гончаренко А.А., Ежов А.А. «Программный комплекс для распознавания жестового языка на основе структурной и параметрической адаптации сверточной нейронной сети» // Программный комплекс для распознавания жестового языка на основе структурной и параметрической адаптации сверточной нейронной сети, 2018. №RU 2018666854. С. 1-20.

4. Документация по MEDIPIPE / материалы сайта компании GitHub. [Электронный ресурс]: https://google.github.io/mediapipe/getting_started/install.html#installing-using-docker (дата обращения 22.03.2022).
5. Описание MEDIPIPE / материалы сайта компании Habr. [Электронный ресурс]: <https://habr.com/ru/post/502440/> (дата обращения 22.03.2022).
6. Воронов В.И., Воронова Л.И., Генчель К.В. «Применение параллельных алгоритмов в нейронной сети для распознавания жестового языка» / Актуальные проблемы инфотелекоммуникаций в науке и образовании (апино 2018). VII Международная научно-техническая и научно-методическая конференция. Сборник научных статей. В 4-х томах. Под редакцией С.В. Бачевского. // Издательство: Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича (Санкт-Петербург), статья в сборнике трудов конференции, 2018. С. 207-212.
7. Воронов В.И., Воронова Л.И., Генчель К.В. Применение параллельных алгоритмов в нейронной сети для распознавания жестового языка. // Телекоммуникации и информационные технологии, 2018. №1. С. 126-130.
8. Библиотека MEDIPIPE. / материалы сайта компании MediaPipe. [Электронный ресурс]: <https://pypi.org/project/mediapipe/> (дата обращения 22.03.2022).
9. Документация JUPITER NOTEBOOK / материалы сайта компании Jupiter. [Электронный ресурс]: <https://jupyter.org/documentation> (дата обращения 22.03.2022).
10. Дополнение к MEDIPIPE [Электронный ресурс]: <https://codepen.io/mediapipe/pen/RwGWYJw/> материалы сайта компании CodePen. (дата обращения 22.03.2022).
11. MediaPipe Online / материалы сайта компании MediaPipe. [Электронный ресурс]: <https://mediapipe.dev/> (дата обращения 22.03.2022).
12. Королькова О.О. Вариативность исполнения жестов в русском жестовом языке (на материале видеословаря русского жестового языка) // В мире научных открытий, 2013. №11-6 (47). С. 187-193.
13. Шакиров Р.И., Артемов М.Д., Воронова Л.И. Подсистема подготовки данных для программно-аппаратного комплекса распознавания жестового языка // Телекоммуникации и информационные технологии. 2021. Т. 8. № 2. С. 101-108.
14. Воронов В.И., Быков А.Д., Воронова Л.И. Проектирование подсистемы детектирования лиц и интерфейса работы с базой данных в программно-аппаратном комплексе биометрической идентификации на основе нейросетевого распознавания лиц // Т-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 4. С. 31-38.
15. Voronov V.I., Zharov I.A., Bykov A.D., Trunov A.S., Voronova L.I. Designing a neural network identification subsystem in the hardware-software complex of face recognition // T-Comm. 2020. Т. 14. № 5. С. 69-76.
16. Трунов А.С., Воронов В.И., Воронова Л.И. Интеграция унаследованных приложений в экосистему научных исследований "Эни" // Т-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 8. С. 33-41.
17. Мохаммад Н., Воронова Л.И. Моделирование кластеризации беспроводной сенсорной сети нейросетевым конструктивным методом // Системы синхронизации, формирования и обработки сигналов. 2021. Т. 12. № 3. С. 4-19.
18. Безумнов Д.Н., Воронова Л.И., Плотникова В.С. Разработка исследовательского стенда для изучения моделей и алгоритмов управления группой наземных мобильных роботов // REDS: Телекоммуникационные устройства и системы. 2021. Т. 11. № 3. С. 3-9.
19. Воронов В.И., Воронова Л.И., Скрыбин В.И., Лукманов К.Д. Разработка лабораторного практикума "подсистема управления безопасностью в программно-аппаратном комплексе умный дом" // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2020. Т. 9. № 2. С. 20-26.
20. Мартынов Д.А., Воронова Л.И. Применение перцептрона для распознавания дактилем русского жестового языка // DSPA: Вопросы применения цифровой обработки сигналов. 2020. Т. 10. № 2. С. 37-46.

USING MACHINE LEARNING IN GESTURE RECOGNITION TASKS

Anton A. Subbotin,

Graduate MTUCI, Moscow, Russia,

anton140299@gmail.com

Voronova L. Ivanovna,

Head Department of ISA&A, Doctor of Physical and Mathematical Sciences,

Professor, MTUCI, Moscow, Russia,

voronova.lilia@yandex.ru

Abstract

Due to the rapid progress in the field of information technology, there are more advanced devices and software that complement and improve the lives of ordinary people and citizens with disabilities. This article explores the use of neural networks in the field of gesture recognition using the MediaPipe framework. The technology design, implementation, training of the neural network, enrichment of the database with gesture images with correlation to classes, testing of the gesture recognition system using the MediaPipe framework were carried out. The result of the study is a neural network that is able to recognize certain gestures of the Russian Sign Language.

Keywords: *Gestures, machine learning, neural network, MediaPipe, Jupyter Notebook, sign language, translator.*

АНАЛИЗ ДАТАСЕТА GOEMOTIONS ДЛЯ ДАЛЬНЕШЕЙ РАЗРАБОТКИ МОДЕЛЕЙ РАСПОЗНАВАНИЯ ЭМОЦИЙ НА БАЗЕ МАШИННОГО ОБУЧЕНИЯ

Юсифов Эльнур Сеймурович,
студент МТУСИ, Москва, Россия,
nurik040404@gmail.com

Синева Ирина Сергеевна,
доцент кафедры ТВиПМ, к.ф.-м.н., МТУСИ, Москва, Россия,
iss@mtuci.ru

Аннотация

Целью работы являлся анализ датасета *GoEmotions* для использования полученных результатов в дальнейшей работе по классификации эмоций. Для уменьшения выбросов данных был предобработан датасет *GoEmotions*: удалены аннотации со степенью согласованности меньше двух, из них выбраны самые популярные, была проведена очистка текста от пунктуации и прочих выбросов, а также проведен стемминг. Полученный текст был представлен в виде векторов с помощью *TF-IDF* меры, а самые важные признаки были выделены с помощью логарифмического отношения шансов. По результатам исследования были определены следующие проблемы: сложная смежная структура классов, сильный дисбаланс классов, обсуждение этих проблем и возможные способы их решения будут представлены в дальнейшем.

Ключевые слова: *machine-learning*, эмоции в тексте, векторное представление текста, анализ данных, *GoEmotions*.

Описание датасета

GoEmotions – это датасет из более чем 58 тысяч обработанных комментариев, извлеченных из *Reddit*, с аннотациями к 28 категориям эмоций [1]. Итоговый размер датасета составляет 211 225 записей.

Датасет содержит следующие категории эмоций: восхищение, веселье, гнев, раздражение, одобрение, забота, замешательство, любопытство, желание, разочарование, неодобрение, отвращение, смущение, волнение, страх, благодарность, горе, радость, любовь, нервозность, оптимизм, гордость, осознание, облегчение, раскаяние, печаль, удивление.

Необработанный набор данных, разбитый на три файла *csv*, включает все аннотации, а также метаданные о комментариях. Каждая строка представляет собой аннотацию одного оценщика для одного примера. Этот файл включает следующие столбцы:

- *text*: Текст комментария (с замаскированными лексемами, как описано в статье).
- *id*: Уникальный идентификатор комментария.
- *author*: Имя пользователя *Reddit* автора комментария.
- *subreddit*: *Subreddit*, к которому принадлежит комментарий.
- *link_id*: Идентификатор ссылки комментария.
- *parent_id*: Идентификатор родителя комментария.
- *created_utc*: Временная метка комментария.
- *rater_id*: Уникальный идентификатор аннотатора.
- *example_very_unclear*: отметил ли аннотатор пример как очень неясный или трудный для обозначения (в данном случае он не выбрал никаких обозначений эмоций).
- отдельные столбцы, представляющие каждую из категорий эмоций, с бинарными метками (0 или 1)

Анализ датасета

В ходе исследования была поставлена задача провести эксперименты и верифицировать методологию в сопоставлении с выполненными исследователями из *Google* в их статье [1]. В работе авторы описывают создание датасета, анализ данных и классификацию эмоций с помощью нейронных сетей (авторы используют архитектуру *BERT*[2]), а также делают выводы по результатам исследования.

Общая методология, используемая в данной статье, базируется на работе [4].

Всего в эксперименте принимало участие 82 человека. Около 1% всех аннотаций были помечены оценщиками, как неясные.

В работе была проанализирована согласованность среди оценщиков, где выяснилось, что в 92% примеров 2 и более оценщика согласны по крайней мере с одним обозначением эмоций.

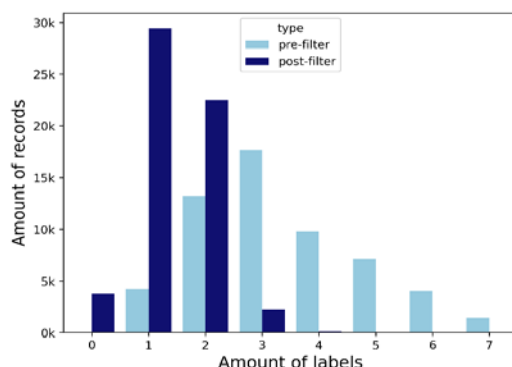


Рис. 1. Количество маркировок эмоций у текстов до и после фильтрации по согласованности больше двух

Так же было рассчитано логарифмическое отношение шансов i -го слова находиться в j множестве слов эмоции. Благодаря этому можно построить таблицу показывающую степень принадлежности каждого слово к конкретному множеству слов эмоции.

Логарифмическое отношение шансов для слов рассчитывается по формулам (1-3):

$$L_i^j = \frac{X_i^j + X_i}{(N^j + N) + (X_i^j + X_i)} \tag{1}$$

$$\sigma_i^j = \sqrt{(X_i^j + X_i)^{-1} + (X_i^{\bar{j}} + X_i)^{-1}} \tag{2}$$

$$\delta_i^j = \frac{\ln(L_i^j) - \ln(L_i^{\bar{j}})}{\sigma_i^j} \tag{3}$$

Где X – множество содержащее пары (слово, количество) во всех текстах, X^j – его подмножество определяющее такие пары для j эмоции, $N^j = \|X^j\|$ – сумма количества слов в подмножестве X^j , N – сумма количества слов в множестве X , а δ_i^j – логарифмическое отношение шансов i -го слова находиться в j множестве слов эмоции. Самые популярные слова для каждой из эмоций приведены в таблице 1.

Таблица 1

Самые популярные слова для каждой категории эмоций

Эмоция	Ассоциации	
	Слово	Частота
Admiration	Great	0.014
Amusement	Lol	0.043
Anger	Fuck	0.013
Annoyance	Annoying	0.002
Approval	Agree	0.003
Caring	You	0.062
Confusion	Confused	0.005
Curiosity	Curious	0.004
Desire	Wish	0.033
Disappointment	Disappointing	0.002
Disapproval	Not	0.024
Disgust	Disgusting	0.011
Embarrassment	Embarrassing	0.008
Excitement	Excited	0.012
Fear	Scared	0.010
Gratitude	Thanks	0.049
Grief	Died	0.025
Joy	Happy	0.022
Love	Love	0.072
Nervousness	Nervous	0.009
Pride	Proud	0.044
Realization	Realize	0.005
Relief	Glad	0.025
Remorse	Sorry	0.076
Sadness	Sad	0.016
Surprise	Wow	0.019
Neutral	[NAME]	0.023

Предобработка датасета

Были выбраны колонки *text*, *rater_id*, а также бинарные метки, представляющие каждую из категорий эмоций. Строки со столбцом *example_very_unclear* являющимся истинным были исключены из датасета, так как не обладают бинарной разметкой. Во время исследования из более чем 200 тысяч аннотаций были выбраны те, частота эмоций которых была максимальной при согласованности между оценщиками больше двух.

Комментарии пользователей так же были предобработаны с помощью регулярных выражений (рис. 2).

```
punct_chars = list((set(string.punctuation) | {
    ">", "€", "-", "_", "~", "|", "€", "™", "™", "™", "™", "™", "™",
    "™"
}) - set(["#"]))
punct_chars.sort()
punctuation = "".join(punct_chars)
replace = re.compile("[%s]" % re.escape(punctuation))

def text_preprocessor(text):
    """Clean text."""
    if isinstance(text, float):
        return []
    # lower case
    text = text.lower()
    # eliminate urls
    text = re.sub(r"http\S*|\S*\.\com\S*|\S*www\S*", " ", text)
    # eliminate @mentions
    text = re.sub(r"\s@\S+", " ", text)
    # substitute all other punctuation with whitespace
    text = replace.sub(" ", text)
    # replace all whitespace with a single space
    text = re.sub(r"\s+", " ", text)
    # strip off spaces on either end
    text = text.strip()
    words = text.split()
    return [w for w in words if len(w) > 2]
```

Рис. 2. Функция предобработки текста

Данная функция переводит текст в нижний регистр, удаляет ссылки, различную пунктуацию и пробелы. Затем был проведен стемминг слов в предложениях, чтобы привести их в инфинитивную форму.

Так же в ходе работы эмоции были объединены в более крупные группы согласно 6 базовым эмоциям П. Экмана (рис. 3) добавив к этому списку нейтральную.

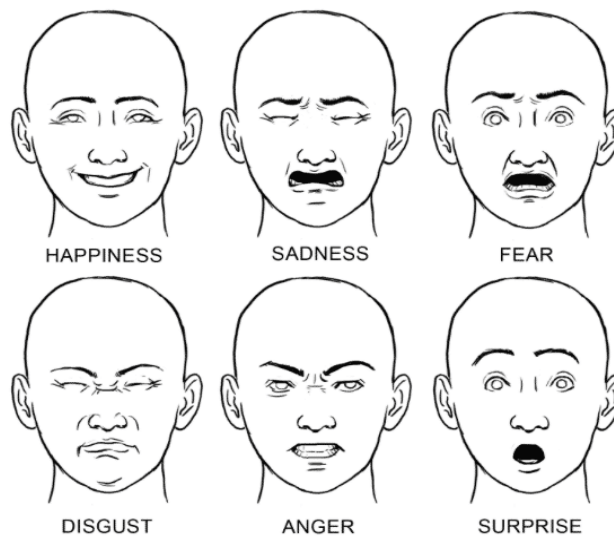


Рис. 3. 6 базовых эмоций П. Экмана

Диаграмма переразметки представлена на рисунке 4.

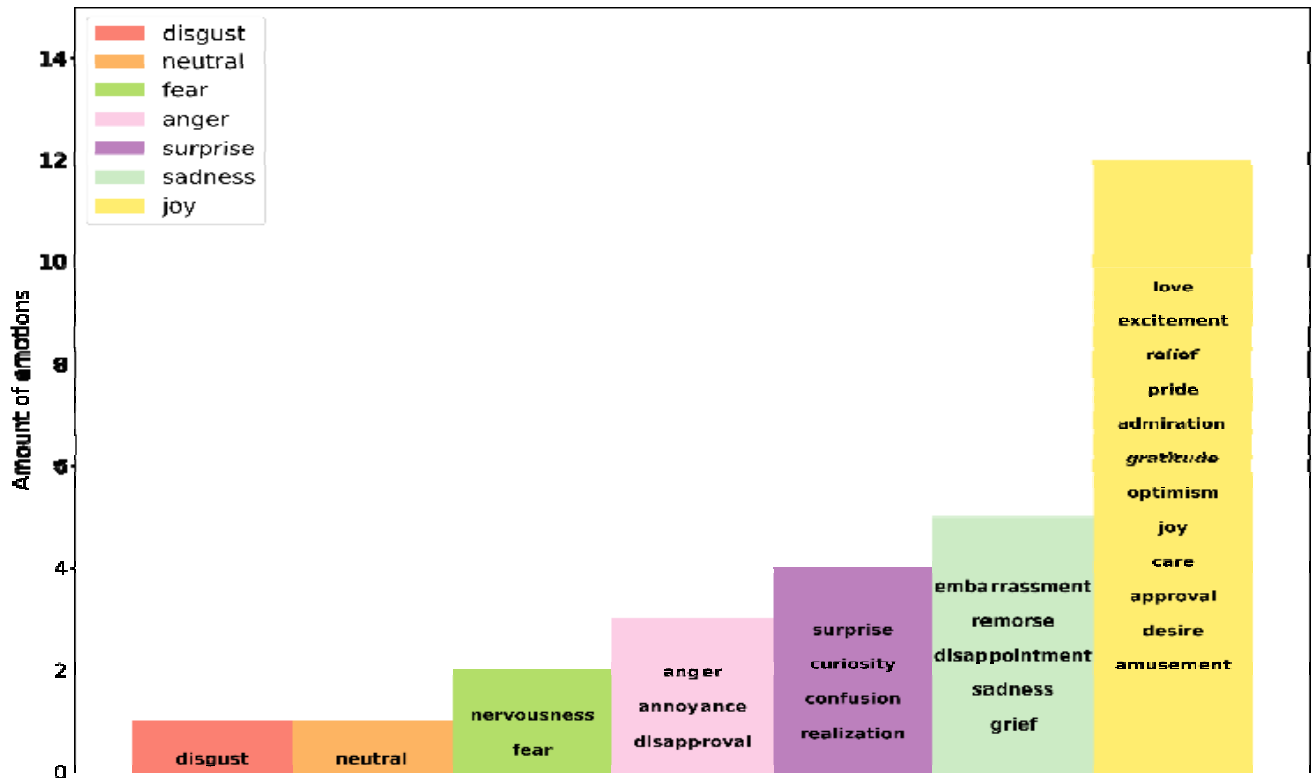


Рис. 4. Диаграмма переразметки 28 эмоций к 7 эмоциям

В ходе предобработки было произведено разделение датасета на обучающую (67%) и валидационную (33%) выборки. В датасете наблюдался сильный дисбаланс классов, он показан на рисунке 5.

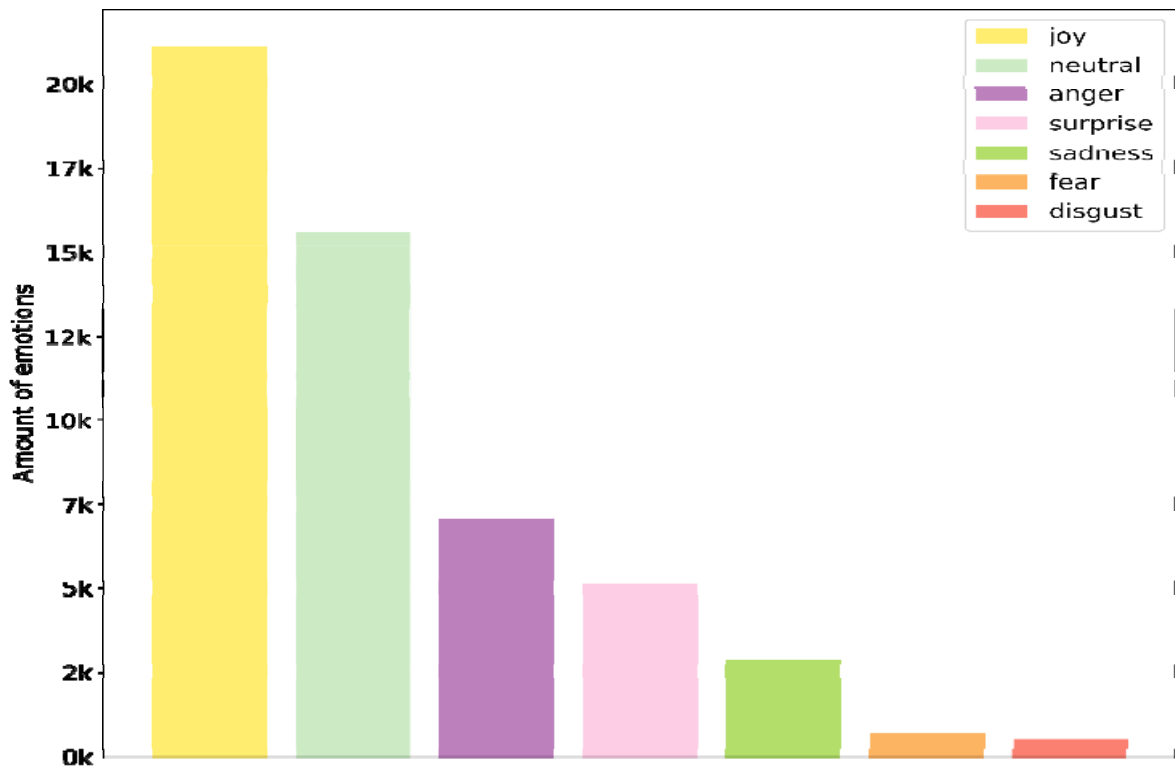


Рис. 5. Дисбаланс классов в датасете

Тексты обучающей и валидационной выборки были представлены в виде векторов с помощью меры *TF-IDF*. Существуют также различные методы текстового представления, основанные на *word2vec* [3], *CBoW* и т.д. Все эти методы представляют собой "черный ящик", где неясно, что происходит на каждом слое.

Выводы

Полученный предобработанный текст позволяет убрать шумы при проведении дальнейшей классификации: удаление пунктуации приводит текст к единой форме, где между словами находятся только пробелы, приведение к нижнему регистру избавляет от лишних признаков (ошибки при вводе, слова с заглавной буквы и т.п.), стемминг позволяет привести все слова к форме инфинитивов. Всё это уменьшает итоговую размерность векторного представления текста, а также приводит текст к единой форме.

Представленный датасет имеет высокий дисбаланс классов, что в некотором роде затрудняет его дальнейшую классификацию. В дальнейших статьях будут разобраны методы решения этой проблемы.

Полученные логарифмические отношения шансов были использованы в качестве критерия важности признаков в векторном представлении текста. В результате были выделены признаки, важность которых выше среднего. Они были извлечены из векторов и в последствии размерность этих представлений была уменьшена до трёхмерного пространства с помощью алгоритма *LSA* на базе которого был построен рисунок 6.

По рисунку 6 видно, что полученное пространство признаков классификации в рамках классов имеет сложную смежную структуру, что в будущем может сказаться на результатах.



Рис. 6. Наиболее важные признаки, извлеченные из набора данных с помощью логарифмического отношения шансов, представлены в трехмерном пространстве с помощью *LSA*

В качестве дальнейшего направления работы была поставлена цель упростить этап классификации описанный в [1], приведя его к классическим алгоритмам машинного обучения. Для этого будут использованы результаты, полученные в рамках данного исследования.

Литература

1. D. Demszky, D. Movshovitz-Attias, J. Ko GoEmotions: A Dataset of Fine-Grained Emotions // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. С. 4040-4054.
2. A. Vaswani, N. Shazeer, N. Parmar Attention is all you need // Advances in neural information processing systems. 2017. №30. С. 5998-6008.
3. T. Mikolov, I. Sutskever, K. Chen Distributed representations of words and phrases and their compositionality // Advances in neural information processing systems. 2013. №26. С. 3111-3119.
4. S. Hassan, J. Ahamed, K. Ahmad Analytics of machine learning-based algorithms for text classification // Sustainable Operations and Computers. 2022. №3. С. 238-248.

ANALYSIS OF GOEMOTIONS DATASET FOR FURTHER DEVELOPMENT OF EMOTION RECOGNITION MODELS BASED ON MACHINE LEARNING

Elnur S. Yusifov,
Student MTUCI, Moscow, Russia,
nurik040404@gmail.com

Irina S. Sineva,
Associate Professor of the Department of PT&AM, Ph.D. in Physics and Mathematics,
MTUCI, Moscow, Russia,
iss@mtuci.ru

Abstract

Goal of research was to analyze the GoEmotions dataset in order to use the results in further work on emotion classification. To reduce outliers, the GoEmotions dataset was preprocessed: annotations with a consistency of less than two were removed, the most popular ones were selected from them, the text was cleaned of punctuation and other outliers, and stemming was performed. The resulting text was represented as vectors using TF-IDF measure, and the most important features were highlighted using log-ratio odds. Based on the results of the study, the following problems were identified: complex contiguous class structure, strong class imbalance, discussion of these problems and possible ways to solve them will be presented in the next paper.

Keywords: machine-learning, emotions in text, vector representation of text, data analysis, GoEmotions

«Цифровые технологии радиосвязи и телерадиовещания»

СНИЖЕНИЕ ИСКАЖЕНИЙ ЗВУКОВОГО СИГНАЛА ПРИ ЕГО ПЕРЕДИСКРЕТИЗАЦИИ

Гавриков Олег Владимирович,
магистрант МГУСИ, Москва, Россия,
oleg.gavrikov13@gmail.com

Мальгин Алексей Дмитриевич,
магистрант МГУСИ, Москва, Россия,
malgin.alexey1999@gmail.com

Орлов Владимир Георгиевич,
главный специалист отдела ОНИРС, к.т.н., МГУСИ, Москва, Россия,
v.g.orlov@mtuci.ru

Попов Олег Борисович,
профессор кафедры ТиЗВ, к.т.н., МГУСИ, Москва, Россия,
o.b.popov@mtuci.ru

Самарин Евгений Васильевич,
студент МГУСИ, Москва, Россия,
7706053@gmail.com

Аннотация

В статье рассмотрены основные способы преобразования частоты дискретизации звуковых сигналов: с переходом во временную область, с переходом в частотную область, а также способ с переходом к аналоговому представлению сигнала. В результате анализа данных способов был сделан вывод о том, что наибольшая эффективность их использования достигается при применении тестовых сигналов, на которых была проведена оценка параметров звукового сигнала. Однако при передаче реального сигнала вносятся существенные искажения. С учётом этого предложена модель алгоритма изменения частоты дискретизации в частотной области, минимально изменяющего форму сигнала и снижающего искажения

Ключевые слова: аналоговый сигнал, передискретизация, частота, звуковой сигнал, преобразование, искажение, цифровой сигнал.

Введение

Преобразование аналоговых сигналов в цифровые является одной из базовых мер для обеспечения удобства работы с информацией в современном мире. Процесс преобразования аналогового сигнала в цифровой возможен благодаря дискретизации – операцией выделения отсчетов аналогового сигнала, то есть его мгновенных значений с определенным интервалом времени, и далее квантованием по уровню. Дискретизация позволяет представить непрерывный сигнал в виде совокупности дискретных значений. При этом, период времени между соседними выборками имеет одинаковую длительность, которая соответствует периоду дискретизации (T_D), а обратная ей величина будет являться частотой дискретизации ($F_D = 1/T_D$). Частота дискретизации любого аналогового сигнала может быть определена по теореме Котельникова [1].

На рисунке 1 представлена упрощенная структурная схема канала вещания, на которой показаны его звенья с их разрядностью и частотой дискретизации. На данной схеме видно, что канал вещания можно разделить на три тракта: тракт формирования программ звукового вещания – ТФП (РД-1), тракт первичного распределения программ – ТПРП (РД-2) и тракт вторичного распределения программ – ТВРП (РД-3) [10].

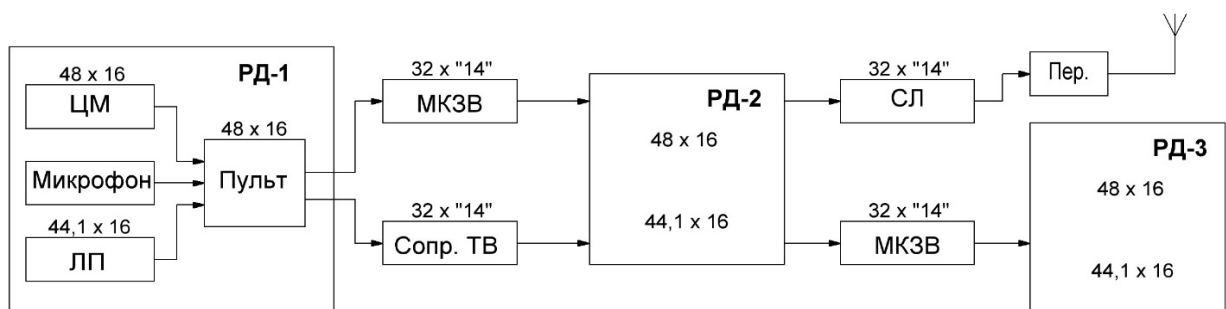


Рис. 1. Структурная схема тракта звукового вещания

На рисунке 1: ЦМ – цифровой магнитофон, ЛП – лазерный проигрыватель, РД – радиодом, Пер. – передатчик, СЛ – соединительная линия, МКЗВ – междугородний канал звукового вещания.

При анализе схемы на рисунке 1 можно отметить, что частота дискретизации меняется при переходе между звеньями и трактами канала звукового вещания, что в свою очередь подтверждается международными рекомендациями и стандартами. В настоящее время для звуковых сигналов стандартизированы несколько частот дискретизации, каждая из которых предназначена для использования в определённых системах звуковой техники:

- 48 кГц – используется в студиях тракта формирования программ (ТФП) для транспортирования высококачественного сигнала без потерь (ИКМ 16 разрядов), например, от цифрового магнитофона (ЦМ) к цифровому пульта звукорежиссёра;

- 44,1 кГц – применяется в проигрывателях компакт-дисков, которые используются для хранения архивных программ в радиодоме при 16-разрядном ИКМ с линейной шкалой;

- 32 кГц – используется в системах передачи и распределения радиопрограмм по сетям вещания компрессированного звукового сигнала (ИКМ 10-12 разрядов), который по субъективному восприятию не отличается от сигнала с 14-разрядной шкалой дискретизации;

- 31,25 кГц – применяется в цифровом стерео-сопровождении телевизионного сигнала [9].

В статье будет рассмотрен и промоделирован способ изменения частоты дискретизации с промежуточным переходом в частотной области, который позволяет обрабатывать частотные коэффициенты в области частот с минимальной энергией, что позволяет минимизировать искажение сигнала на выходе. Данный алгоритм называется алгоритмом преддискретизации и был разработан на кафедре «Телевидения и звукового вещания» МТУСИ.

1. Основные способы изменения частоты дискретизации

Известно множество способов изменения частоты дискретизации звукового сигнала, однако наиболее эффективными и используемыми являются три способа: цифровой способ преобразования во временной области; цифровой способ преобразования в частотной области; аналоговый способ цифрового преобразования частоты дискретизации.

Аналоговый способ преобразования является наименее эффективным, так как в процессе его применения, цифровой сигнал с определенной частотой дискретизации необходимо подать на вход ЦАП, после чего, полученный аналоговый сигнал подать АЦП, в котором и задаются параметры новой частоты дискретизации и разрядности преобразованного цифрового сигнала. Данный способ обладает высокой вычислительной сложностью из-за большого количества преобразований и соответствующих операций, а также связан с появлением шумов квантования из-за нелинейности АЦП при использовании неравномерной шкалы квантования, появлением джиттера и других негативных факторов, обусловленных не идеальными характеристиками устройств АЦП и ЦАП [4].

Как известно для измерения влияния шума квантования на полезный сигнал используют параметр, называемый отношением сигнал/шум (ОСШ или SNR) [8], который представляет собой отношение мощности полезного сигнала к мощности шума (1).

$$\frac{P_c}{P_{KB}} = 10 \lg \frac{12 \cdot (U_c / \sqrt{2})^2}{(U_c / 2^{n-1})^2} = 6 \cdot n + 1,8 \text{ [дБ]}, \quad (1)$$

где: n – разрядность квантования, U_c – амплитуда гармонического сигнала.

Из формулы (1) следует, что на качество сигнала влияет разрядность квантования, поэтому для достижения на выходе ЦАП после цифрового преобразования идеального аналогового звука, с такими же характеристиками, как у аналогового сигнала на входе АЦП, необходимо квантовать его с разрядностью $n = 24$ бит и частотой дискретизации 192 кГц [12,14].

Помимо того, что использование таких сложных устройства АЦП и ЦАП из-за их высокой стоимости в системах звукового вещания нерентабельно, отсутствуют каналы для передачи таких сигналов по сетям звукового вещания.

Как видно на рисунке 1 максимальное качество звукового сигнала нормируется использованием в микшерном пульте РД в тракте ТПФ (РД 1) цифрового ИКМ преобразования с $F_D = 48$ КГц и 16 битовой шкалой квантования. Эти параметры существенно ниже упомянутых параметров, требуемых для обеспечения идеального качества вещательного звукового сигнала при цифровом преобразовании. Помимо низкой помехозащищённости сигнала из-за недостаточно высоких значений F_D и n одной из причин снижения качества аудиосигнала являются искажения, вносимые непосредственно АЦП-ЦАП оборудованием, так как даже при идеальном квантовании возникают искажения, достигающие 3 дБ, а с учётом невозможности идеальной реализации ЦАП-АЦП преобразователей (например, плохая помехозащищённость на этапе разработки, конструирования и серийного производства), искажения могут возрастать до 4,5 дБ. [13]. Помимо самих преобразователей ЦАП-АЦП, в их обвязку часто входят элементы ФНЧ, на входе и выходе которые вызывают искажения амплитудных и фазо-частотных характеристик.

Исходя из вышеизложенного, следует, что использование цифровых методов преддискретизации может исключить внесение в сигнал искажений аппаратного происхождения, а также поможет избежать искажений, возникающих из-за повторных квантования и преобразования звукового сигнала в трактах канала звукового вещания (рис. 1). При этом цифровым методам преобразования частоты дискретизации в сравнении с аналоговыми свойственны следующие преимущества:

- цифровой сигнал более устойчив к помехам малого уровня (например, электромагнитным наводкам), что дает возможность выпускать цифровую технику с абсолютно одинаковыми характеристиками, в то время как подобное недостижимо в аналоговых устройствах в виду большой схмотехнической сложности;
- возможность интеграции нескольких цифровых устройств в СБИС или ПЛИС, что в свою очередь ведет к микроминиатюризации аппаратных средств;
- возможность использовать, как универсальные алгоритмы для преобразования частот дискретизации по типу «из любой в любую» (кратных и некратных), так и специальные быстрые алгоритмы для преобразования заранее известного набора ЧД, которые используются в каналах звукового вещания или содержатся в документах ГОСТ [9,10].

Кратность частот на входе/выходе условного передискретизатора является основой для цифрового преобразования частоты дискретизации. Процесс нахождения наименьшей кратной целочисленной частоты присутствует практически всегда в тракте радиовещания, в той или иной форме – в виде прибора или алгоритма.

Ниже приведена в таблице 1 приведены данные кратности основных частот дискретизации, используемых в трактах канала звукового вещания [4].

Таблица 1

Кратность основных частот дискретизации для канала звукового вещания

F _{2д} \ F _{1д}	48	44,1	32	31,25
48	1	160/147	3/2	192/125
44,1	147/160	1	441/320	882/625
32	2/3	320/441	1	128/125
31,25	125/192	625/882	125/128	1

Структурная схема алгоритма изменения частоты дискретизации во временной области приведена на рисунке 2, [7].

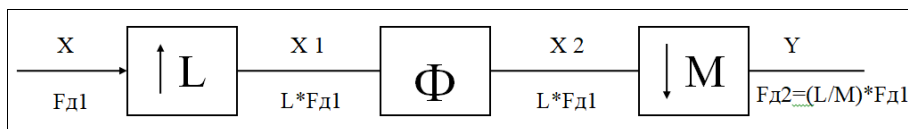


Рис. 2. Структурная схема алгоритма преобразования ЧД во временной области [13]

2. Изменение ЧД в устройствах ЦАП и АЦП

Устройства ЦАП и АЦП также позволяют изменить частоту дискретизации. Принцип АЦП, основанный на приеме аналогового сигнала и его последующей модуляции дельта сигналом с высокой частотой дискретизации, широко используется в современной технике [12]. При этом с использованием только одного опорного напряжения, кратного или равного шагу дискретизации, можно фиксировать значения уровня входящего сигнала и записывать их на выходе АЦП в цифровом (бинарном) формате. При этом положительные и отрицательные значения сигнала, преобразовываются в 16-24 разрядную последовательность, записанную в ЗУ в виде последовательности нулей и единиц [15]. Минимальное значение уровня сигнала с выхода АЦП рассчитывается как $U_{ВХmin}=U_{Оп} / N$, где N – разрядность системы. На рисунке 3 приведены графики сигнала на входе и выходе АЦП с меньшей и (слева) большей (справа) разрядностью N [16].

Данный способ преобразования аналогового сигнала в цифровой позволяет упростить процесс кодирования. При этом существенное преимущество данного способа заключается в отсутствии необходимости использования сложного аналогового входного ФНЧ (anti-aliasing filter), стоимость которого в устройстве АЦП достигает 80%. Даже при частоте дискретизации 2,8 МГц используемый на выходе ЦАП простой ФНЧ, собранный только из резисторов и конденсаторов, обеспечит требуемую линейность в диапазоне звуковых частот.

3. Способ изменения ЧД с минимальными потерями качества сигнала

Предлагаемый способ изменения ЧД основан на использовании преобразования частоты дискретизации в частотной области и заключается в манипуляциях спектральными отсчетами цифрового сигнала в частотной области с последующим преобразованием новых спектральных отсчетов во временную область.

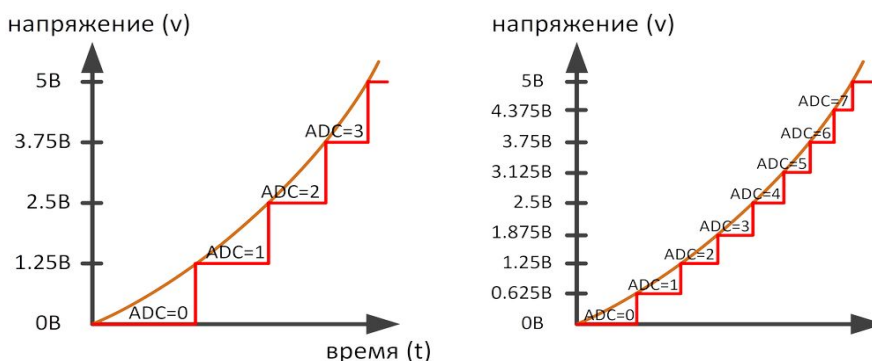


Рис. 3. Графики сигнала на входе/выходе и выхода АЦП [12]

На рисунке 4 представлена структурная схема предложенного алгоритма преобразования ЧД. Она состоит из 4 блоков, соответствующих этапам преобразования: от синтеза самого тестового сигнала до преобразования во временную область.



Рис.4. Структурная схема алгоритма преобразования частоты дискретизации в частотной области

Для эффективного использования данного алгоритма необходимо учитывать, что каждый частотный коэффициент отражает разное количество энергии, что предполагает возможность режекции коэффициентов с наименьшим количеством энергии. При этом предлагается, при добавлении коэффициентов для сохранения их целочисленного значения на интервале наблюдения, который должен оставаться одинаковым на обеих частотах, просто вставить нулевые значения соответствующих добавляемых коэффициентов.

4. Моделирование разработанного алгоритма

На основе исследований, приведенных [2, 8, 17-22], сигнал ошибки дискретизации имеет прямую зависимость от длительности интервала наблюдения. На основе данных, полученных в результате исследований, был составлен график зависимости вероятности ошибки передискретизации от временной длины и типа окна (рис. 5). Согласно исследованиям [4], было выявлено, что наименьшее искажение при изменении частоты дискретизации имеет оконная функция Наттолла. Поэтому, данная оконная функция применена в этой работе. Исследования также показали, что при наложении на сигнал оконной функции Хэмминга, заметные искажения появляются лишь на краях интервала наблюдения, в то время как искажения в его середине в два раз меньше, в сравнении с оконной функции Наттолла [11].

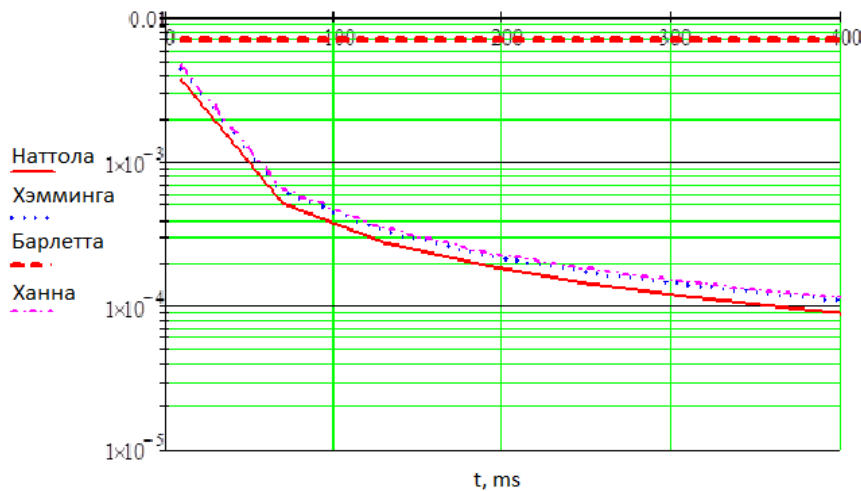


Рис. 5. График зависимости вероятности ошибки передискретизации от длины и типа окна

При необходимости выполнить минимизацию шумов в основной части интервала наблюдения целесообразно использование функции Хэмминга. Согласно данным исследования в [3] и результатам, приведённым на рисунке 5, оптимальная длина интервала наблюдения составляет примерно 200 мс. С учётом этого в данной работе была выбрана длина окна равная 240 мс.

В настоящее время существует множество программных средств для математического моделирования различных процессов в области электроника, радиотехника, механики и т.д. Системы моделирования могут быть графическими, в которых решение задач моделирования происходит с помощью составления последовательности структурных блоков, или математическими, где вводятся конкретные формулы и выражения. Помимо этого, широко используются системы моделирования на основе различных языков программирования.

В качестве компьютерной программы для моделирования поставленной в данной статье задачи использовался Matlab без применения Simulink. Данная программа была выбрана из-за свойственной ей универсальности, гибкости и большого набором встроенных инструментов. Помимо этого, она весьма эффективна при работе с матрицами и массивами данных, в качестве основного базиса представления информации, что отвечает специфики обработки аудио сигнала [2].

Моделирование алгоритма выполнялось в программной среде моделирования Matlab для реального вещательного сигнала с частотой дискретизации $F_d = 44,1$ кГц, разрядностью 16 бит и длительностью 2517 мс. Сигнал является речевым, и его временная диаграмма приведена на рисунке 6.

После того, как файл с исходным сигналом в несжатом формате WAV загрузился в среду Matlab, было произведено последовательное выполнение алгоритма по следующей схеме изменения частоты дискретизации: 44,1 кГц – 48 кГц – 32 кГц – 48 кГц – 32 кГц.

Применение конкретной оконной функции частично осуществлялось с помощью библиотечной функции Matlab. В данном случае использовалась взвешивающая функция Наттола с длительностью окна 240 мс и половинным перекрытием (50%). Сама оконная функция вызывается следующей командой: «nuttalwin()». Для применения взвешивающей функции к каждому отдельному окну (фрейму) с перекрытием использовалась сторонняя функция «v_enframe», являющаяся составной частью «Voicebox», который был создан в Имперском Колледже Лондона [12]. Для половинного перекрытия использовался конечный цикл «for».

После применения взвешивающей оконной функции с перекрытием происходит переход в частотную область, при этом количество частных отсчетов соответствует количеству временных.

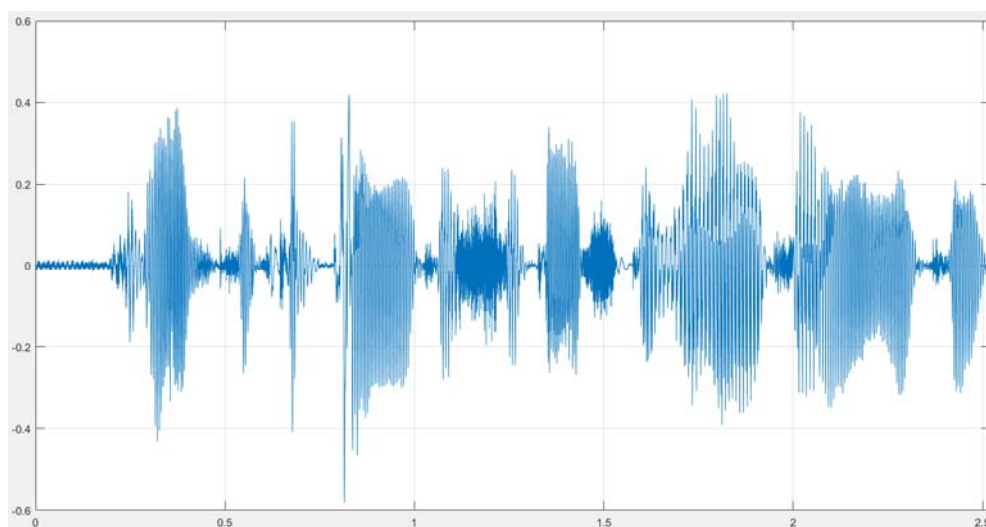


Рис.6. Исходный вещательный сигнал для передискретизации

Для того чтобы осуществлять манипуляцию отсчётами в частотной области необязательно использовать конструкции с циклами, как это обычно принято для операций над матрицами в других средах программирования. Можно проводить удаление или добавление частотных составляющих в спектр с помощью простых операций с матрицами, среда Matlab позволяет это сделать.

При ЧД 44,1 кГц сигнал состоит из 111000 дискретных отсчётов; при частоте 48 кГц должно получиться 120820 отсчётов, а при частоте 32 кГц должно быть 80544 отсчёта. Преобразование сигнала от частоты дискретизации 44,1 кГц к ЧД 48 кГц осуществляется с помощью операции добавления 9820 нулевых отсчётов в частотной области. Переход от 48 кГц к 32 кГц производится с помощью вычитания 40276 частотных отсчётов с минимальной мощностью в области высоких частот. Соответственно для обратного перехода к частоте дискретизации 48 кГц в частотной области обратно возвращаются 40276 отсчёта, но уже нулевых. Таким же образом производится и последующее преобразование к ЧД 32 кГц. С целью достижения чистоты эксперимента, после каждого этапа преобразования сигнал возвращался во временную область.

Одним из параметров оценки эффективности алгоритма может являться время выполнения операций. Для вычисления этого показателя в Matlab присутствует встроенная функция оценки времени. Она вызывается с помощью команд: «tic...toc». Время выполнения операций в случае четырёх этапов передискретизации составляет 190 мс. Это считается достаточно хорошим результатом, так как в нашем случае при длительности фрейма 240 мс за данный временной интервал должен успеть обработаться следующий фрейм, при этом в случае одного этапа передискретизации время обработки будет гораздо меньше.

После прохождения исходного вещательного сигнала всех четырёх этапов изменения частоты дискретизации на выходе генерируется файл в формате WAV 16 бит с частотой $dF_d = 32$ кГц. Для сравнения данный сигнал показан на рисунке 7 в программе Cool Edit Pro вместе с исходным сигналом.



Рис. 7. Исходный сигнал (сверху) и сигнал после 4-х этапов передискретизации

Для сравнения, 4-х этапная передискретизация была также проведена с помощью встроенной функции Matlab «resample()» и встроенной функции программы Cool Edit Pro – «Convert Sample Type». Для оценки различий спектров сигналов были получены спектры итогового сигнала и передискретизированных после 4-х этапов передискретизации с использованием разных алгоритмов (рис. 8).

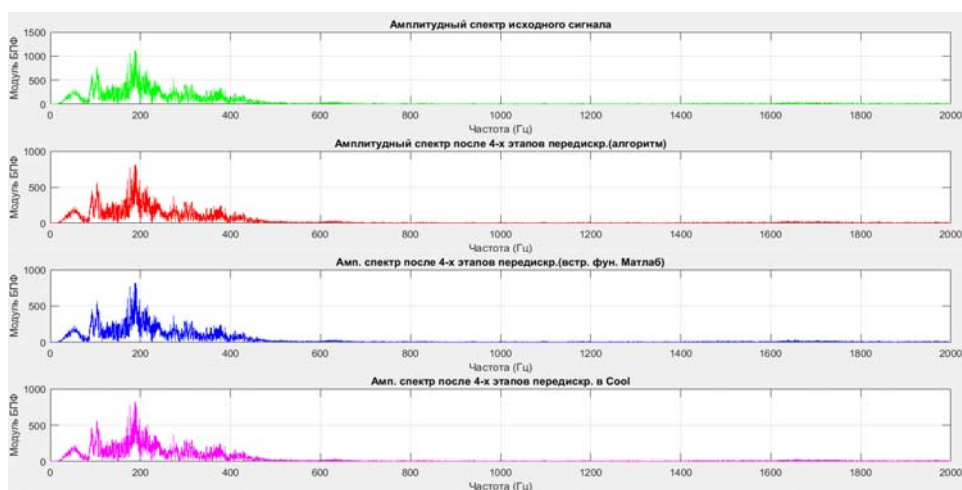


Рис. 8. Спектры исходного сигнала и после 4-х этапной передискретизации с использованием разных алгоритмов

Также, для наглядности сравним спектры исходного сигнала и сигнала после всех этапов изменения ЧД с помощью предложенного алгоритма в программе Cool Edit Pro (рисунок 9).

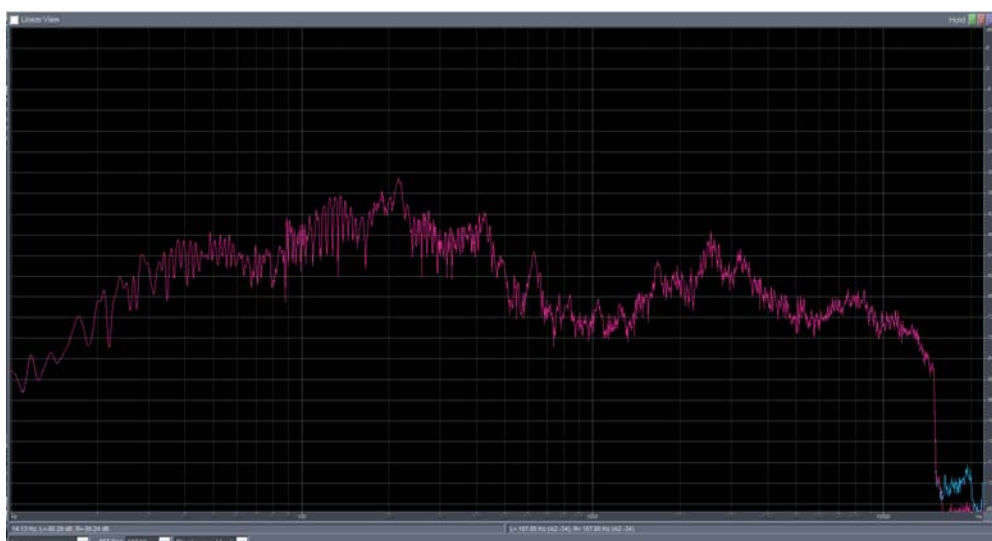


Рис.9. Сравнение спектров исходного сигнала и после передискретизации с помощью предложенного алгоритма

На рисунке 9 видно, что передискретизация с использованием предложенного алгоритма не привела к серьезным изменениям спектра, заметно только уменьшение амплитуды шумов на частотах выше 16 кГц. Такие изменения спектра не слышны человеческому уху.

Выведем теперь спектры исходного сигнала и сигнала после передискретизации с помощью встроенной функции в Matlab (рис. 10).

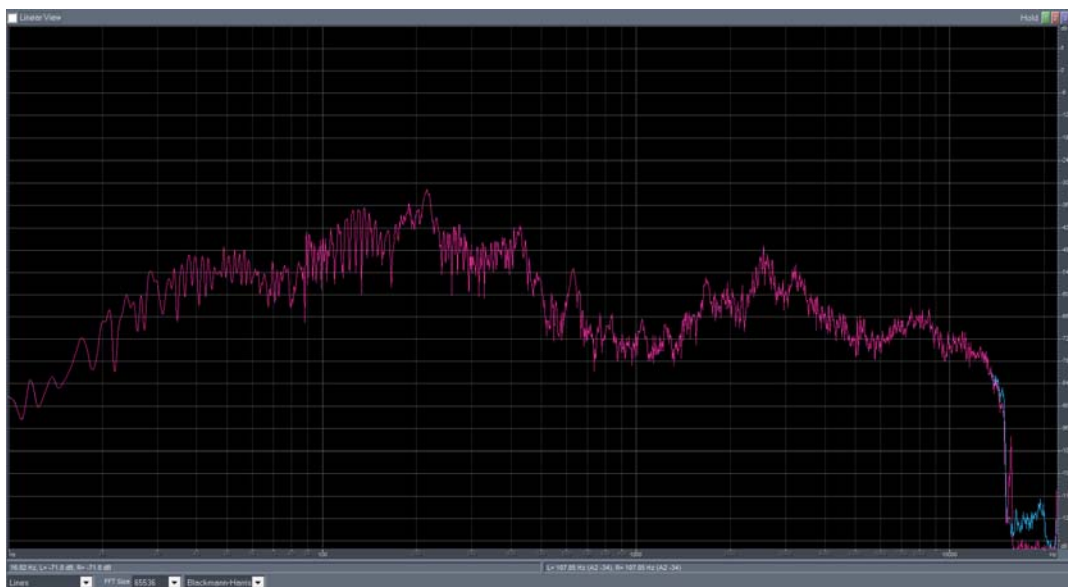


Рис. 10. Сравнение спектров исходного сигнала и после передискретизации с помощью Matlab

По рисунку 10 видно, что передискретизированный сигнал с помощью функции Matlab больше отличается от исходного, так как в области 15,5 кГц виден резкий скачок амплитуды (на 20 дБ) относительно исходного. Скорее всего данная высокочастотная составляющая тоже не будет слышна человеческому уху.

Сравним также спектр сигнала преобразованного в самом Cool Edit Pro с исходным спектром сигнала (рис. 11).

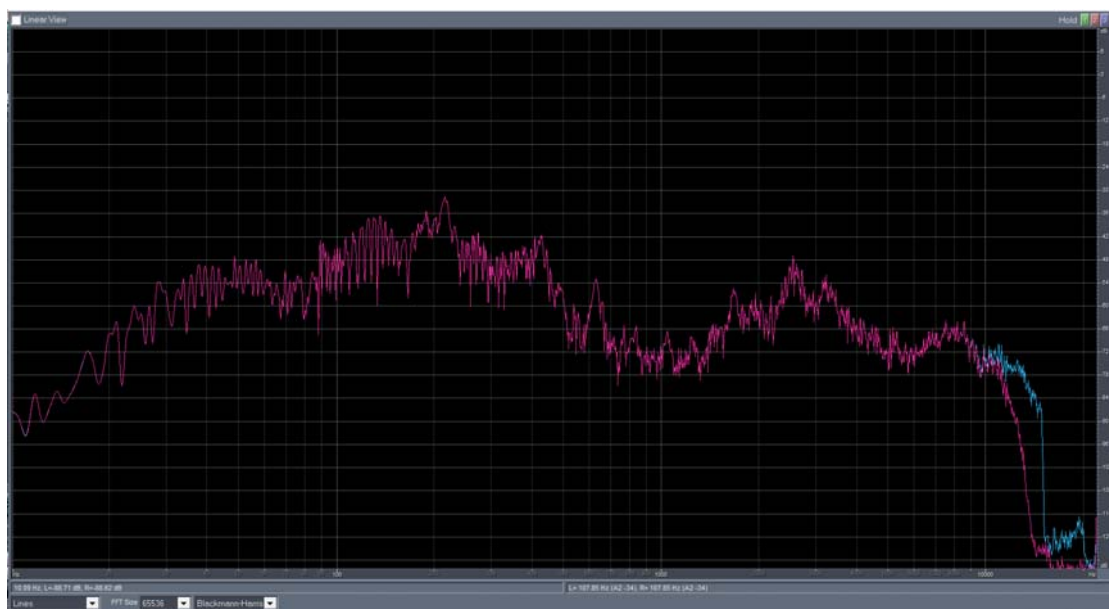


Рис.11. Сравнение спектров исходного сигнала и сигнала после передискретизации в Cool Edit Pro

Как видно на рисунке 11 спектр передискретизированного сигнала с помощью Cool Edit Pro существенно отличается от спектра исходного сигнала. Это проявляется в виде сокращения диапазона частот – после 10 кГц спектр начинает резко уменьшаться. Данные изменения спектра чётко определяются на слух.

Далее рассмотрим сравнение исходного сигнала с сигналом ошибки при использовании каждого из алгоритмов передискретизации. Сигнал ошибки вычислялся, как было рассмотрено выше, с помощью функции «Mix Paste» в Cool Edit Pro, позволяющей сложить сигналы в противофазе.

На рисунке 12 показано сравнение спектров исходного сигнала и сигнала ошибки для предложенного алгоритма, а на рисунке 13 представлены статистические данные сигнала ошибки.

На спектрограммах на рисунках 12 и 13 видно, что шумовая составляющая маскируется по уровню полезным сигналом и находится в зоне низких уровней. Это свидетельствует о том, что человеческое ухо не услышит сигнал ошибки обусловленный передискретизацией, так как он перекрывается на всех частотах звукового диапазона.

Аналогичные спектры только для встроенной функции Matlab приведены на рисунках 14 и 15.

Если рассматривать показатели шумового сигнала для встроенной функции Matlab на рисунке 14, то можно отметить, что упомянутая ранее составляющая звукового сигнала на частоте 15,5 кГц не маскируется полезным сигналом. Это видно, как на спектрограмме (рис. 14), так и на статистической гистограмме (рис. 15) [7]. Составляющую шумового сигнала на частоте 15,5 кГц вполне различима натренированным экспертом.



Рис. 12. Спектры исходного сигнала и сигнала ошибки при использовании предложенного алгоритма передискретизации

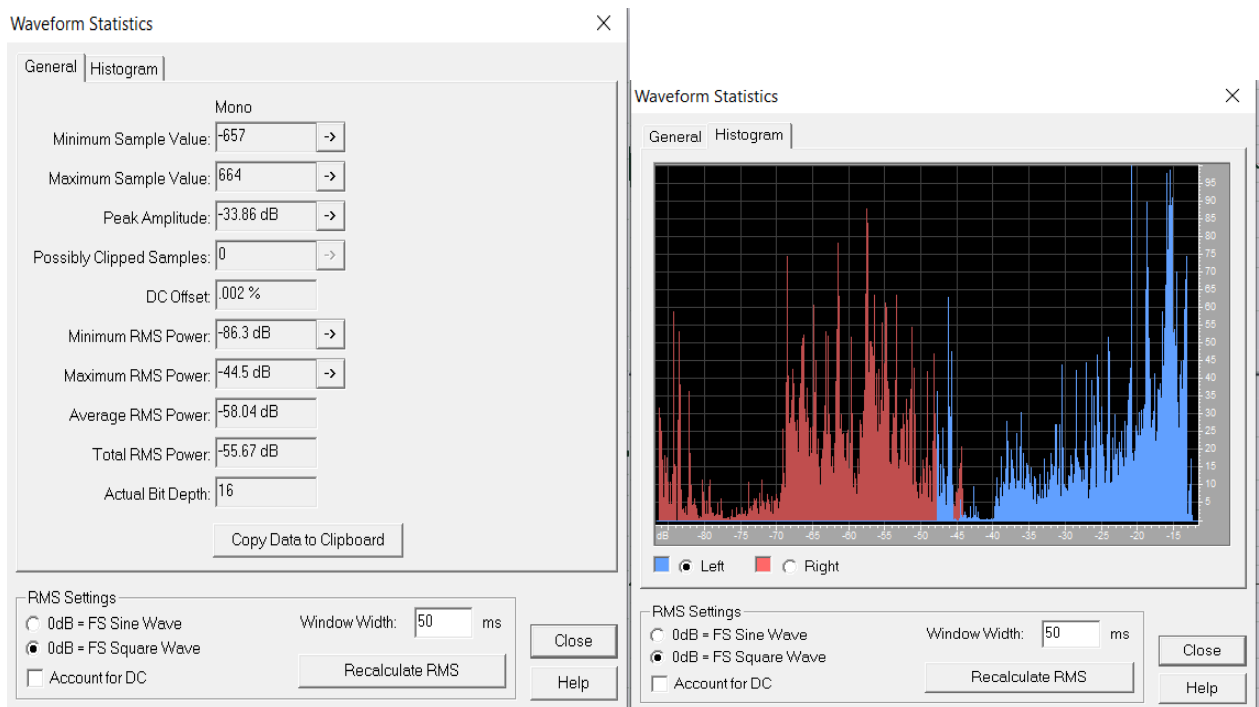


Рис.13. Статистические данные сигнала ошибки при использовании предложенного алгоритма передискретизации

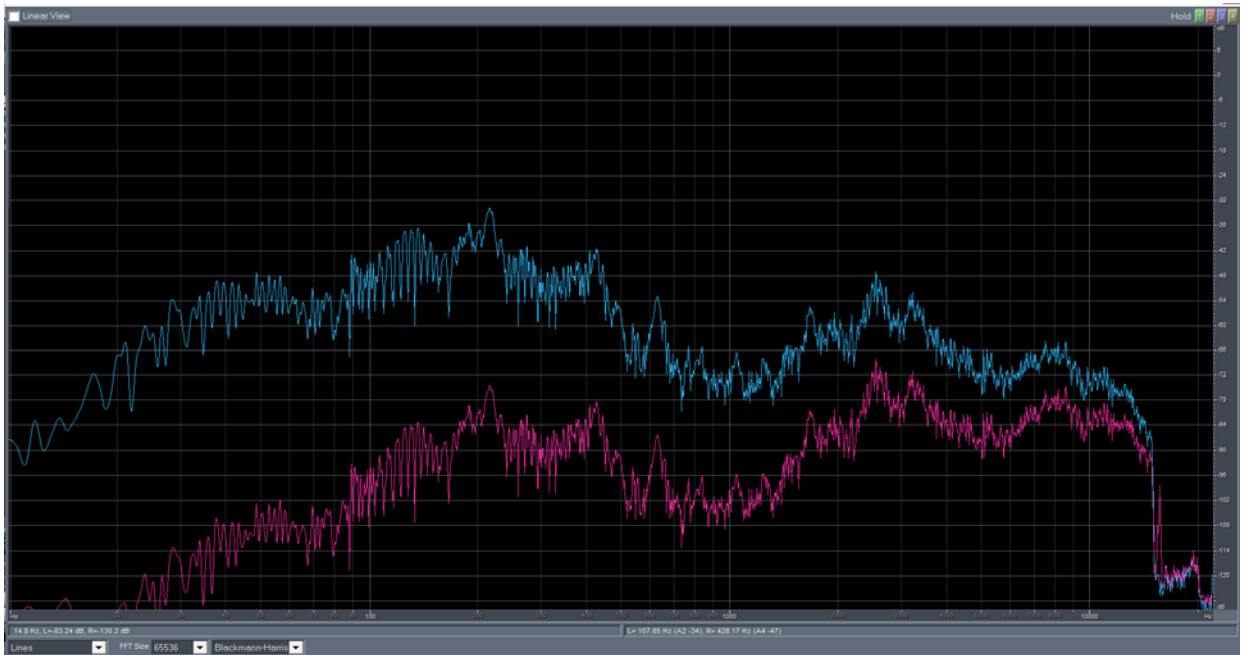


Рис.14. Спектры исходного сигнала и сигнала ошибки при использовании встроенной функции Matlab

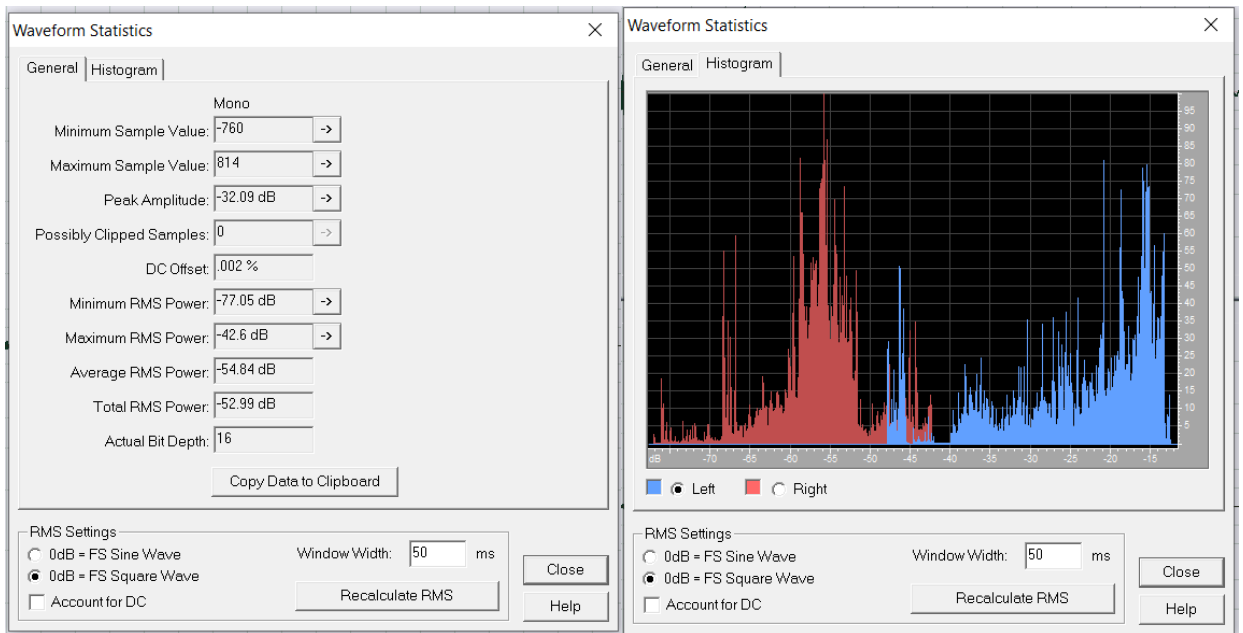


Рис.15. Статистические данные сигнала ошибки при использовании встроенной функции Matlab

Рассмотрим также показатели шумового сигнала после передискретизации с использованием встроенной функцией Cool Edit Pro (рис. 16, 17).

Ожидаемо, что, как и упоминалось ранее, наименьшей маскировкой обладает шумовой сигнал после передискретизации в программе Cool Edit Pro.

На рисунке 16 видно, что шумовой сигнал начинает превосходить по амплитуде полезный начиная с 4,5 кГц.

Из рисунка 17 следует, что гистограмма шумового сигнала (показано красным цветом) начинается уже с уровня – 27 дБ и связанные с этим искажения, скорее всего, будут слышны для человеческого уха.

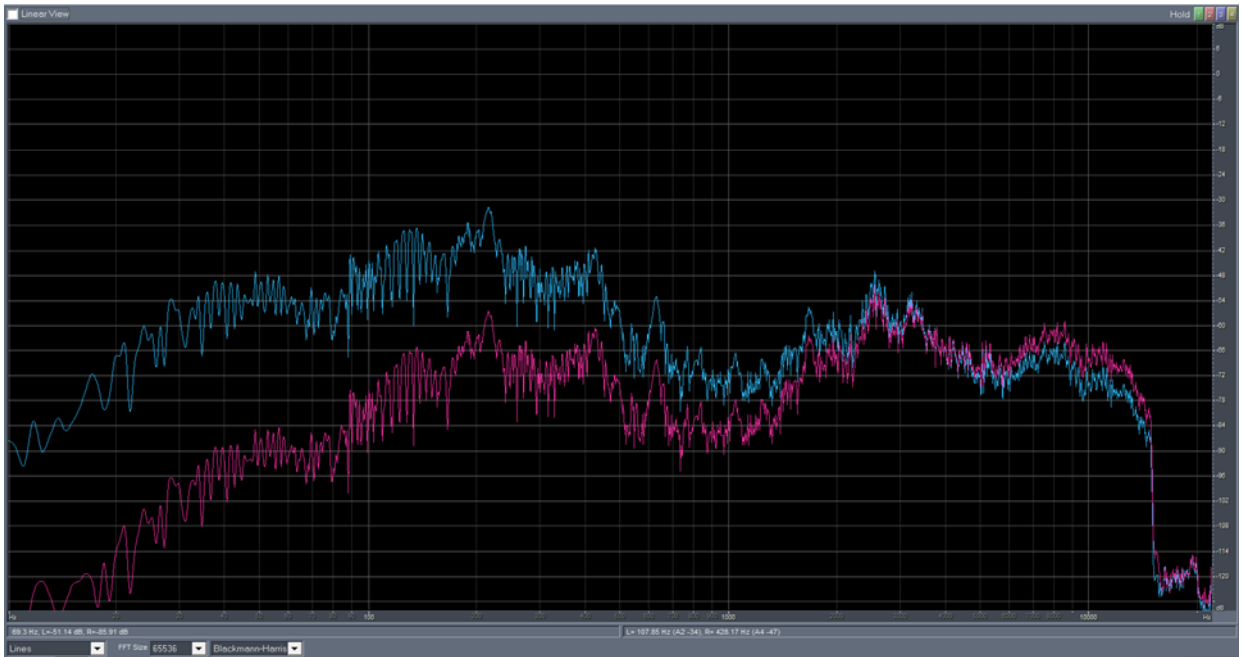


Рис. 16. Спектры исходного сигнала и сигнала ошибки при передискретизации в программе Cool Edit Pro

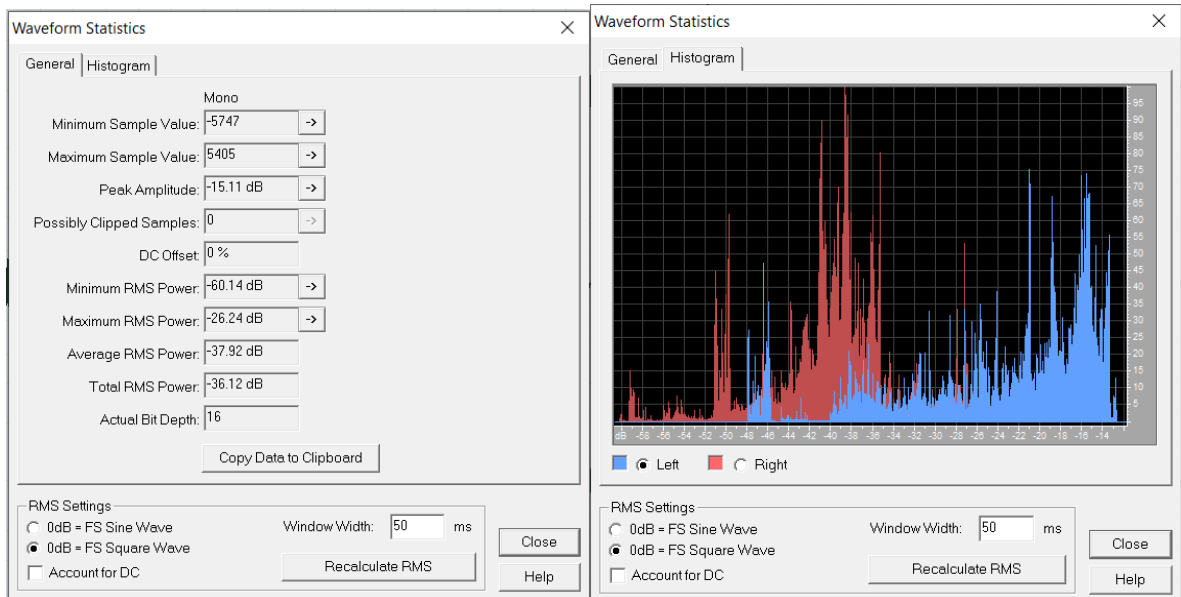


Рис. 17. Статистические данные сигнала ошибки при передискретизации в программе Cool Edit Pro

В таблице 2 представлены сравнительные данные RMS (Root Mean Square – среднеквадратическое значение, аналог ОСМ) и пиковой амплитуды шумовых сигналов при передискретизации разными способами.

Таблица 2

RMS и значение пиковой амплитуды шумовых сигналов при передискретизации разными способами

Параметры	Предложенный алгоритм	Функция Matlab	Встроенная функция Cool Edit Pro
RMS, дБ	-58,04	-54,84	-37,92
Пиковая амплитуда, дБ	-33,86	-32,09	-15,11

Из таблицы 2 следует, что наименьшее значение шума по обоим измеренным параметрам наблюдается при передискретизации с использованием предложенного алгоритма. Уменьшение составляет 3,2 дБ (~2 раза) в сравнении передискретизацией с использованием функции Matlab и 20 дБ – при использовании встроенной функции Cool Edit Pro.

ОСШ после использования предложенного алгоритма передискретизации составляет 37 дБ если считать по средним значениям RMS. С учётом того, что нет информации об ОСШ исходного сигнала, который при измерениях был загружен из интернет-архива радиопередач Радио России [5], в котором программы хранятся в сжатом аудио формате, можно утверждать, что ОСШ исходного сигнала меньше студийного 16 битного сигнала с ОСШ 96-98 дБ [6]. При этом можно предположить, что ОСШ исходного сигнала с учётом его хранения в сжатом формате, составляет приблизительно 50 дБ и это значение приблизительно соответствует динамическому диапазону исходного аудио файла. Соответственно, ОСШ ухудшилось приблизительно на 13 дБ.

Заключение

Основываясь на экспериментальных данных, которые были получены на кафедре ТиЗВ МТУСИ, в предложенном алгоритме передискретизации была выбрана оконная функция Наттолла и длина интервала наблюдения 240 мс, обеспечивающие наилучшее значение ОСШ. Экспериментально доказано, что на минимизацию искажений напрямую влияет увеличение длительности интервала наблюдения. Установлено, что остаточный шумовой сигнал при использовании предложенного алгоритма передискретизации имеет наименьшую мощность по сравнению с результатами, полученными с использованием встроенных функций Matlab и Cool Edit Pro. Также, выявлено, что шумовой сигнал по параметру среднего значения RMS составляет всего 0,02% от среднего RMS исходного сигнала, что полностью соответствует требованиям ГОСТ 52742-2007 [10], по условиям которого сигнал остатка преобразования должен составлять не больше 2% от исходного.

Литература

1. Greiner R.A., Eggers Jeff. The Spectral Amplitude Distribution of Selected Compact Discs. // J. Audio Eng.Soc., vol.37, No.4, 1989 April.
2. VOICEBOX: Speech Processing Toolbox for MATLAB Department of Electrical and Electronic Engineering of Imperial College London: URL: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html> (дата обращения: 10.05.2021).
3. Попов О.Б., Рихтер С.Г., Хрянин Е.А. Вопросы объективизации измерений параметров качества звуковых вещательных сигналов // Метрология и измерительная техника в связи, 2003. №2(32). С. 27-29.
4. Блейхут Р. Теория и практика кодов, контролирующих ошибки. Пер. с англ.: И.И. Грушко, В.М. Блиновский. Под редакцией: К.Ш. Зигангирова. М.: Мир, 1986. 576 с.
5. Влияние цифрового тракта на качество передачи звуковых сигналов. Маскировка // Хелпикс. Орг – Интернет помощник. URL: <https://helpiks.org/8-12472.html> (дата обращения: 06.05.2021).
6. Всё, что вы хотели знать о децибелах, но боялись спросить : обучающие материалы Rohde&Schwarz. 2018. 44 с.
7. Гилки, Роберт Х. Сравнение спектральной и временной информации при маскировании акустических сигналов. Auditory processing of complex sounds. London, 1987. С. 26-35, пер. С-66334.
8. Гольденберг Л.М., Матюшкин Б.Д., Поляк М.Н. Цифровая обработка сигналов. Справочник/ М.: Радио и связь, 1985. 312 с.
9. ГОСТ Р 50757-95. Сигналы передач звукового вещания государственных и независимых телерадиокомпаний, передаваемые на вход трактов первичного распределения. Основные параметры. Методы измерений.
10. ГОСТ Р 52742-2007. Каналы и тракты звукового вещания. Типовые структуры. Основные параметры качества. Методы измерений.
11. Дворкович В.П., Дворкович А.В. Оконные функции для гармонического анализа сигналов. М.: Техносфера, 2014. 112 с.
12. Попов О.Б., Рихтер С.Г. Цифровая обработка сигналов в трактах звукового вещания. Учебное пособие для вузов. - 2-е изд., стереотип. М.: Горячая Линия – Телеком, 2015. 342 с.
13. Попов О.Б., Рихтер С.Г. О возможных подходах к измерению качества передачи в адаптивных вещательных каналах // Метрология и измерительная техника в связи, 1998, №5. С. 24-27.
14. Орлов В.Г. Достоверность методов оценки качества звучания речи в системах связи. Материалы 8-й международной научно-технической конференции «Перспективные технологии в средствах передачи информации, 21-22 мая 2009. Владимир: Изд-во Владим. гос у-та, 2009. Том 2. С. 141-145.
15. Блейхут Р. Теория и практика кодов, контролирующих ошибки. Пер. с англ.: И.И. Грушко, В.М. Блиновский. Под редакцией: К.Ш. Зигангирова. М.: Мир, 1986. 576 с.
16. Орлов В.Г., Гитлиц М.В., Соловьёв В.Г., Шауро А.В. Влияние скорости цифрового преобразования адаптивной дельта-модуляции на информационные параметры кодируемых речевых сигналов // Радиотехника. 1987. № 3. С. 63-65.
17. Акимочев Е.А., Орлов В.Г. Производственно-вещательный комплекс радиостанции FM диапазона "Радио 1" // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2018. № 1. С. 3-6.
18. Тактакишвили В.Г., Попов О.Б., Абрамов В.А., Борисов А.А. Методы компактного представления, оценки и обработки звуковых сигналов на основе их комплексного представления // Т-Comm: Телекоммуникации и транспорт. 2019. Т. 13. № 2. С. 11-17.
19. Абрамов В.А., Попов О.Б., Терехов А.Н., Люкина Е.В. Контроль шумового загрязнения городской среды на основе использования ненормированных модуляционных параметров шума // Т-Comm: Телекоммуникации и транспорт. 2018. Т. 12. № 12. С. 54-60.

20. *Абрамов В.А., Малов А.В., Попов О.Б., Черников К.В.* Прогнозирование качества передачи сигнала вещания по короткому реальному звуковому сигналу // Т-Сотт: Телекоммуникации и транспорт. 2017. Т. 11. № 2. С. 19-24.
21. *Абрамов В.А., Попов О.Б., Борисов А.А., Черников К.В.* Статистика звукового сигнала, представленного комплексными модулирующими функциями // Т-Сотт: Телекоммуникации и транспорт. 2017. Т. 11. № 6. С. 29-32.
22. *Точный Ю.М., Попов О.Б.* Алгоритм изменения частоты дискретизации сигнала звукового вещания // Телекоммуникации и информационные технологии. 2017. Т. 4. № 2. С. 26-29.

EDUCTION OF DISTORTION OF THE AUDIO SIGNAL DURING ITS OVERSAMPLING

Oleg V. Gavrikov,
Graduate MTUCI, Moscow, Russia,
oleg.gavrikov13@gmail.com

Alexey D. Malgin,
Graduate MTUCI, Moscow, Russia,
malgin.alexey1999@gmail.com

Vladimir G. Orlov,
Chief specialist of the Department of OoRWoS, PhD., MTUCI, Moscow, Russia,
v.g.orlov@mtuci

Oleg B. Popov,
Professor of the Department of TiZV, PhD., MTUCI, Moscow, Russia,
o.b.popov@mtuci.ru

Evgeny V. Samarin,
Student MTUCI, Moscow, Russia,
7706053@gmail.com

Abstract

The article discusses the main ways of converting the sampling frequency of audio signals. These are methods: with a transition to the time domain, with a transition to the frequency domain, as well as a method with a transition to an analog representation of the signal. As a result of the analysis of these methods, it was concluded that the greatest efficiency of using such techniques is achieved when using test signals on which the channel parameters were evaluated, but when transmitting a real signal, significant distortions are introduced. A model of an algorithm for changing the sampling frequency in the frequency domain, which almost does not change the shape of the signal, is given.

Keywords: analog signal, oversampling, frequency, audio signal, conversion, distortion, digital signal.

АНАЛИЗ И СРАВНЕНИЕ ВАРИАНТОВ ПОСТРОЕНИЯ УСИЛИТЕЛЕЙ МОЩНОСТИ ДЛЯ МОБИЛЬНЫХ СЕТЕЙ 5G

*Донцов Дмитрий Евгеньевич,
Магистрант МТУСИ, Москва, Россия,
elysium_10@mail.ru*

*Горгадзе Светлана Феликсовна,
профессор кафедры РОС, д.т.н., МТУСИ, Москва, Россия,
s.f.gorgadze@mtuci.ru*

Аннотация

Рассматривается проблематика повышения энергетической эффективности передающих трактов радиоборудования систем 5G. Проводится описание и сравнительный анализ существующих методов построения высокоэффективных линейных усилителей мощности, применительно к построению радиоборудования систем 5G. Обсуждается целесообразность применения метода У. Догерти, как одного из возможных и реализуемых способов повышения КПД линейных усилителей мощности.

Ключевые слова: OFDM, пик-фактор, высокоэффективный линейный усилитель мощности, метод У. Догерти, метод Л. Кана, метод дефазирования.

Введение

В современных телекоммуникационных системах с целью повышения информационной эффективности радиоканалов используют все более сложные методы модуляции несущей частоты, а результирующий информационный сигнал характеризуется как амплитудной, так и фазовой модуляцией, и зачастую напоминает шум в некоторой ограниченной полосе частот с характерным значительным по величине пик-фактором. Кроме того, довольно часто один и тот же высокочастотный усилительный тракт используется для усиления нескольких сигналов при амплитудной модуляции результирующего суммарного сигнала. Это характерно как для базовых, так и для мобильных станций сотовых систем, но в существенно большей степени относится к базовым станциям. При усилении таких сигналов с целью уменьшения нелинейных искажений усилители должны работать с пониженной мощностью и КПД, в режиме *backoff mode* [1,2].

Для улучшения энергетических показателей усилителей сигналов с меняющейся амплитудной огибающей необходимо увеличивать время работы их анодный элемент (АЭ) в режимах с высоким КПД. С этой целью были предложены методы построения усилителей: схема Л.Р. Кана с удалением и восстановлением огибающей, автоматическое регулирование режима (АРР), методы Догерти и Ширекса, а также усилители с дельта-сигма модуляцией или широтно-импульсной модуляцией (ШИМ) на нелинейных компонентах [1]. В данной статье рассмотрим и сравним наиболее перспективные варианты построения усилителей мощности для перспективных мобильных сетей.

Метод Л. Кана

Метод Л.Р. Кана основан на возможности отдельного формирования и усиления низкочастотной амплитудной огибающей $U_m(t)$ и высокочастотной фазомодулированной $\cos(\omega_o t + \varphi(t))$ составляющей сигнала вида $S(t) = U_m(t)\cos(\omega_o t + \varphi(t))$. При этом амплитудная огибающая $U_m(t)$ сначала удаляется из сигнала, а затем восстанавливается, поэтому в англоязычной литературе этот метод получил название Envelope Elimination and Restoration (EER).

Соответствующий усилитель является двухканальным (рис. 1).

Его высокочастотный входной сигнал делится с помощью делителя мощности, после чего два одинаковых сигнала подаются на два разных канала схемы. В верхнем канале сигнал детектируется, в результате чего получается модулирующий сигнал в основной полосе частот. Он подается на мощный ключевой усилитель, на выходе которого устанавливается ФНЧ для восстановления огибающей. В другом канале сигнал проходит через ограничитель уровня, устраняющего амплитудную модуляцию. Полученный высокочастотный сигнал содержит информацию лишь о фазе усиленного сигнала. Он задерживается на некоторое время, и затем поступает либо на усилитель радиочастоты, работающий в высокоэффективном режиме (классы E, F), либо на предварительный усилитель с существенно более низким коэффициентом усиления.

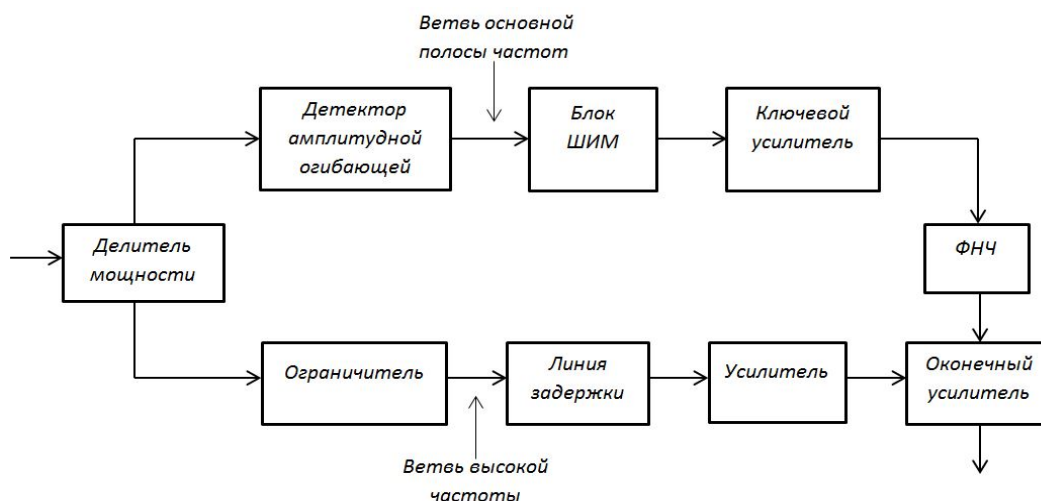


Рис. 1. Блок-схема усилителя Л. Кана

В первом случае мощный сигнал огибающей выводится в качестве модулирующего сигнала на выходной электрод АЭ окончного усилителя с относительно низким коэффициентом усиления с целью амплитудной модуляции высокочастотной составляющей, подаваемой на входной электрод. Оконечный усилитель используется в основном для амплитудной модуляции, и в нем на высоком уровне мощности синтезируется усиливаемый сигнал (рис. 2).

Во втором случае сигнал амплитудной огибающей подается на выходные электроды АЭ окончного усилителя, состоящего из двух усилительных каскадов – предвыходного и выходного. Одновременно на вход предвыходного каскада подается высокочастотный сигнал с выхода предварительного усилителя. Предвыходной каскад используется в основном для амплитудной модуляции, а высокоэффективное усиление сигнала производится в выходном каскаде, в результате чего реализуется комбинированная модуляция и одновременное усиление сигнала. Таким образом, работа всех каскадов тракта радиочастоты в высокоэффективном режиме обеспечивает возможность предельного использования мощности АЭ и высокий КПД. Но качественные и количественные показатели усилителя в существенной степени определяются инерционностью канала усиления огибающей, групповая задержка в котором может составлять до 10-20 микросекунд из-за необходимости использования ФНЧ [4]. В то же время задержка сигнала в широкополосном тракте усиления радиочастотной составляющей составляет обычно несколько наносекунд, и его дополнительная задержка приблизительно на 10 микросекунд – сложная техническая задача.

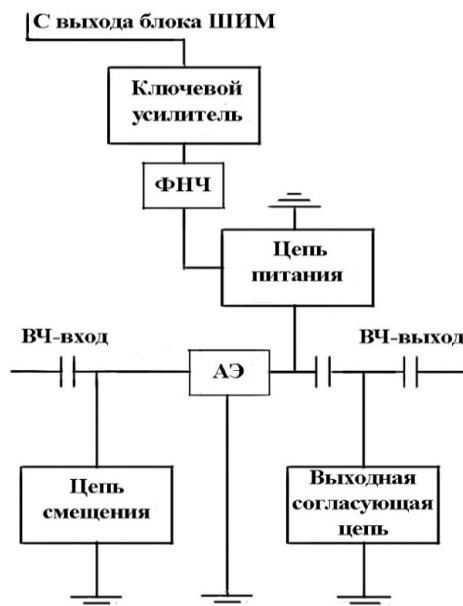


Рис. 2. Оконечный усилительный каскад схемы Л. Кана с мощным модулирующим сигналом, подаваемым на выходной электрод АЭ

Так, в [3,5] рассматривается линия задержки на каскаде операционных усилителей, на входе которой частота сигнала понижается с помощью опорного сигнала гетеродина и полосового фильтра, используемого с целью выделения необходимых частотных составляющих. Затем частота вновь преобразуется до частоты исходного сигнала. Возможно применение и вариантов высокочастотных линий задержки, например на поверхностных акустических волнах. В этом

случае ширина полосы частот задерживаемого широкополосного сигнала ограничивается амплитудно-частотной характеристикой линии, которая определяется параметрами входных и выходных встречно-штыревых преобразователей и схем согласования. Таким образом, на практике возможна реализация лишь узкополосной линии задержки радиочастотной составляющей усиливаемого сигнала, что приводит к появлению его паразитной модуляции, как фазовой, так и амплитудной.

Определенной проблемой является полоса пропускания НЧ тракта, которая должна быть минимум в 5 раз больше, чем ширина канала радиосигнала. При этом тактовая частота ШИМ-регулятора должна быть в 7 раз больше полосы пропускания НЧ тракта, т.е. тактовая частота ШИМ-регулятора должна быть минимум в 35 раз больше, чем ширина канала усиливаемого сигнала.

Амплитудная модуляция в оконечном усилителе

Считаем, что оконечный усилитель схемы Л. Кана состоит из одного резонансного усилительного каскада как это показано на рисунке 1. Амплитудная модуляция осуществляется введением низкочастотного модулирующего напряжения $U_{мод}(t) = U_m(t) - \langle U_m(t) \rangle$ последовательно с его постоянным напряжением питания $E_{п0}$, которое примем равным $\langle U_m(t) \rangle$. Таким образом, суммарное напряжение, подаваемое на выходной электрод АЭ:

$$(E_{п} = E_{п0} + U_{мод}(t) = E_{п0} (1 + (U_{мод}(t))/E_{п0})) \quad (1)$$

Модуляция реализуется в основном в перенапряженном режиме, и лишь при максимальном значении $E_{п}$ режим граничный. Импульс тока, определяемый входным высокочастотным напряжением с постоянной амплитудой $U_{вх}$ и фиксированным напряжением смещения $E_{см}$, имеет постоянный угол отсечки:

$$\cos \theta = \frac{(E_{отс} - E_{см})}{U_{вх}}, \quad (2)$$

где $E_{отс}$ – напряжение отсечки выходного тока.

Остаточное напряжение на выходном электроде $e_{вых.гр} = E_{пmax} - U_{выхmax}$. По мере уменьшения $E_{п}$ в импульсе тока $i_{вых}(\omega t)$ появляется провал, что иллюстрируется на рисунке 3. Этот рисунок может использоваться в качестве иллюстрации расчета модуляционных характеристик, выполненного С.И. Евтяновым [6].

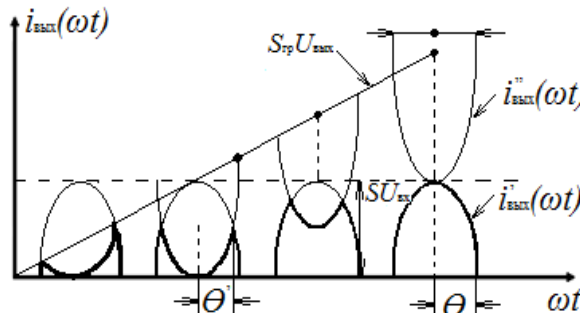


Рис. 3. Иллюстрация изменения формы импульса выходного тока АЭ при изменении напряжения питания

Очевидно, что амплитуда 1-ой гармоники $I_{вых1}$ выходного тока является функцией лишь $E_{п}$ и $U_{вых}$, поскольку $U_{вх}$, $E_{см}$, θ и сопротивление нагрузки $R_{н}$ резонансного усилителя не изменяются, то есть являются постоянными. Тогда в соответствии с рисунком 3:

$$I_{вых1}(E_{п}, U_{вых}) = S U_{вх} \gamma_1(\theta) - (S U_{вх} + S_{гр} U_{вых}) \gamma_1(\theta'), \quad (3)$$

где: S – крутизна проходной статической характеристики АЭ,

$S_{гр}$ – крутизна линии граничного режима,

$U_{вых}$ – амплитуда выходного напряжения,

$\gamma_1(\cdot)$ – коэффициент Берга,

θ' – «верхний» угол отсечки импульсов выходного тока.

Кроме того, из анализа рисунка 3 следует:

$$\cos \theta' = \frac{S_{гр} E_{п} - S(E_{см} - E_{отс})}{S U_{вх} + S_{гр} U_{вых}} \quad (4)$$

Вместе с тем:

$$I_{\text{ВЫХ1}}(U_{\text{ВЫХ}}) = \frac{U_{\text{ВЫХ}}}{R_{\text{н}}} \quad (5)$$

Затем, с целью исследования функции $I_{\text{ВЫХ1}}(E_{\text{п}})$ использовался графический подход, и в соответствии с (3) строилось семейство кривых $I_{\text{ВЫХ1}}(U_{\text{ВЫХ}})$, где в качестве параметра выступали выбираемые значения $E_{\text{п}}$, [5]. В той же системе координат строилась прямая (5). Точки пересечения графиков, построенных по (4) при различных значениях $E_{\text{п}}$, и прямой (5) определяют модуляционную характеристику $I_{\text{ВЫХ1}}(E_{\text{п}})$ усилителя. Аналогично можно найти зависимость постоянной составляющей импульсов тока на выходе АЭ от напряжения питания $I_{\text{ВЫХ0}}(E_{\text{п}})$, заменив в (3) $\gamma_1(\theta)$ и $\gamma_1(\theta')$ на $\gamma_0(\theta)$ и $\gamma_0(\theta')$ соответственно. Результат построения функций $I_{\text{ВЫХ1}}(E_{\text{п}})$ и $I_{\text{ВЫХ0}}(E_{\text{п}})$ показан на рисунке 4; предполагалось, что максимальное значение $E_{\text{п}}$ (в данном случае 25 В) соответствует граничному режиму, $E_{\text{п0}} = E_{\text{п}}/2$, импульс тока $i_{\text{ВЫХ}}(\omega t)$ $i_{\text{ВЫХ}}(\omega t)$ начинает «уплощаться» на уровне $S_{\text{гр}}e_{\text{ВЫХ.гр}}$ при $E_{\text{п}} = e_{\text{ВЫХ.гр}}$, этот импульс имеет форму трапеции с $\theta \approx \theta'$ при $U_{\text{ВЫХ}} = 0$ и $E_{\text{п}} = 0,5e_{\text{ВЫХ.гр}}$, а также $\theta = 90^\circ$.

Из анализа рисунка 4 следует, что функция $I_{\text{ВЫХ1}}(E_{\text{п}})$ близка к линейной, а $I_{\text{ВЫХ0}}(E_{\text{п}})$ существенно нелинейна.

На основании рисунка 4 построены характеристики $P_1(E_{\text{п}})$, $P_0(E_{\text{п}})$, $P_{\text{рас}}(E_{\text{п}})$ и $\eta_1(E_{\text{п}})$ в соответствии с выражениями:

$$P_1 = I_{\text{ВЫХ1}}^2 R_{\text{н}}; P_0 = I_{\text{ВЫХ0}} E_{\text{п}}; P_{\text{рас}} = P_0 - P_1; \eta_1 = P_1/P_0. \quad (6)$$

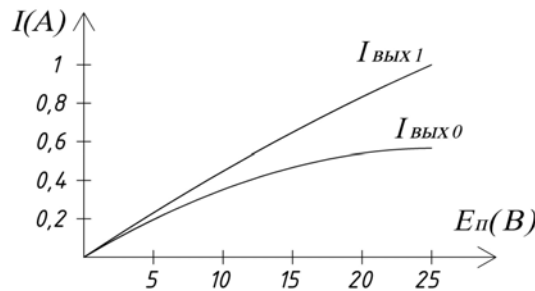


Рис. 4. Зависимости первой гармоники и постоянной составляющей выходного тока от напряжения питания

Отметим, что если бы функции $I_{\text{ВЫХ1}}(E_{\text{п}})$ и $I_{\text{ВЫХ0}}(E_{\text{п}})$ были линейными, то $P_1(E_{\text{п}})$, $P_0(E_{\text{п}})$, $P_{\text{рас}}(E_{\text{п}})$ имели бы вид квадратичных парабол, но в данном случае они приблизительно соответствуют параболам степени 3/2. При этом КПД довольно высокий во всем диапазоне изменения $E_{\text{п}}$ из-за использования перенапряженного режима.

Исследуем зависимость коэффициента усиления по мощности K_{P} от $E_{\text{п}}$. С этой целью будем полагать, что $E_{\text{п}}$ не влияет на первую гармонику тока на входе АЭ, а также функция $I_{\text{ВЫХ1}}(E_{\text{п}})$ линейна. Тогда:

$$K_{\text{P}} = K_{\text{Pmax}} \left(\frac{E_{\text{п}}}{E_{\text{пmax}}} \right)^2 \quad (7)$$

Таким образом, достоинством рассматриваемого вида модуляции является высокий КПД, сочетающийся с близкой к линейной модуляционной характеристикой $I_{\text{ВЫХ1}}(E_{\text{п}})$. Вместе с тем $I_{\text{ВЫХ0}}(E_{\text{п}})$ существенно нелинейна. Кроме того, как следует из (7), коэффициент усиления по мощности пропорционален квадрату напряжения питания, и его среднее значение мало. Входная проводимость АЭ тоже может существенно зависеть от $E_{\text{п}}$, что приведет к паразитной амплитудной и фазовой модуляции входного, а, следовательно, и выходного сигнала усилителя. Для ослабления этого эффекта можно уменьшить связь между усилителем высокочастотной составляющей и оконечным усилительным модулирующим каскадом, что очевидно приведет к еще большему уменьшению коэффициента усиления. Последнее обстоятельство не является критичным для данного варианта построения усилителя Л. Кана, поскольку усиление уже выполнено, и основной функцией оконечного усилителя является амплитудная модуляция высокочастотной составляющей, по возможности с минимальными искажениями.

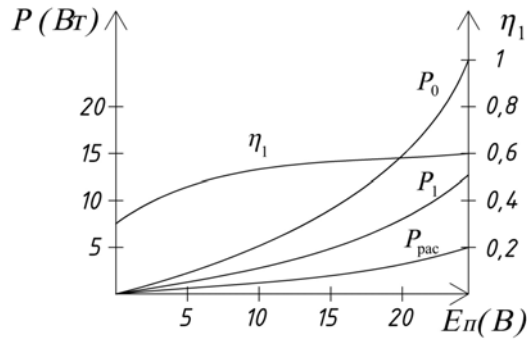


Рис. 5. Характеристики мощности и КПД оконечного усилителя

КПД усилителя Л. Кана, рассмотренного выше, определяется характеристиками основных усилительных каскадов – усилителя низкочастотной огибающей, усилителя высокочастотной составляющей и оконечного модулирующего усилительного каскада. Действительно, полагая, что полезная мощность сигнала на выходе оконечного модуляционного каскада составляет $P_{1МВЧ}$, а коэффициент усиления по мощности мал, определим его КПД как:

$$\eta_{1МВЧ} = \frac{P_{1МВЧ}}{(P_{0ог} + P_{1ВЧ})}, \tag{8}$$

где $P_{0ог}$ – напряжение питания, $P_{1ВЧ}$ – мощность входного сигнала.

Но $P_{0ог}$ вырабатывается в тракте усиления огибающей, то есть:

$$P_{0ог} = \eta_{1ог} P_0, \tag{9}$$

где $\eta_{1ог}$ – КПД усилителя огибающей, P_0 – напряжение питания.

Аналогично определим, что:

$$P_{1ВЧ} = \eta_{1ВЧ} P_0, \tag{10}$$

где $\eta_{1ВЧ}$ – КПД усилителя высокочастотной составляющей.

Тогда:

$$\eta_{1МВЧ} = \frac{P_{1МВЧ}}{\eta_{1ог} P_0 + \eta_{1ВЧ} P_0} = \frac{P_{1МВЧ}}{P_0} \frac{1}{\eta_{1ог} + \eta_{1ВЧ}} \tag{11}$$

С учетом последнего выражения, полагая, что усилители огибающей и высокочастотной составляющей потребляют одинаковую мощность P_0 , получим для КПД рассматриваемого варианта усилителя Л. Кана:

$$\eta_{1К} = \frac{P_{1МВЧ}}{2P_0} = \eta_{1МВЧ} \frac{\eta_{1ог} + \eta_{1ВЧ}}{2}, \tag{12}$$

Из последнего выражения следует, что если составляющие сигнала с амплитудной модуляцией усиливаются высокоэффективными ключевыми усилителями с КПД порядка 95%, а средний КПД модулирующего каскада составляет приблизительно 56% (рисунок 5), то КПД данного усилителя составит приблизительно 50-55%. Интерес представляет рассмотрение комбинированной амплитудной модуляции в оконечном и предоконечном каскадах. В работе [3] показано, что в этом случае КПД схемы Кана теоретически может достигать приблизительно 70%.

Отметим, что для данного варианта усилителя Л. Кана в частотном диапазоне приблизительно до 900...1200 МГц интерес также представляет рассмотрение усилителя класса Е в оконечном модулирующем каскаде, поскольку в ключевых режимах амплитуда выходного сигнала зависит от значения напряжения питания. В результате этого можно усиливать сигналы с амплитудной модуляцией в ключевом режиме, если менять напряжение питания в соответствии с амплитудой усиливаемого сигнала. Предполагаемое повышение КПД всей схемы может составить несколько десятков процентов по сравнению с резонансным усилителем, рассмотренным выше.

Метод У. Догерти

Данный метод построения УМ был предложен в 30-х годах прошлого века американским инженером-электриком Уильямом Догерти [7]. Способ построения первоначально применялся в мощных (до нескольких сотен киловатт) ламповых ВЧ передатчиках с амплитудной модуляцией. Однако данный метод не утратил актуальности, и в настоящее

время используется в построении передающих устройств базовых станций мобильной связи третьего и четвертого поколений, а также в цифровых стандартах телевидения для сигналов, характеризующихся высоким пик-фактором.

Идея усилителя Догерти, структурная схема которого представлена на рисунке 6, состоит в следующем: устройство состоит из двух усилителей – основного, который работает во всем диапазоне входного напряжения, и пикового, диапазон входного напряжения которого составляет от половины до максимального значения. Оба работают на общую нагрузку.

Работа усилителя Догерти при $U_{вх}=0$ до $U_{вх}=0,5U_{вхmax}$ происходит следующим образом: при данном режиме пиковый усилитель закрыт и не потребляет энергию, а всю полезную мощность выдает основной или т.н. усилитель несущей частоты (УНЧ). УП (пиковый усилитель) работает в недонапряженном режиме класса АВ, но напряжение смещения выбирается близкому к классу В. Это одновременно позволяет приблизиться по энергоэффективности к классу В (КПД которого больше, чем при работе в классе А), и к линейности класса А. При относительно небольших значениях входного напряжения его форма полностью повторяется на выходе усилительного прибора [8]. При таком режиме работы УНЧ будет работать как источник тока. Сигнал поступает на нагрузку, а побочные составляющие фильтруются в выходной фильтрующей системе (ВФС).

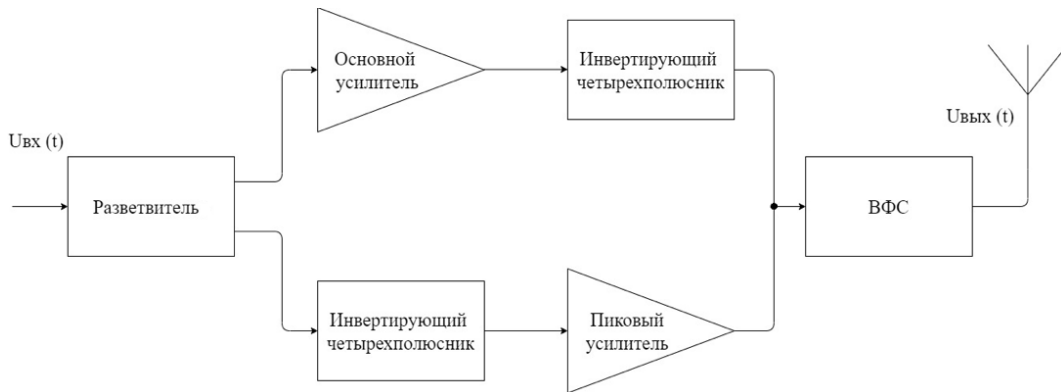


Рис. 5. Структурная схема усилителя Догерти

При увеличении входного напряжения будет увеличиваться выходной ток транзистора, при этом будет увеличиваться полное нагрузочное сопротивление, подключенное через инвертирующий четырехполосник [7]. При половинном значении выходного тока УП выходной импеданс будет максимальным, вместе с тем падение напряжения на нем будет так же максимальным. Так как пиковый усилитель закрыт, то КПД усилителя Догерти (рисунок 7) будет равным КПД основного усилителя, при этом он будет развивать 25% максимальной мощности передающего устройства. КПД при таком значении будет максимальным, транзистор будет работать в граничном режиме [1].

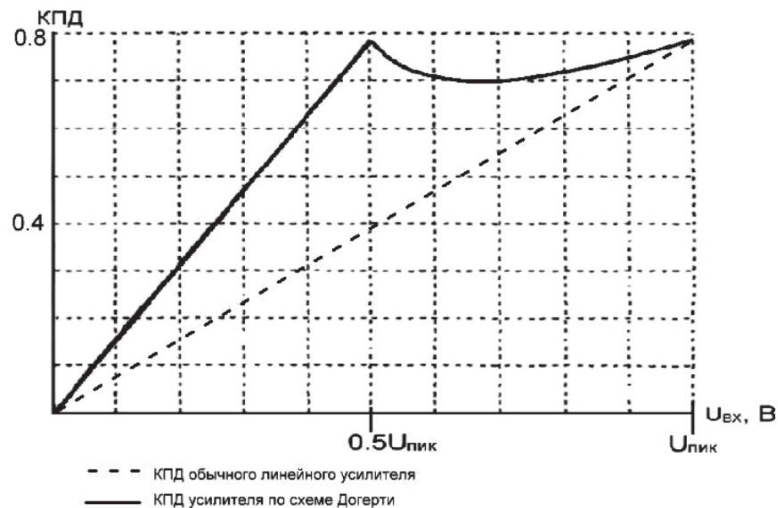


Рис. 6. Зависимости КПД от напряжения возбуждения для линейного усилителя и по схеме Догерти [9]

При напряжении возбуждения $U_{вх}=0,5U_{вхmax}$ до $U_{вхmax}$ работа усилителя Догерти состоит в следующем: при увеличении входного напряжения открывается пиковый усилитель, который работает в нелинейном режиме класса С, КПД которого превосходит усилитель класса В. При таком классе работы величина выходного тока возрастает быстрее уровня напряжения возбуждения. Одновременно с этим выходной импеданс основного усилителя падает, но падение напряжения на нем остается неизменным, что не дает перейти УП в перенапряженный режим, для которого характерен высокий уровень искажений. В таком случае основной усилитель работает уже как источник напряжения, в то же

время пиковый усилитель работает как источник тока. При достижении максимального входного напряжения оба усилителя работают в граничном режиме, причем развивают одинаковую мощность [7].

Таким образом, определяющим фактором в высоком КПД усилителя Догерти является динамическое изменение полного нагрузочного сопротивления, зависящее от напряжения возбуждения. На рисунке 7 виден провал на графике КПД описываемого усилителя, который возникает вследствие открывания пикового усилителя, при этом общий КПД остается большим, нежели у обычного линейного усилителя.

Несомненно, главным преимуществом усилителя по схеме Догерти относительно усилителей классов В и АВ (самых распространенных при проектировании передающих устройств) является более высокий КПД одновременно с приемлемой линейностью на всем участке рабочего напряжения. Также стоит отметить выигрыш в два раза в энергоэффективности усилителя Догерти при наиболее вероятных значениях сигнала в диапазоне $0,1 \dots 0,3$ от максимального входного напряжения.

Помимо этого, можно выделить и другие достоинства метода, а именно: возможность использования как в мало-мощных, так и в мощных передающих устройствах; саморегулируемость и простота построения; меньшее число побочных спектральных составляющих относительно обычных линейных усилителей класса АВ.

Однако из-за использования инвертирующих четырехполосников в схеме вытекает главный недостаток – узкополосность. В качестве четырехполосников обычно применяется четвертьволновая линия передачи, эквивалентной схемой которой является П-контур в большинстве случаев фильтра нижних частот [1]. Ее фазовый сдвиг зависит от частоты сигнала, вследствие чего передатчик работает в узком диапазоне с невозможностью перестройки на другую полосу частот [1].

Помимо этого, из существенных недостатков можно выделить искажения из-за излома амплитудной характеристики при включении пикового транзистора. Решением данной проблемы является применение адаптивной предкоррекции и других способов линеаризации. Другими недостатками построения усилителя по схеме Догерти являются: более низкий коэффициент усиления (около 3дБ) по сравнению с усилителями класса АВ; сложное построение; трудоёмкость настройки и нахождения оптимальной рабочей точки усилительных приборов [9].

Заключение

В настоящее время в системах мобильной связи при построении высокоэффективных усилителей сигналов со сложными комбинированными методами модуляции наиболее часто используется метод Догерти. Развиваются цифровые усилители с $\Sigma\Delta$ -модуляцией на нелинейных элементах. Это объясняется тем, что они более просты в исполнении, в сравнении с усилителями на основе метод Л. Кана и Ширекса. Метод АРР, хотя и позволяет в отдельных случаях более простыми способами получить более высокий КПД, но часто приводит к появлению паразитной фазовой модуляции усиливаемого сигнала (амплитудно-фазовой конверсии), компенсировать которую гораздо сложнее по сравнению с амплитудными искажениями.

Основным источником нелинейных искажений сигнала в усилителе Догерти является излом его амплитудной характеристики, соответствующий моменту «включения» вспомогательного АЭ. Это объясняется тем, что статическая характеристика последнего не является идеальной, и имеет естественную кривизну. Для борьбы с этим недостатком часто используются цифровые методы коррекции АМ/АМ и АМ/РМ характеристик усилителя, поскольку традиционные способы линеаризации, связанные с использованием вариантов отрицательной обратной связи по радиочастотному сигналу как правило не могут использоваться на СВЧ, так как скорость изменения фазы высокой частоты в этом случае довольно высока, что делает весьма сложной задачу сохранения исходного соотношения фаз.

Литература

1. Горгадзе С.Ф. СВЧ усилители мощности для мобильной связи и радиодоступа. М.: Горячая линия – Телеком, 2022. 342 с.
2. Литвяков В.С, Орлов В.Г. Эволюция и тенденции совершенствования радиоинтерфейсов в сетях Geran/Utran/E/Utran // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. Ростов-на-Дону: ПЦ «Университет» СКФ МТУСИ, 2020. С. 86-92.
3. Kahn, L.R. Single-Sideband Transmission by Envelope Elimination and Restoration. IRE Proc. Vol.40, July 1952, pp. 803-806.
4. Kenington, P.B. High Linearity RF Amplifier Design, Norwood. MA: Artech House, 2000.
5. Raab, F.H. Intermodulation Distortion in Kahn-Technique Transmitters // IEEE Trans. On Microwave Theory and Techniques, Vol.44, December 1996, pp. 2273-2278.
6. Евтянов С.И. Ламповые генераторы. М.: Связь, 1967.
7. Doherty, W.H. A New High Efficiency Power Amplifier for Modulated Waves. IRE Proc. Vol.24, September 1936, pp. 1163-1182.
8. Кочемасов В.Н., Косичкина Т.П. Усилители мощности по схеме Догерти // Электроника НТБ. 2019. №3. С. 144-152.
9. Разин К. О., Иванюшкин Р.Ю. Проблематика построения усилителей мощности передатчиков эфирного цифрового телевидения на основе метода У. Догерти с введением автоматической регулировки режима по питанию // Материалы Международной научно-технической конференции. М.: INTERMATIC, 2017. С. 1131-1134.

ANALYSIS AND COMPARISON OF POWER AMPLIFIER OPTIONS FOR 5G SIGNALS

Dmitry E. Dontsov,
Graduate MTUCI, Moscow, Russia,
elysium_10@mail.ru

Svetlana F. Gorgadze,
Professor of the Department of RCQ, Doctor of Technical Sciences, MTUCI, Moscow, Russia,
s.f.gorgadze@mtuci.ru

Abstract

The problem of increasing the energy efficiency of radio equipment 5G transmitters is considered. A description and comparison of some existing methods for the construction of high-efficiency linear power amplifiers for the construction of 5G radio equipment is provided. The feasibility of the W. Doherty method is discussed as a possible and feasible way to increase the efficiency of linear power amplifiers.

Keywords: *OFDM, peak-to-average power ratio, high-efficiency linear power amplifier, Doherty amplifier, Polar Modulated power amplifier, LINC power amplifier*

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ОБНАРУЖЕНИЯ СИГНАЛОВ

*Бугаева Герел Мерезновна,
магистрант МТУСИ, Москва, Россия,
gerel.bugaeva@mail.ru*

*Набиев Владислав Ринатович,
магистрант МТУСИ, Москва, Россия,
nabiev_vlad@mail.ru*

*Репинский Владимир Николаевич,
доцент кафедры ИСУиА, к.т.н., МТУСИ., Москва, Россия,
vnrepinski@gmail.com*

Аннотация

Некоторые аспекты, рассматриваемые в этом исследовании, включают методы обнаружения с постоянной вероятностью ложных тревог (ПВЛТ), реализованные традиционным способом, и их отличие от реализации искусственных нейронных сетей (ИНС). Затем представлено краткое описание помех и целей, смоделированных с помощью распределений Вейбулла и Рэля. Особое внимание уделяется ИНС как детектору. Для оценки производительности представлено сравнение его с CFAR детектором.

Ключевые слова: сигнал, статистические, нейросетевые обнаружители, имитационные модели.

Введение

Импульсы радара воздействуют на различные объекты, а ответ принимается специально ориентированной антенной. Время, за которое возврат достигает антенны, пропорционально расстоянию (дальности) до обнаруженного объекта. Отраженные сигналы усиливаются и фильтруются приемником. После этого аналого-цифровой преобразователь преобразует аналоговый сигнал в дискретные временные отсчеты. Затем цифровой процессор сигналов выполняет обнаружение целей и контроль ложных тревог. Обнаружение радиолокационного сигнала в обычных радиолокационных системах основано на статистических методах. Эти методы требуют значительных вычислительных ресурсов и часто оптимальны только для одного типа распределения помех. Более поздние подходы к обнаружению радиолокационных сигналов также рассматривают искусственные нейронные сети (ИНС) в качестве нелинейных детекторов. Различные архитектуры ИНС были обучены для решения задачи обнаружения радиолокационных сигналов. Многоуровневая архитектура показывает наилучшие характеристики, особенно в условиях смешанных помех, по сравнению с алгоритмом постоянной вероятности ложных тревог (CFAR). Одной из ключевых проблем при разработке детектора является выделение целей из шума и помех [12]. На выходе детектора должно быть принято решение, является ли отражение на входе радара целью или нет. Классический метод принятия такого решения – алгоритм постоянных вероятностей ложных тревог.

Постоянная вероятность ложных тревог (Constant false alarm rate, CFAR) представляет собой адаптивный алгоритм, применяемый в сфере радиолокации в задачах обнаружения цели при условии присутствия существенного шума. Алгоритм CFAR предназначен для оценки минимального порогового значения мощности, при превышении которого принятый сигнал следует считать относящимся к цели. При чрезмерно низком значении порога будет увеличиваться число ошибок ложного срабатывания, при высоком, напротив, количество ошибок ложного срабатывания будет сведено к минимуму, однако возрастет число ошибок пропуска.

В большинстве задач обнаружения полезного радиолокационного сигнала при проектировании РЛС величина порогового значения устанавливается исходя из требований к вероятности возникновения ошибок ложных тревог [1]. В зависимости от значения порога, система может правильно обнаружить одну или несколько целей, пропустить обнаружение или ошибочно обнаружить цель из-за шума (ложные тревоги). Часто используемым алгоритмом, выполняющим такую операцию, является процессор с постоянной вероятностью ложных тревог, обычно реализуемый с помощью цифровой фильтрации. Алгоритм CFAR оценивает энергию помех в целевой тестовой ячейке и регулирует порог обнаружения в соответствии с изменением энергии помех в различных положениях тестовой ячейки. Затем амплитуда центральной ячейки (CUT) в этом регистре сравнивается с порогом. Порог обнаружения вычисляется таким образом, чтобы приемник поддерживал постоянную заранее заданную вероятность ложной тревоги P_{fa} [2, 4].

На рисунке 1 изображен детектор CFAR, который можно описать как сдвиговой регистр с подаваемым на него входным сигналом. Амплитуда в CUT сравнивается с уровнями колебаний в процессоре CFAR, масштабированным скалярным коэффициентом, определяемым как отношение порогового уровня к среднему значению амплитуд принятого колебания.

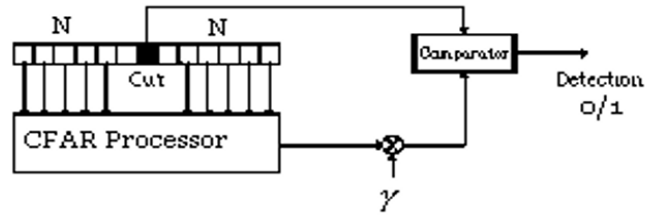


Рис. 1. Детектор с постоянной вероятностью ложных тревог

Производительность алгоритма ухудшается, когда фон помех не входит в выборку амплитуд оценки. В подавляющем большинстве реальных ситуаций уровень шума не является постоянной величиной а является функцией пространственных и, в ряде случаев, временных координат. В этом случае пороговое значение должно изменяться по ходу выполнения алгоритма поиска цели для поддержания постоянной ошибки ложных тревог. На самом деле меняется не только распределение помех, но и уровень помех (масштаб). В таком случае алгоритм, основанный только на усреднении энергии помех, демонстрирует плохую производительность. Кроме того, статистические подходы требуют достаточных знаний о данных для построения моделей, и часто это очень сложно [7]. Этим объясняется разработка нескольких нелинейных методов для улучшения процесса обнаружения целей.

В последние годы исследователи также рассматривают возможность принятия решения об обнаружении сигналов радарными с использованием искусственных нейронных сетей (ИНС) [7]. Фактически, проблему такого обнаружения можно рассматривать как задачу распознавания образов, что очень хорошо соответствует возможностям, которые предлагает ИНС [5, 8, 9]. Эти подходы дают лучшие и более надежные результаты, чем традиционные статистические подходы.

В [8] была проанализирована многослойная сеть прямого распространения с различными отношениями сигнал/помеха, и результаты были сравнены с показателями традиционных обнаружителей. В качестве входных параметров использовались число ячеек в эталонном окне и входные статистические характеристики, такие как: статистическое среднее значение по эталонному окну, средняя мощность и дисперсия. Показано, что детекторы ИНС постоянно обеспечивают повышенную производительность по сравнению с обычными детекторами, когда размер эталонного окна мал. Кроме того, получены аналогичные надежные оценки производительности в других условиях, например, наличие мешающих целей, мертвых ячеек (ячейки без значения или с искаженным значением) и разрывов помех. [9]

Цель работы состоит в том, чтобы выяснить, можно ли обучить ИНС обнаружению радиолокационных сигналов в условиях смешанных помех и лучше ли ее характеристики по сравнению с обычными радарными детекторами с постоянной вероятностью ложных тревог. Для её достижения проведен анализ и синтез материалов научных публикаций и литературных источников по рассматриваемой теме.

Статистическая природа помех

Как упоминалось ранее, пороговый детектор был бы оптимальным для одного типа плотности вероятности помех, но потребовалась бы априорная информация о плотности вероятности, что на практике представляет собой непростую задачу, решаемую применением адаптивного приема с обучением. Два параметра распределения определяют его форму b и масштаб. Изменение параметра формы приводит к различным негауссовым распределениям [7]. Особым случаем с $b=2$ и $a=1$ является распределение Рэля, которое является хорошим приближением для сигналов, отраженных от цели [7,11]. распределение вероятностей Вейбулла определяется [7]:

$$g(x) = \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^a} \tag{1}$$

На рисунке 2 изображено распределение (плотность) вероятностей Вейбулла для различных значений b .

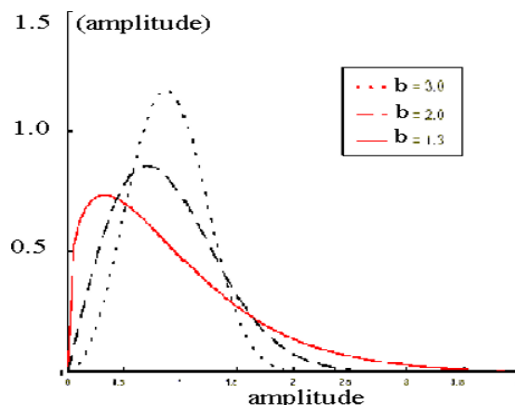


Рис. 2. Распределения вероятностей Вейбулла. Случай $b=2$ соответствует распределению Рэля. Чем меньше b , тем длиннее хвост кривой на графике [3]

Искусственные нейронные сети

Перцептрон — простейшая модель нейрона. Веса ($w_i, i=1, \dots, n$) масштабируют входные сигналы. Также предполагается коэффициент смещения w_0 . На рисунке 3 изображен перцептрон.

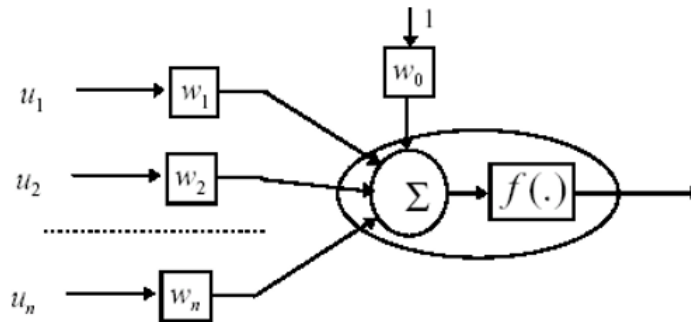


Рис. 3. Модель нейрона

Обычно используемыми функциями активации являются сигмоида и гиперболический тангенс (tgh). Сигмоида имеет следующее выражение:

$$f(x) = \frac{1}{1 + e^{-xw}} \tag{2}$$

ИНС представляют собой реализации, основанные на входных шаблонах, отображаемых на выходные шаблоны, и обычно большинство современных ИНС имеют многоуровневую архитектуру и способность моделировать отношения ввода-вывода.

Способность ИНС моделировать отношения ввода-вывода без каких-либо знаний об этих отношениях называется модельной свободой [10]. Как правило, сеть имеет один входной и один выходной слой, а также один или несколько скрытых слоев. Если поток информации через сеть идет только в одном направлении без петель обратной связи, мы называем это сетью прямой связи. Использование многоуровневой архитектуры перцептрона, показанной на рисунке 4, означает, что может быть некоторая нелинейность между входным и выходным шаблонами, когда в сети достаточно скрытых слоев [10].

Чтобы распознавать закономерности, сеть необходимо обучить с помощью обучающего алгоритма. Обратное распространение — это хорошо известный и в значительной степени успешный алгоритм обучения с учителем для сетей, где входные данные известны априори. Название «обратное распространение» происходит от того факта, что градиенты ошибок скрытых единиц получают путем их обратного распространения, связанного с выходными единицами. Ошибки вычисляются как некоторая функция от расстояния между фактическим выходом и целевым выходом.

Многоуровневые архитектуры доказали свою функциональную способность решать сложные задачи в области обработки информации, но их трудно обучать и масштабировать.

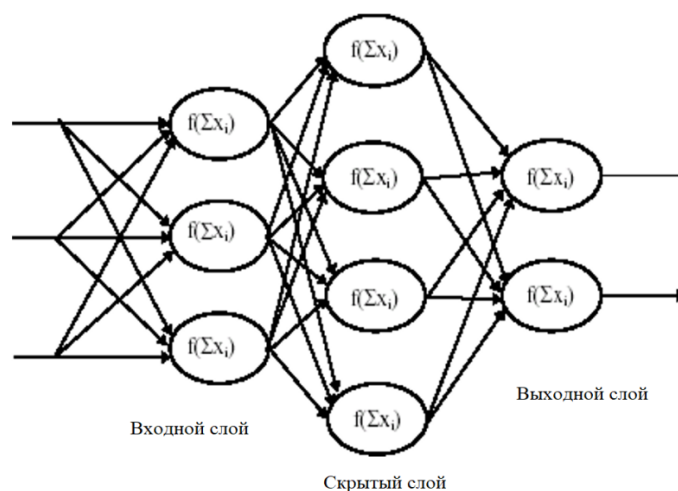


Рис. 4. Архитектура многоуровневого перцептрона

Решение проблемы с ИНС обычно включает следующие шаги:

- Выбор топологии сети, соответствующей характеру проблемы.
- Выбор функций активации, соответствующих характеру проблемы.
- Обучение нейронной сети с помощью процедуры обучения с учителем.
- Тестирование сети по набору образцов, отличных от тех, которые использовались для обучения.

Обнаружение ИНС

Простой детектор с одним входным слоем и одним персептроном, как показано на рисунке 5 имеет сигмоидальную функцию, которая преобразует значения, меньшие или равные нулю, в нулевой выход (цель отсутствует) и в значения больше нуля до единицы (цель присутствует). Предположим, что веса соединений из эталонных ячеек инициализируются вручную с помощью $1/2N$ [7]. Если далее вес на соединении инициализируется единицей, ИНС выдаст результат больше нуля, если значение больше среднего - значения всех окружающих ячеек.

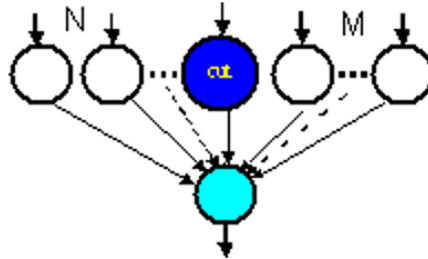


Рис. 5. Детектор на ИНС

Состояние активации первого выходного блока равно:

$$\sum_{t=1}^{N+M} w_t x_t = \frac{1}{2} * \frac{1}{N} \left(\sum_{t=1}^N x_t + \sum_{t=N+1}^{N+1+M} x_t \right), \tag{3}$$

в данном случае с $N=M$, что является тем же результатом, что и усредненный по ячейкам детектор [12]. Таким образом, показано, что теоретически оба детектора должны иметь одинаковую производительность [7]. Далее следует выяснить, имеет ли ИНС преимущества перед обычными детекторами за счет использования простых и сложных архитектур.

Разновидности смеси сигнал/помеха

Чтобы добиться высокой эффективности обнаружения, ИНС-детектор должен быть обучен для распознавания в условиях действия разнообразных помех, в противном случае детектор будет эффективен только в одном конкретном случае. Для этого были синтезированы четыре вида помех, все для распределения вероятностей Вейбулла с $a = 1$ и другим параметром формы b . В частности, $b=2$ является частным случаем теории Вейбулла, называемой Рэлеевским распределением вероятностей.

Случай 1: помеха во времени представляет собой стационарный случайный процесс.

Были обучены четыре структуры из 31 входного нейрона в его скрытом слое. Таким образом, мы получаем специализированную структуру для каждого b параметра Вейбулла. Входной сигнал масштабировался и проходил через сдвиговый регистр. Параметр масштаба $g/(2N)$ применялся к первым 15 и последним 15 входным образцам (рисунок 5). Образец 16 оставили без масштабирования. В таблице 1 показано количество итераций для обучения ИНС.

Таблица 1

Количество итераций для обучения ИНС

Параметр Вейбулла	Итерации	Pfa
a=1 b=2	800	1e-2
a=1 b=1.6666	840	1e-2
a=1 b=1.3333	900	1e-2
a=1 b=1	520	1e-2

Рисунки 6, 7, 8 и 9 иллюстрируют производительность четырех структур ИНС только с одним нейроном в его скрытом слое и выходным нейроном по сравнению с обычным. В качестве основного оценочного параметра примем условную вероятность правильного обнаружения (при условии наличия цели), которая в большинстве реальных систем лежит в пределах 0.5 – 0.99, тогда как условная вероятность ложной тревоги в пределах $10^{-4} - 10^{-10}$. В отличие от систем цифровой связи, где ошибка приводит к обнаружению альтернативного символа и максимальная вероятность ошибки не может превышать 0.5, здесь ошибка – это пропуск цели, никак не влияющий на эффективность обнаружения.

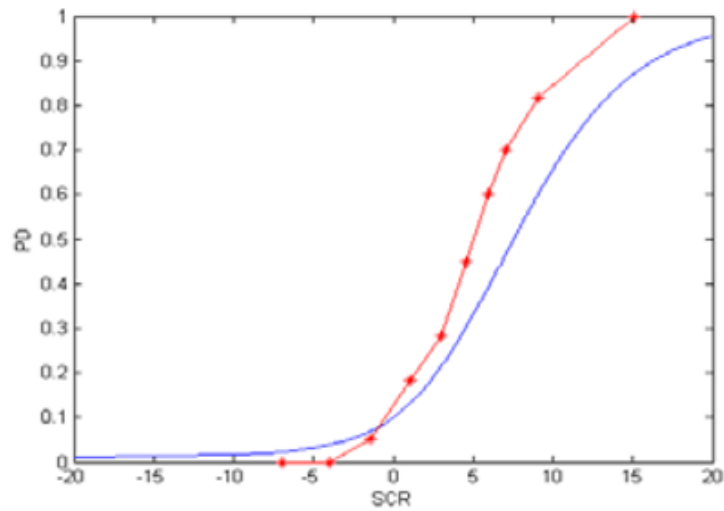


Рис. 6. Условная вероятность правильного обнаружения сигнала в зависимости от отношения сигнал/помеха для ИНС и обычного. Параметры Вейбулла $a=1$, $b=2$ (случай Рэлея).

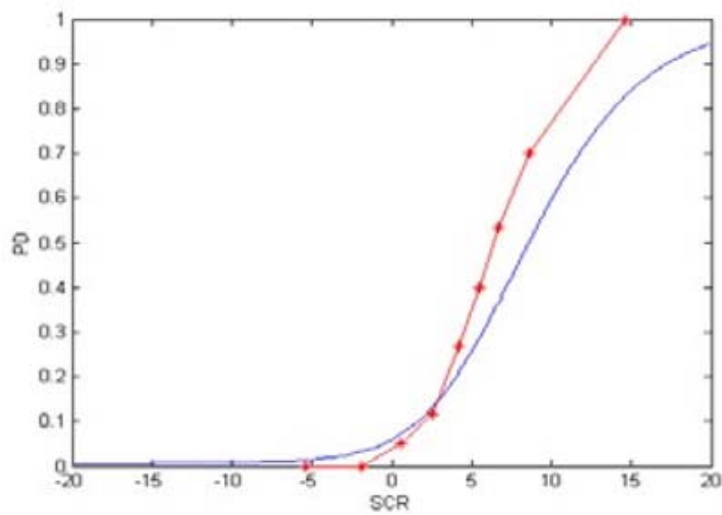


Рис. 7. Условная вероятность правильного обнаружения сигнала в зависимости от отношения сигнал/помеха для ИНС и обычного. Параметры Вейбулла $a=1$, $b=1,6666$

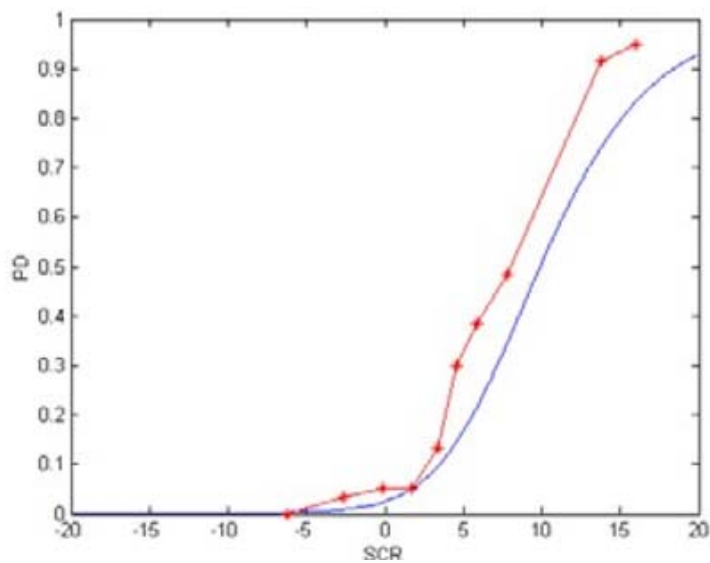


Рис. 8. Условная вероятность правильного обнаружения сигнала в зависимости от отношения сигнал/помеха для ИНС и обычного. Параметры Вейбулла $a=1$, $b=1,3333$

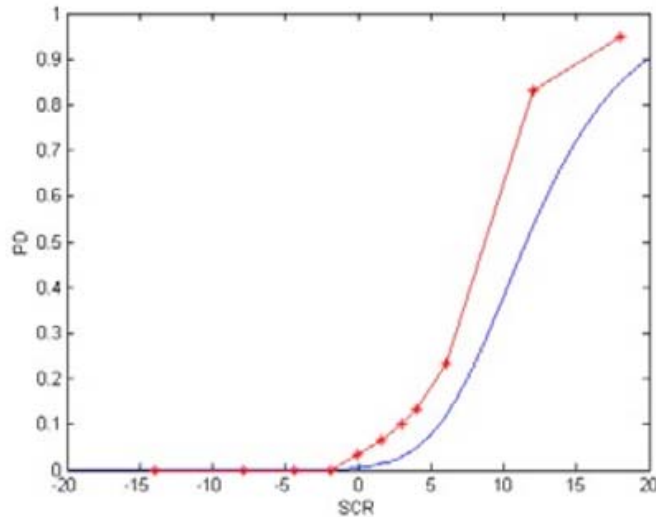


Рис. 9. Условная вероятность правильного обнаружения сигнала в зависимости от отношения сигнал/помеха для ИНС и обычного. Параметры Вейбулла $a=1, b=1$

Случай 2: Смешанная помеха представляет собой нестационарный процесс или, в крайнем случае, локально-стационарный, имеющий непараметрическое описание.

В данном случае задача состоит в том, чтобы найти единую структуру, обеспечивающую хорошую производительность в условиях смешанных помех. ИНС для случая 2 состоит из одного скрытого слоя с 4 нейронами (рисунок 10). Входной сигнал был масштабирован, как и в случае 1, параметр g принимает наихудший случай, который соответствует параметру Вейбулла $b=2$.

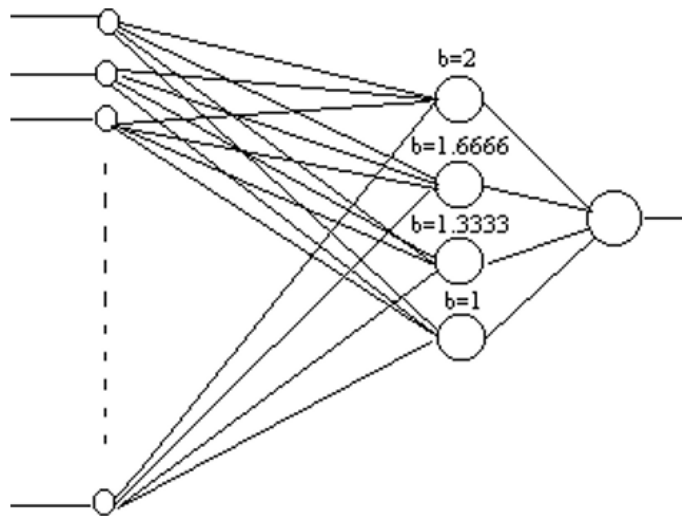


Рис. 10. Структура ИНС для случая 2 (среда смешанных помех)

В таблице 2 приведено количество итераций, которое потребовалось для обучения этой структуры:

Таблица 2

Количество итераций, которое потребовалось для обучения

Итерации	P_{fa}
1840	$1e-2$

Результаты расчетов вероятности обнаружения для разных параметров модели Вейбулла приведены на рисунках 11, 12, 13 и 14 (синяя линия).

Для сравнения результаты для гомогенной помехи показаны красным цветом.

Затем вероятность обнаружения обычного детектора сравнивается с обобщенной структурой ИНС с четырьмя элементами в ее скрытом слое, показанной черной пунктирной линией.

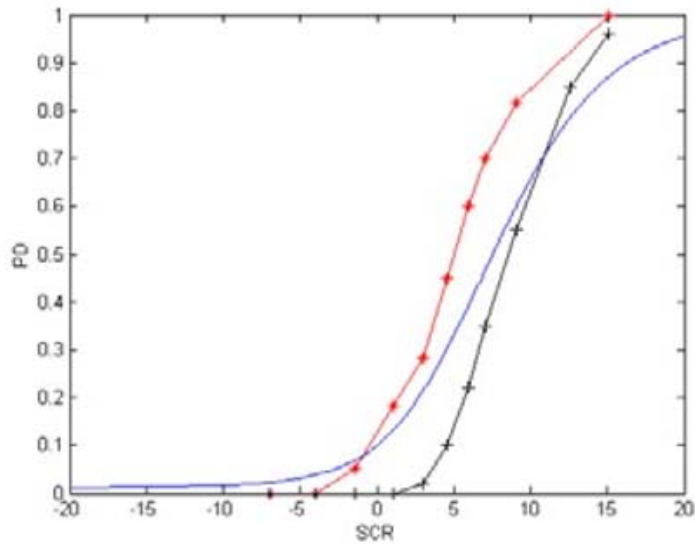


Рис. 11. Вероятность правильного обнаружения сигнала в зависимости от отношения сигнал/помеха для ИНС и детектора CFAR. Параметры Вейбулла $a=1$, $b=2$

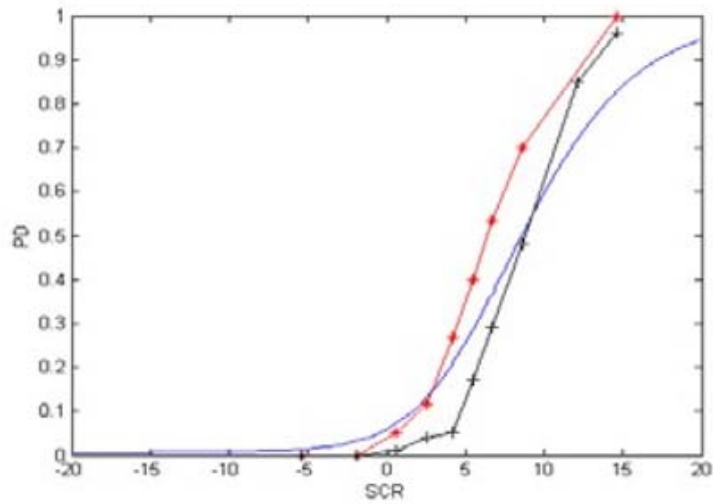


Рис. 12. Вероятность правильного обнаружения сигнала в зависимости от отношения сигнал/помеха для ИНС и детектора CFAR. Параметры Вейбулла $a=1$, $b=1,6666$

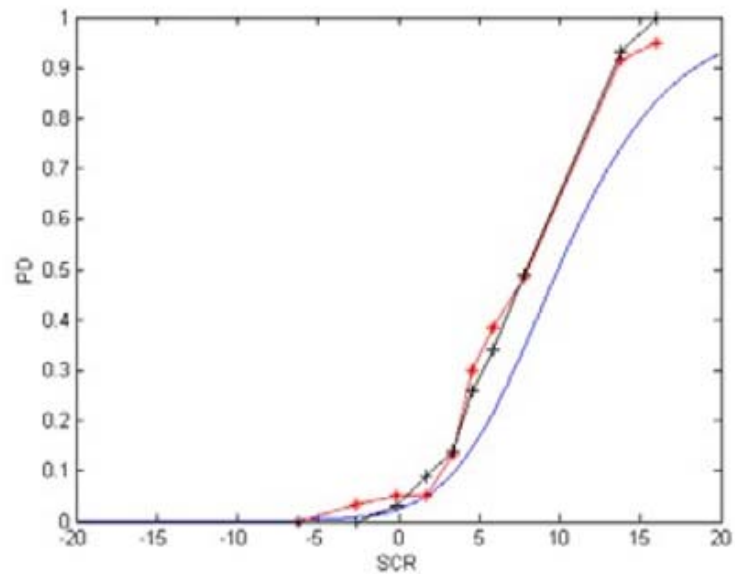


Рис. 13. Вероятность правильного обнаружения сигнала в зависимости от отношения сигнал/помеха для ИНС и детектора CFAR. Параметры Вейбулла $a=1$, $b=1,3333$

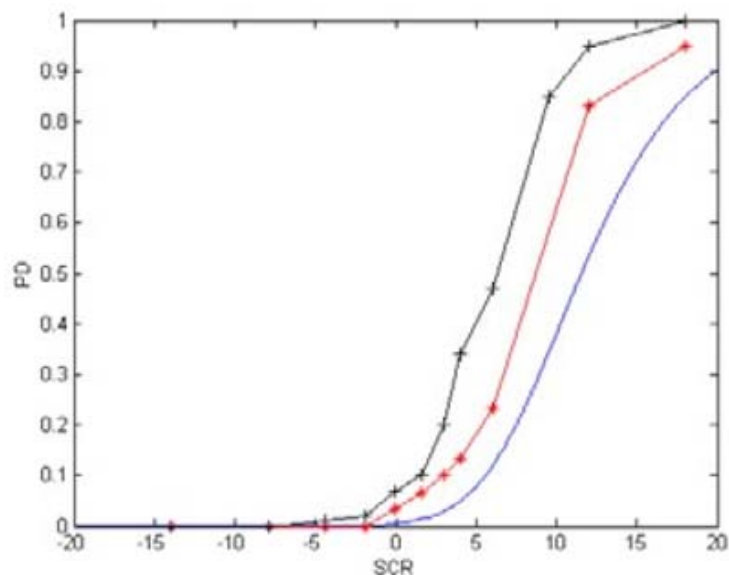


Рис. 14. Вероятность правильного обнаружения сигнала в зависимости от отношения сигнал/помеха для ИНС и детектора CFAR. Параметры Вейбулла $a=1$, $b=1$

Выводы

1. Обычный детектор CFAR оптимален только для гомогенных помех, и статистические подходы требуют достаточных знаний о помехах для построения оптимальных обнаружителей. Изменения в помеховой обстановке приведут к изменениям в эффективности обнаружения. Надежность в этом смысле означает, что частота ложных тревог должна оставаться постоянной, даже когда происходят изменения в окружающей среде. На практике такого добиться довольно трудно.

2. В случае гомогенной помехи производительность специализированных ИНС-структур приближается к производительности порогового обнаружителя.

3. Для случая гетерогенной помехи можно сделать вывод, что одна структура ИНС, обученная для среды со смешанными помехами, обеспечивает лучшую производительность для случаев $b = 1,3333$ и $b = 1$, чем обычная. При $b=1,6666$ указанная структура соответствует традиционной.

4. Необходимо иметь в виду, что, в то время как эффективность порогового обнаружителя сильно зависит от статистических характеристик смеси сигнала и шума, ИНС детектор инвариантен к ним. Следовательно, многоуровневые архитектуры обеспечивают лучшие результаты, поскольку подчеркивают способность ИНС синтезировать в единой архитектуре возможности специализированных простых сетей, адекватным этой цели.

Литература

1. *M. J. Minkler G.*, CFAR: The principles of automatic radar detection in clutter, Baltimore: Magellan Book Co., 1990, p. 384.
2. *Татузов А.Л.* Нейронные сети в задачах радиолокации. М.: Радиотехника, 2009. 432 с.
3. *Тихонов В.И., Харисов В.Н.* Статистический синтез и анализ радиотехнических устройств и систем. М.: Радио и связь, 2015. 608 с.
4. *Левин Б.Р.* Теоретические основы статистической радиотехники. М.: Радио и связь, 1989. 718 с.
5. *Акимов П.С.* и др. Обнаружение радиосигналов / под ред. А. А. Колосова. М.: Радио и связь, 1989. 288 с.
6. *Расторгуев С.П.* Информационная война. М.: Радио и связь, 1999. 416 с.
7. *Kück M.* Constant False Alarm Rate Detection of Radar Signals with Artificial Neural Networks. Höskolan Skövde, 1996.
8. *Круглов В.В., Дли М.И., Голунов Р.Ю.* Нечеткая логика и искусственные нейронные сети. М.: ФИЗМАТЛИТ, 2001. 224 с.
9. *Круглов В.В., Борисов В.В.* Искусственные нейронные сети. Теория и практика. М.: Горячая линия – Телеком, 2002. 352 с.
10. *Хайкин С.* Нейронные сети: полный курс. М.: Издательский дом “Вильямс”, 2006. 1104 с.
11. *Бардин К.В.* Проблема порогов чувствительности и психофизические методы. М.: Наука, 1976. 456 с.
12. *Быков В.В.* Цифровое моделирование в статистической радиотехнике. М.: Сов. радио, 1971. 325 с.

USING AN ARTIFICIAL NEURAL NETWORK FOR SIGNAL DETECTION

Gerel M. Bugaeva

Graduate MTUCI, Moscow, Russia
gerel.bugaeva@mail.ru

Vladislav R. Nabiev

Graduate MTUCI, Moscow, Russia
nabiev_vlad@mail.ru

Vladimir N. Repinski

Associate Professor of the Department of ISU&A, Ph.D., MTUCI, Moscow, Russia
vnrepinski@gmail.com

Abstract

Some aspects considered in this study include constant false alarm probability (CFLT) detection methods implemented in the traditional way and how they differ from the implementation of artificial neural networks (ANNs). Then, a brief description of interference and targets modeled with the Weibull and Rayleigh distributions is presented. Particular attention is paid to ANN as a detector. To evaluate the performance, a comparison of it with a conventional detector is presented and an evaluation of two ANN architectures was carried out.

Keywords: *signal, statistical, neural network detectors, simulation models.*

АНАЛИЗ АЛГОРИТМОВ АДАПТИВНОЙ ФИЛЬТРАЦИИ В СИСТЕМЕ БЕЗ ОПОРНОГО СИГНАЛА

Припутин Владимир Сергеевич,
доцент кафедры РТС, доц., к.т.н., МТУСИ, Москва, Россия,
v.s.priputin@mtuci.ru

Мартынов Сергей Владимирович,
магистрант МТУСИ, Москва, Россия,
MartynovSergey34@yandex.ru

Аннотация

Настоящая статья посвящена анализу алгоритмов адаптивной фильтрации полезных сигналов без использования опорного сигнала в условиях сложной сигнально-помеховой обстановки. В статье приведены результаты сравнительного анализа и моделирования таких алгоритмов адаптивной фильтрации, как LMS, NLMS, RLS. Результаты проведенного имитационного моделирования показали, что алгоритм NLMS является самым оптимальным.

Ключевые слова: цифровая обработка сигналов, адаптивное помехоподавление, алгоритмы адаптивной фильтрации, алгоритм LMS, алгоритм NLMS, алгоритм RLS.

Введение

Технологии цифровой обработки сигналов (ЦОС) появились давно и к настоящему времени нашли массовое применение в различных сферах жизнедеятельности человека. Аппарат методов ЦОС в настоящее время достаточно хорошо развит. Инженерам и исследователям доступно огромное количество разнообразных методов цифровой фильтрации, а также реализующих их алгоритмов. Среди всего многообразия алгоритмов цифровой фильтрации отдельное место занимают алгоритмы адаптивной фильтрации. Данные алгоритмы нашли широкое применение при решении задач, связанных с подавлением помех в таких областях, как акустика, радиолокация, радионавигация и связь, медицина (при проведении ЭКГ и МРТ), а также в других областях.

Настоящая статья посвящена сравнительному анализу алгоритмов адаптивной фильтрации для подавления помех в системе без опорного сигнала. При синтезе адаптивного фильтра, осуществляющего оптимальную оценку полезного сигнала в условиях сложной сигнально-помеховой обстановки, могут быть использованы различные алгоритмы адаптивной фильтрации. В данной работе проанализированы следующие алгоритмы адаптивной фильтрации: наименьших квадратов LMS, нормализованный алгоритм наименьших квадратов NLMS и рекурсивный алгоритм наименьших квадратов RLS.

1. Адаптивное подавление помех

Помехоподавление – это метод оценки полезного сигнала на основе анализа наблюдаемого сигнала, представляющего собой аддитивную смесь искомого сигнала и помех (шумов). Если параметры полезного сигнала и помех неизвестны или непрерывно изменяются во времени, то возникает необходимость применения адаптивных фильтров. В отличие от традиционных цифровых фильтров с фиксированными параметрами, адаптивные фильтры не имеют постоянных коэффициентов, они непрерывно обновляются в процессе работы. Различают алгоритмы адаптивной фильтрации с опорным сигналом и без него.

Адаптивное подавление помех – это подход, используемый для оценки полезного сигнала $d(n)$ по искаженному помехами наблюдению $x(n) = d(n) + v_1(n)$. Обычно метод использует первичный вход, содержащий искаженный сигнал, и эталонный вход, содержащий помехи, каким-то неизвестным образом коррелирующий с первичными помехами $v_1(n)$. Эталонный вход $v_2(n)$ можно выделить и вычесть из основного входа, чтобы получить оценку сигнала $\hat{d}(n)$. Принцип подавления помех показан на рисунке 1.

В процессе вычитания помех из наблюдаемого сигнала оцениваемый полезный сигнал может быть существенным образом искажен. Для того, чтобы этого не происходило требуется, чтобы оценка помех $\hat{v}_1(n)$ была точной копией $v_1(n)$. Поскольку свойства каналов передачи неизвестны и непредсказуемы, параметры процессов фильтрации и вычитания управляются адаптивным процессом. В частности, используется адаптивный фильтр, способный регулировать свою импульсную характеристику для минимизации сигнала ошибки, который, в свою очередь, зависит от выходного сигнала фильтра [1, 8-11].

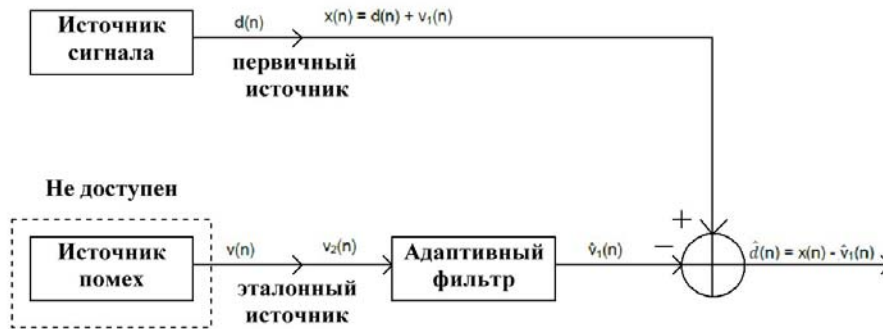


Рис. 7. Принцип адаптивного подавления помех

Без какой-либо информации о $d(n)$ или $v_1(n)$ невозможно отделить полезный сигнал от помех. Если известен (задан) опорный сигнал $v_2(n)$, который коррелирует с первичными помехами $v_1(n)$, то его можно использовать для оценки полезного сигнала $d(n)$ по наблюдению [2]. Опорный сигнал можно получить, поместив один или несколько приемных элементов в поле помех. К сожалению, во многих приложениях эталонный сигнал недоступен. В частности, в нестационарных средах требуемая статистика $v_1(n)$ и $v_2(n)$, как правило, неизвестна. Однако, даже при отсутствии опорного сигнала все еще можно синтезировать устройство подавления помех.

2. Адаптивный фильтр

Адаптивные фильтры — это фильтры с возможностью адаптации к неизвестной среде. Эти фильтры можно использовать в приложениях, где входной сигнал неизвестен или не обязательно стационарен. В отличие от традиционных методов проектирования цифровых фильтров адаптивные фильтры не имеют постоянных коэффициентов фильтрации [2]. Адаптивные фильтры имеют возможность регулировки своей импульсной характеристики. Они требуют малых или вообще не требуют априорных знаний о характеристиках сигнала и помех.

Адаптивный фильтр состоит из двух отдельных частей: цифрового фильтра с регулируемыми коэффициентами $W_n(z)$ и адаптивного алгоритма, который используется для настройки или изменения коэффициентов фильтра [5]. Существуют различные типы алгоритмов адаптивной фильтрации. Адаптивный фильтр может быть построен на базе как КИХ-, так и БИХ-фильтров. Блок-схема адаптивного фильтра показана на рисунке 2.

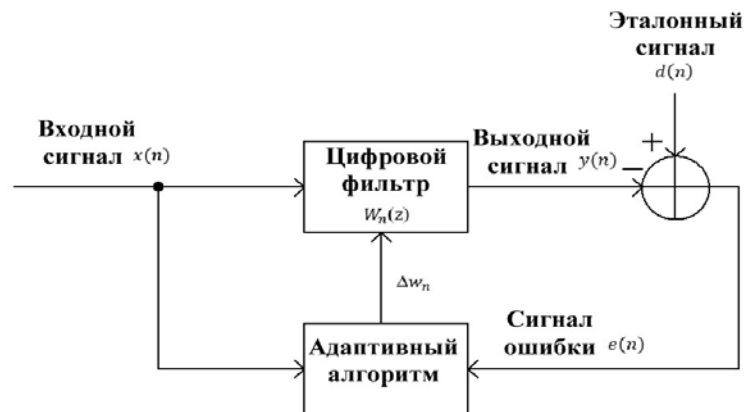


Рис. 2. Блок-схема работы адаптивного фильтра

Процесс фильтрации включает в себя вычисление выходного сигнала фильтра $y(n)$ в ответ на входной сигнал фильтра $x(n)$. Выходной сигнал фильтра сравнивается с требуемым откликом $d(n)$, генерируя оценку ошибки $e(n)$, как показано на рисунке 4. Затем сигнал ошибки обратной связи $e(n)$ используется адаптивным алгоритмом для изменения коэффициентов фильтра w_n , с целью минимизации ошибки оценки по некоторому критерию оптимизации.

Алгоритму адаптивного фильтра для расчета новых коэффициентов по каждой выборке отсчетов наблюдаемого сигнала требуется так называемое уравнение обновления коэффициентов. Уравнение обновления коэффициентов

$$w_{n+1} = w_n + \Delta w_n$$

является частью адаптивного алгоритма, где Δw_n — поправка на выборке n , которая применяется к коэффициентам фильтра w_n для формирования нового набора коэффициентов w_{n+1} на выборке $n+1$ [2].

При синтезе адаптивного фильтра необходимо определить порядок формирования поправки ΔW_n . Основная цель состоит в том, чтобы последовательность исправлений уменьшала ошибку наименьших квадратов или среднеквадратичную ошибку. Адаптивный фильтр, согласно Монсону Х. Хейсу [2], должен обладать следующими свойствами:

- 1) В стационарной среде адаптивный фильтр должен производить последовательность поправок ΔW_n таким образом, чтобы W_n сходились к решению Винера;
- 2) Для расчета ΔW_n не обязательно знать статистику сигнала. Оценка статистики должна быть «встроена» в адаптивный фильтр;
- 3) Для нестационарных сигналов фильтр должен уметь адаптироваться к изменяющейся статистике последних и «отслеживать» решение по мере его развития во времени.

Наиболее важным при реализации адаптивного фильтра является требование к формированию сигнала ошибки $e(n)$. Без сигнала ошибки фильтр не может адаптироваться к изменяющимся условиям функционирования, поскольку последовательность ошибок определяет как должны быть скорректированы коэффициенты фильтра.

2.1. Алгоритмы адаптивной фильтрации

Адаптивные алгоритмы используются для настройки коэффициентов цифрового фильтра таким образом, чтобы сигнал ошибки был минимальным по заданному критерию. Существуют различные типы алгоритмов адаптивной фильтрации, показанные на рисунке 3.

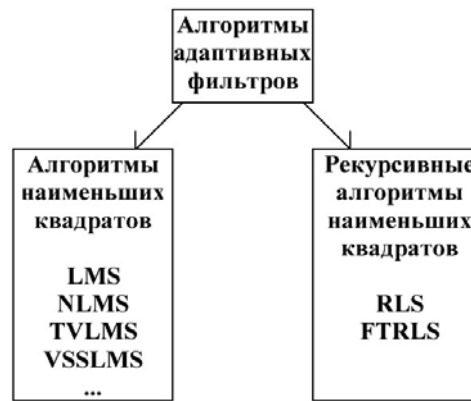


Рис. 3. Классификация алгоритмов адаптивной фильтрации

Алгоритмы наименьших квадратов LMS адаптируют коэффициенты фильтра до тех пор, пока разница между желаемым и фактическим сигналом не будет минимизирована, что связано с получением наименьших среднеквадратичных значений сигнала ошибки [4]. Помимо наименьшего среднеквадратичного LMS, к этому классу алгоритмов относятся нормализованное наименьшее среднеквадратичное NLMS, изменяющееся во времени наименьшее среднеквадратичное TVLMS и наименьшее среднеквадратичное значение с переменным размером шага VSSLMS. Алгоритмы LMS основаны на методе стохастического градиентного спуска в том смысле, что коэффициенты фильтра адаптируются только на основе ошибки в текущий момент времени.

Алгоритмы рекурсивной адаптивной фильтрации, такие как рекурсивный метод наименьших квадратов RLS или быстрый поперечный рекурсивный метод наименьших квадратов FTRLS, рекурсивно находят коэффициенты, которые минимизируют взвешенную линейную функцию стоимости наименьших квадратов, относящуюся к входным сигналам. Этот подход отличается от метода LMS наименьших квадратов, целью которого является уменьшение среднеквадратичной ошибки. Адаптация коэффициентов фильтра основана на всех данных об ошибках, но при использовании коэффициента забывания более старые данные могут быть уменьшены по сравнению с более новыми данными [4].

Наиболее часто используемые адаптивные алгоритмы для подавления помех — это LMS наименьшего среднеквадратичного значения, нормализованное NLMS наименьшего среднего квадрата и рекурсивный алгоритм RLS наименьшего квадрата. Поэтому три разных алгоритма будут сравниваться и исследоваться более подробно.

Алгоритм наименьших квадратов LMS

LMS является одним из самых простых и наиболее широко используемых алгоритмов адаптивной фильтрации. Алгоритм LMS основан на методе стохастического градиентного спуска для нахождения вектора коэффициентов, который минимизирует функцию потерь [2]. Параметры алгоритма LMS меняются для каждой новой выборки.

В адаптивном фильтре наискорейшего спуска уравнение обновления весового вектора имеет вид, где * представляет собой комплексно-сопряженное выражение [7]:

$$w_{n+1} = w_n + \mu E \{ e(n) x^*(n) \}$$

Алгоритм наискорейшего спуска показывает практическое ограничение в том, что ожидание $E\{e(n)x^*(n)\}$ обычно неизвестно. Следовательно, его необходимо заменить такой оценкой, как выборочное среднее:

$$\hat{E}\{e(n)x^*(n)\} = \frac{1}{L} \sum_{l=0}^{L-1} e(n-l)x^*(n-l)$$

При включении этой оценки в алгоритм наискорейшего спуска обновление весового вектора становится:

$$w_{n+1} = w_n + \frac{\mu}{L} \sum_{l=0}^{L-1} e(n-l)x^*(n-l)$$

В (1) среднее по ансамблю оценивается с использованием среднего по выборке (L=1).

$$\hat{E}\{e(n)x^*(n)\} = e(n)x^*(n) \tag{1}$$

Наконец, объединение уравнений приводит к уравнению обновления (2), известному как алгоритм LMS:

$$w_{n+1} = w_n + \mu e(n)x^*(n), \tag{2}$$

где w_n – оценка вектора значения веса в момент времени n, $x(n)$ – вектор входного сигнала, $e(n)$ – вектор ошибки фильтра, а μ – размер шага, который определяет скорость сходимости фильтра и общее поведение.

Одной из трудностей при проектировании и реализации адаптивного фильтра LMS является выбор размера шага μ . Этот параметр должен лежать в определенном диапазоне, чтобы алгоритм LMS сходил:

$$0 < \mu < \frac{2}{\lambda_{max}}$$

где λ_{max} – наибольшее собственное значение автокорреляционной матрицы R_x [2].

Нормализованный LMS алгоритм наименьших квадратов – NLMS

Проблема использования алгоритма LMS заключается в подборе размера шага μ . Для решения этой проблемы был разработан алгоритм Normalized LMS. Поправка, примененная к весовому вектору w_n в выборке $n+1$, нормирована относительно входного вектора $x(n)$ на итерации n .

Для стационарных процессов в широком смысле алгоритм LMS сходится в среднем квадрате автокорреляционной матрицы:

$$0 < \mu < \frac{2}{tr(R_x)}$$

Оценка в приведенном выше уравнении может быть рассчитана из:

$$tr(R_x) = (p+1) E\{|x(n)|^2\}$$

Поэтому условие среднеквадратичной сходимости можно заменить на:

$$0 < \mu < \frac{2}{(p+1) E\{|x(n)|^2\}}$$

Где $E\{|x(n)|^2\}$ – мощность входного речевого сигнала $x(n)$. Эту мощность можно оценить, используя среднее значение по времени, например:

$$\hat{E}\{|x(n)|^2\} = \frac{1}{p+1} \sum_{k=0}^p |x(n-k)|^2$$

Это приводит к следующей оценке размера шага для среднеквадратичной сходимости:

$$0 < \mu < \frac{2}{x^H(n)x(n)}$$

Тогда изменяющийся во времени размер шага определяется выражением:

$$\mu(n) = \frac{\beta}{x^H(n)x(n)} = \frac{\beta}{|x(n)|^2},$$

где β – нормализованный размер шага в диапазоне $0 < \beta < 2$. Замена μ в уравнении обновления весового вектора LMS на $\mu(n)$ приводит к нормализованному алгоритму LMS NLMS в (3).

$$w_{n+1} = w_n + \beta \frac{x^*(n)}{|x(n)|^2} e(n) \quad (3)$$

Нормализованный алгоритм LMS можно рассматривать как алгоритм LMS с изменяющимся во времени параметром размера шага.

Рекурсивный алгоритм наименьших квадратов - RLS

В отличие от алгоритма LMS, целью которого является уменьшение среднеквадратичной ошибки, алгоритм RLS направлен на рекурсивный поиск коэффициентов фильтра, наименьших среднеквадратичных значений сигнала ошибки [3]. По сравнению со среднеквадратичной ошибкой, ошибку наименьших квадратов можно минимизировать непосредственно из данных $x(n)$ и $d(n)$. Рекурсивный алгоритм наименьших квадратов создает набор коэффициентов фильтра W_n (k) для выборки n, которые минимизируют взвешенную ошибку метода наименьших квадратов.

$$\varepsilon(n) = \sum_{i=0}^n \lambda^{n-1} |e(i)|^2$$

Алгоритм RLS включает рекурсивное обновление вектора W_n и обратной матрицы автокорреляции $P(n)$. При оценке вектора усиления $g(n)$ и обратной матрицы автокорреляции $P(n)$ необходимо вычислить произведение:

$$z(n) = P(n-1)x^*(n)$$

Вектор усиления $g(n)$ можно рассчитать следующим образом:

$$g(n) = \frac{1}{\lambda + x^T x(n) z(n)} z(n),$$

где λ – так называемый «фактор забывания», придающий экспоненциально меньший вес более старым образцам ошибок. Следовательно, он также определяется как экспоненциальный весовой коэффициент и находится в диапазоне $0 < \lambda \leq 1$. Включая эти определения, матрица автокорреляции $P(n)$ приводит к:

$$P(n) = \frac{1}{\lambda} [P(n-1) - g(n)z^H(n)]$$

В (4) вычисляется так называемая априорная ошибка $\alpha(n)$, представляющая собой разницу между полезным сигналом $d(n)$ и оценкой $\hat{d}(n)$, полученной путем применения предыдущего набора коэффициентов фильтра W_{n-1} в новый вектор данных $x(n)$.

$$\alpha(n) = d(n) - w_{n-1}^T x(n) \quad (4)$$

Наконец, объединение уравнений приводит к экспоненциально взвешенному алгоритму рекурсивных наименьших квадратов RLS (5).

$$w_n = w_{n-1} + \alpha(n) g(n) \quad (5)$$

Частный случай $\lambda=1$ называется алгоритмом RLS с растущим окном, где все предыдущие ошибки считаются равными по весу в общей ошибке [4]. Поскольку алгоритм RLS включает рекурсивное обновление вектора W_n и обратной матрицы автокорреляции $P(n)$, требуются начальные условия для обоих этих членов [2].

Основные различия рассмотренных алгоритмов адаптивной фильтрации

Скорость сходимости

Скорость сходимости определяется как количество циклов адаптации, необходимых для сходимости алгоритма от некоторого начального состояния к его стационарному состоянию или достаточно близкому к оптимуму, подобно оптимальному решению Винера в смысле среднеквадратичной ошибки [4].

Сходимость адаптивного алгоритма LMS:

Характеристики сходимости адаптивного алгоритма LMS зависят от двух факторов: размера шага μ и разброса собственных значений автокорреляционной матрицы $\chi(Rx)$. Размер шага μ должен лежать в определенном диапазоне:

$$0 < \mu < \frac{2}{\lambda_{max}},$$

где λ_{max} – наибольшее собственное значение автокорреляционной матрицы Rx . Большое значение размера шага μ приведет к более быстрой сходимости, но может быть менее стабильным около минимального значения. Сходимость алгоритма обратно пропорциональна разбросу собственных значений корреляционной матрицы [6].

Сходимость адаптивного алгоритма NLMS:

В NLMS-алгоритме зависимость μ от автокорреляционной матрицы преодолевается за счет использования переменного параметра размера шага, в котором изменение достигается за счет деления на каждой итерации фиксированного размера шага на входную мощность. Переменный размер шага вычисляется как:

$$\mu(n) = \frac{\beta}{x^H(n)x(n)} = \frac{\beta}{|x(n)|^2},$$

где β – нормализованный размер шага с ($0 < \beta < 2$). Использование переменного размера шага устраняет большую часть компромисса между остаточной ошибкой и скоростью сходимости по сравнению с фиксированным размером шага.

Сходимость адаптивного алгоритма RLS:

По сравнению с адаптивным алгоритмом LMS, где поведение сходимости зависит от размера шага μ , скорость сходимости адаптивного фильтра RLS основана на обратной матрице автокорреляции $P(n)$, которая приводит к отбеливанию входных данных касания. Далее это зависит от экспоненциального весового коэффициента λ . Экспоненциальный весовой коэффициент λ должен быть больше нуля и меньше или равен единице $0 < \lambda \leq 1$.

Характеристики ошибок

Адаптивные фильтры пытаются оптимизировать производительность, сводя к минимуму сигнал ошибки между выходным сигналом адаптивного фильтра и желаемым сигналом в соответствии с некоторым критерием. Большое значение ошибки указывает на то, что адаптивный фильтр не может точно отслеживать нужный сигнал. Минимальное значение ошибки гарантирует, что адаптивный фильтр оптимален. Различные алгоритмы адаптивной фильтрации сильно зависят от критерия оптимизации.

Минимальная среднеквадратическая ошибка MSE:

Критерием алгоритма LMS и NLMS является минимальное среднеквадратичное значение сигнала ошибки. MSE определяется как среднее по ансамблю квадрата последовательности ошибок, обозначаемое как:

$$\xi(n) = E\{|e(n)|^2\}$$

Так называемая неправильная настройка – еще одна мера производительности для алгоритмов, использующих минимальный критерий MSE. Неправильная настройка M представляет собой отношение стационарной избыточной среднеквадратической ошибки к минимальной среднеквадратической ошибке, которую можно математически описать как:

$$M = \frac{\xi_{ex}(\infty)}{\xi_{min}}$$

Компромисс между низкой скоростью сходимости и малой среднеквадратичной ошибкой или рассогласованием необходим, поскольку при увеличении размера шага μ скорость сходимости снижается, но увеличивается СКО [8].

Минимальная ошибка метода наименьших квадратов LSE:

В алгоритме RLS должны быть рассмотрены две различные ошибки. Априорная ошибка оценки $\xi(n)$ – это ошибка, которая произошла бы, если бы коэффициенты фильтра не были обновлены, а апостериорная ошибка $e(n)$, с другой стороны, произошла после обновления вектора веса. Критерий оптимизации наименьших квадратов алгоритма RLS зависит в общем случае от функции стоимости $\varepsilon(n)$, основанной на $e(n)$, а не $\xi(n)$.

$$\varepsilon(n) = \sum_{i=0}^n \lambda^{n-1} |e(i)|^2$$

2.2. Адаптивное подавление помех без опорного сигнала

Как уже упоминалось ранее, во многих нестационарных приложениях трудно или невозможно получить опорный сигнал. Например, когда широкополосный сигнал искажается периодическими помехами, эталонный сигнал недоступен. В этом случае можно получить опорный сигнал, задержав зашумленный процесс $x(n) = d(n) + v_1(n)$. Задержанный сигнал $x(n - n_0)$ используется в качестве опорного сигнала для устранения помех. Принцип помехоподавления без опорного сигнала показан на рисунке 4.

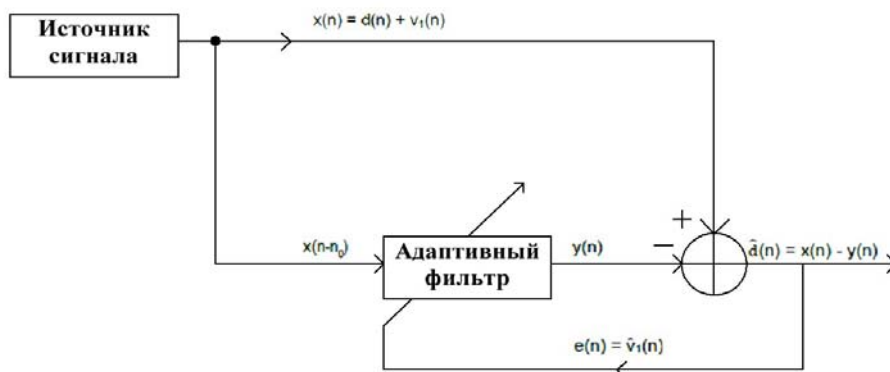


Рис. 4. Принцип адаптивного помехоподавления без опорного сигнала

Выбранная фиксированная задержка n_0 должна иметь достаточную длину, чтобы компоненты широкополосного сигнала на эталонном входе стали некоррелированными с компонентами на первичном входе. Компоненты интерференции из-за их периодического характера останутся коррелированными друг с другом. Среднеквадратическая ошибка формируется путем получения разницы между $x(n)$ и выходным сигналом адаптивного фильтра $y(n)$.

Выход системы $e(n)$ равен:

$$e(n) = d(n) + v_1(n) - y(n)$$

Возводя в квадрат, получаем:

$$e(n)^2 = v_1(n)^2 + [d(n) - y(n)]^2 + 2v_1(n) * [d(n) - y(n)]$$

Учитывая ожидания с обеих сторон:

$$E\{e(n)^2\} = E\{v_1(n)^2\} + E\{[d(n) - y(n)]^2\} + 2E\{v_1(n) * [d(n) - y(n)]\}$$

Кроме того, вход адаптивного фильтра равен $x(n - n_0)$, следовательно:

$$E\{v_1(n) * y(n)\} = E\{[v_1(n) * d(n - k)]\} + E\{[v_1(n) * v_1(n - k)]\}$$

Поскольку помеха $v_1(n)$ некоррелирована как с $d(n)$, так и с $v_1(n - k)$, минимизация среднеквадратичной ошибки $E\{e(n)^2\}$ эквивалентна минимизации $E\{[d(n) - y(n)]^2\}$ среднеквадратическая ошибка между $d(n)$ и выходным сигналом адаптивного фильтра $y(n)$.

$$\min E\{e(n)^2\} = \min E\{v_1(n)^2\} + \min E\{[d(n) - y(n)]^2\}$$

предположении, что $d(n)$ является широкополосным процессом и задержка n_0 больше времени декорреляции, запаздывающий процесс $x(n - n_0)$ не коррелирует с помехами $v_1(n)$, но коррелирует с $d(n)$. В результате получается адаптивный фильтр, который дает минимальную среднеквадратичную оценку широкополосного процесса $d(n)$, а входной сигнал ошибки $e(n)$ соответствует оценке помех $v_1(n)$ [2].

3. Имитационное моделирование алгоритмов

Было проведено имитационное моделирование рассмотренных ранее алгоритмов адаптивной фильтрации в среде Matlab. В процессе моделирования на вход каждого адаптивного фильтра подвались одинаковые выборки отсчетов сигнала, содержащие искомым сигнал $d(n)$ и помеховую последовательность $v_1(n)$.

$$x(n) = d(n) + v_1(n)$$

Чтобы извлечь полезную информацию из зашумленного процесса, выполняются серии вычислительных экспериментов, в которых искомым полезный сигнал представляет собой оцифрованную синусоиду с единичной амплитудой, т.е.

$$d(n) = \sin(n\omega_0 + \varphi),$$

где $\omega_0 = 0.05\pi$.

Широкополосный помеховый сигнал возникает из-за внешнего нежелательного воздействия на датчик. Если помеховый сигнал некоррелирован, не будет информации о будущих или прошлых значениях, и полезный сигнал не может быть оценен. Следовательно, помеховая последовательность $v_1(n)$ моделируется как авторегрессионный процесс AR(1), который генерируется разностным уравнением первого порядка:

$$v_1(n) = 0.8v_1(n-1) + g(n),$$

где: $g(n)$ – белый гауссовский шум с нулевым средним и дисперсией $\sigma_g^2 = 0.25$, который не коррелирует с $d(n)$. Авторегрессионный процесс генерируется путем фильтрации последовательности белого шума в Matlab.

При фильтрации последовательности белого шума последовательные выборки становятся коррелированными. Коррелированный помеховый сигнал теперь можно предсказать с помощью метода адаптивного помехоподавления.

На рисунке 5 показан график из 1000 отсчетов желаемой синусоиды и зашумленного процесса $x(n) = d(n) + v_1(n)$.

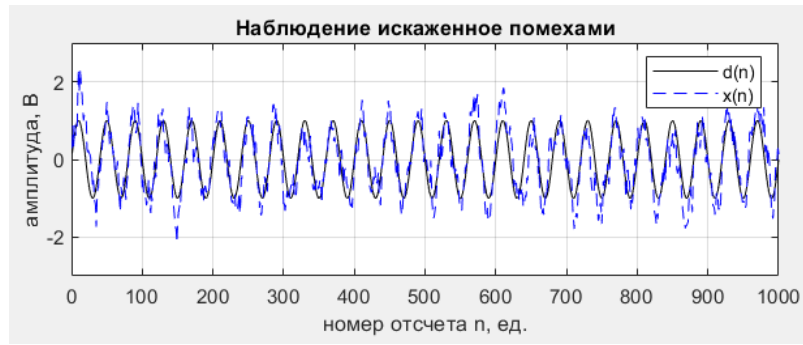


Рис. 5. График эталонной синусоиды и зашумленного сигнала

Измерительная система представляет собой черный ящик, в котором нет коррелированного с помехами опорного сигнала. В этом случае можно получить опорный сигнал, задержав зашумленный процесс $x(n) = d(n) + v_1(n)$. Задержанный процесс $x(n - n_0)$ коррелирует с $d(n)$ и может использоваться в качестве опорного сигнала для оценки полезного сигнала.

$$x(n - n_0) = x(n) \quad , n > n_0$$

Опорный сигнал получается путем задержки $x(n)$ на $n_0 = 25$ отсчетов, что можно увидеть на следующем графике, представленном на рисунке 6.

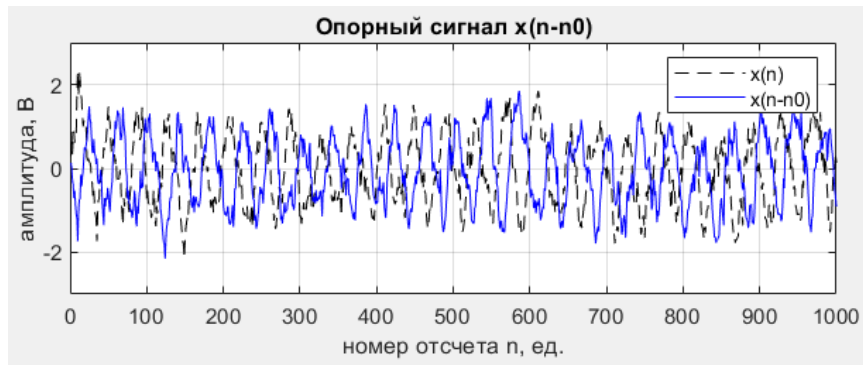


Рис. 6. График опорного сигнала

Моделирование различных алгоритмов адаптивной фильтрации выполнялось с использованием адаптивного помехоподавителя 12-го порядка.

3.1. LMS алгоритм

На рисунке 7 показаны результаты, полученные при применении алгоритма LMS для адаптивного помехоподавления без опорного сигнала. На рисунке можно увидеть оценку полезного сигнала $d(n)$, формируемую адаптивным помехоподавлением с порядком фильтра 12 и размером шага $\mu = 0,002$.

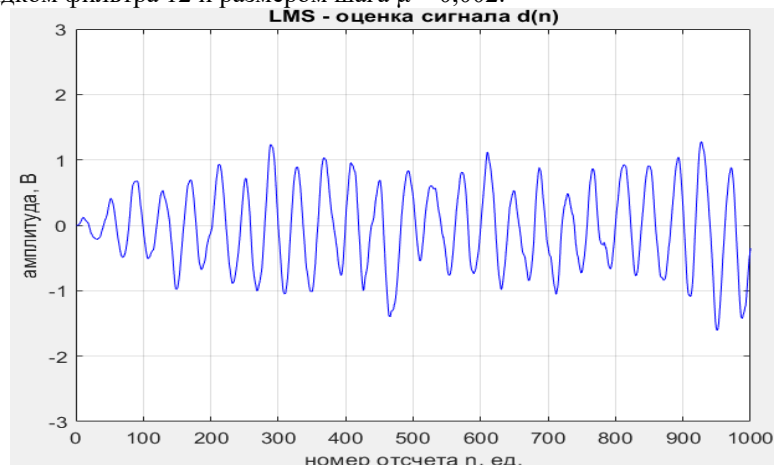


Рис. 7. График оценки полезного сигнала, с применением алгоритма LMS

Анализируя график на рисунке 7, можно сделать вывод, что алгоритм LMS имеет сравнительно слабые показатели помехоподавления. В начале амплитуда оценки сигнала $d(n)$ очень мала по сравнению с полезным сигналом, и форма синусоиды не может быть воспроизведена. Примерно после 200 выборок адаптивный фильтр может дать более точную оценку $d(n)$.

3.2. NLMS алгоритм

Ниже приведены результаты адаптивного помехоподавления с использованием алгоритма NLMS. Оценка полезного сигнала $d(n)$, создаваемого адаптивным помехоподавляющим устройством 12-го порядка с нормированным параметром размера шага $\beta = 0,25$, представлена на графике (рисунок 8).

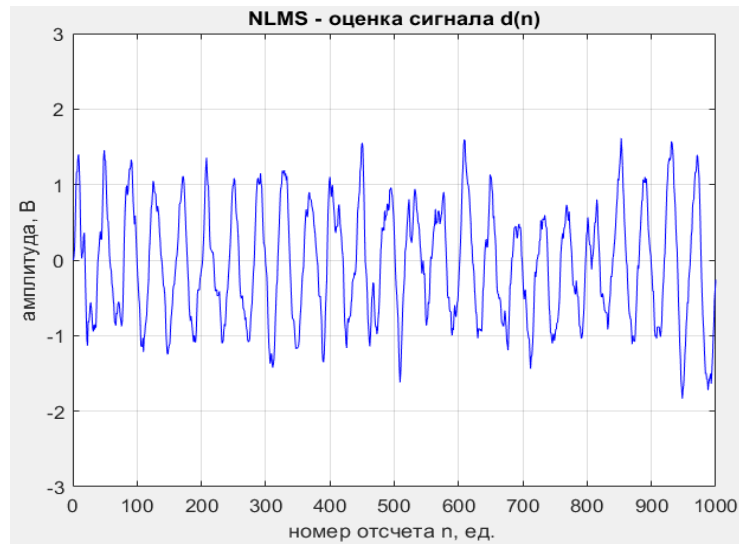


Рис. 8. График оценки полезного сигнала, с применением алгоритма NLMS

Из графика на рисунке 8 следует, что адаптивный фильтр может дать довольно точную оценку $d(n)$ примерно после 150 выборок, однако присутствуют довольно резкие скачки амплитуды. Алгоритм адаптивной фильтрации NLMS достаточно эффективен при адаптивном помехоподавлении.

3.3. RLS алгоритм

Используя адаптивный помехоподавитель 12-го порядка с коэффициентами, которые обновляются с помощью алгоритма RLS, оценка полезного сигнала $d(n)$, которая создается с экспоненциальным весовым коэффициентом $\lambda = 1$, показана на рисунке 9.

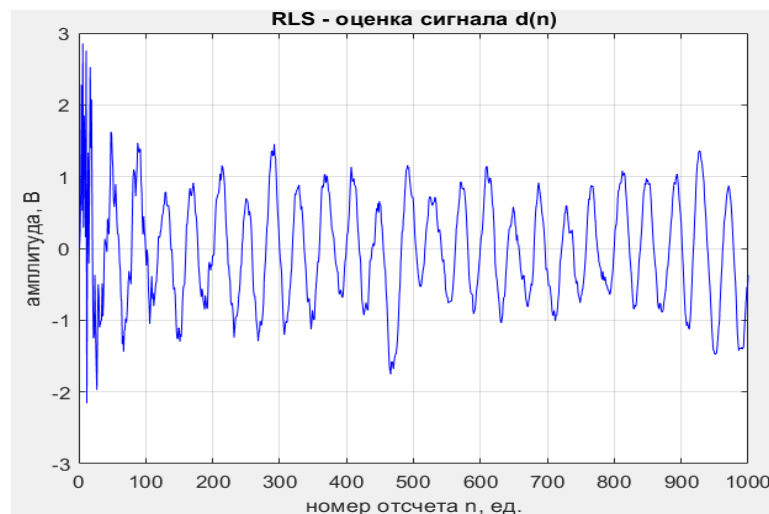


Рис. 9. График оценки полезного сигнала, с применением алгоритма RLS

В начале амплитуда оценки сигнала $d(n)$ очень велика по сравнению с полезным сигналом, и форма синусоиды не может быть воспроизведена. После обработки 200 отчетов входного сигнала адаптивный фильтр начинает давать более точную оценку $d(n)$. Алгоритм RLS показывает хорошую производительность при помехоподавлении.

Для удобства анализа, на рисунке 10 наложены друг на друга результаты оценки полезного сигнала с использованием рассмотренных ранее алгоритмов: LMS, NLMS, RLS.

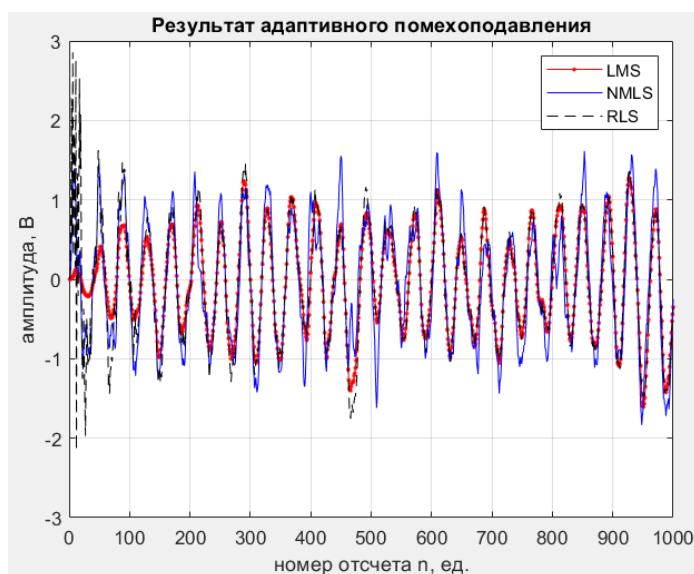


Рис. 10. Графики оценки полезного сигнала, алгоритмов RLS, LMS, NLMS в сравнении друг с другом

Непосредственно сравнивая оценку полезного сигнала $d(n)$ алгоритмов: RLS, LMS и NLMS стоит отметить алгоритм NLMS, как более оптимальный из рассмотренных. Данный алгоритм, при своих недостатках лучше сохраняет амплитуду сигнала на всей кривой оценки сигнала, а также требует меньшее количество выборок для получения достаточно точной оценки $d(n)$.

Выводы

В настоящей статье был проведен сравнительный анализ алгоритмов адаптивной фильтрации без опорного сигнала. Были рассмотрены следующие алгоритмы: алгоритм наименьших квадратов LMS, нормализованный алгоритм наименьших квадратов NLMS и рекурсивный алгоритм RLS наименьших квадратов. По результатам проведенного имитационного моделирования алгоритмов был сделан вывод, что устройство адаптивного помехоподавления без опорного сигнала получает больше преимуществ от использования алгоритма NLMS, нежели от алгоритмов LMS или RLS. Данный алгоритм лучше восстанавливает форму огибающей полезного сигнала в процессе адаптации, а также требует меньший размер временных выборок отсчетов для получения достаточно точной оценки $d(n)$. Алгоритм NLMS с нормализованным размером шага $\beta = 0,25$ видится наилучшим выбором для построения адаптивного помехоподавителя без опорного сигнала.

Литература

1. *A. Singh*. Adaptive Noise Cancellation. 2001. [Электронный ресурс]: <http://www.cs.cmu.edu/~aarti/pubs/ANC.pdf> [Дата обращения: 22. 03. 2022]
2. *M. H. Hayes*. Adaptive filtering in Statistical digital signal processing and modeling. John Wiley & Sons, 1996, pp. 493-570.
3. *G. Yadav and B. Krishna*. Study of different adaptive filter algorithms for noise cancellation in real-Time environment // International Journal of Computer Applications, 2014.
4. *MathWorks*. Compare RLS and LMS Adaptive Filter Algorithms, [Электронный ресурс]: <https://www.mathworks.com/help/dsp/ug/compare-rls-and-lms-adaptive-filter-algorithms.html> [Дата обращения 23 03 2022].
5. *A. Tajammul, D. K. Gupta, V. K. Gupta*. Analysis of Noise Signal Cancellation using LMS, NLMS and RLS Algorithms of Adaptive Filter // Advanced Research in Electrical and Electronic Engineering, vol. 2, 2015.
6. *Джиган В.И.* Адаптивная фильтрация сигналов: теория и алгоритмы. Москва: Техносфера, 2013. 528 с.
7. Studbooks.net: LMS алгоритм [Электронный ресурс]: https://studbooks.net/2369642/tehnika/adaptivnyy_algorithm [Дата обращения 25.03.2022].
8. *Lobov E.M., Priputin V.S., Lobova E.O., Liberman E.A.* Dispersion distortion tracking compensator based on the sigma-point kalman // T-Comm. 2021. Т. 15. № 1. С. 41-47.
9. *Либеровский Н.Ю., Припутин В.С., Чиров Д.С.* Обзор исследований адаптивного формирования диаграммы направленности и цифровой обработки сигналов // Научные исследования в космических исследованиях Земли. 2021. Т. 13. № 6. С. 16-21.

10. Sokolov K.Yu., Priputin V.S., Lobova E.O. Implementation of cosine modulated digital filter bank on processor with ARM architecture // T-Comm. 2020. Т. 14. № 11. С. 57-63.
11. Аджемов С.С., Кучумов А.А., Либеровский Н.Ю., Припутин В.С. Сравнение метода слепого разделения сигналов на базе статистик второго порядка с классическими методами адаптивного подавления помех в случае плохо обусловленных статистик // Системы синхронизации, формирования и обработки сигналов. 2017. Т. 8. № 1. С. 92-97.

ANALYSIS OF ADAPTIVE NOISE CANCELING ALGORITHMS IN A SYSTEM WITHOUT A REFERENCE SIGNAL

Vladimir S. Priputin,
Associate Professor of the Department of RTS, Ph.D., MTUCI, Moscow, Russia,
v.s.priputin@mtuci.ru

Sergei V. Martynov,
Student MTUCI, Moscow, Russia,
MartynovSergey34@yandex.ru

Abstract

The article discusses the basic concept of adaptive noise cancelation. The purpose of the paper is to study adaptive filtering algorithms in a noisy system without a reference signal. The article describes the basic principles and scope of adaptive noise cancelation technology. Algorithms LMS, NLMS, RLS of adaptive filtering in a system without a reference signal are examined and modelled. The study analyses presented algorithms and reveals their features in comparison with each other. The results of the research showed that the NLMS algorithm is the most optimal one.

Keywords: Digital Signal Processing, adaptive noise cancelation, adaptive filtering algorithms, LMS algorithm, NLMS algorithm, RLS algorithm.

СИНТЕЗ НОВЫХ СИГНАЛОВ ДЛЯ ФИЛЬТРА БАТТЕРВОРТА ПО КРИТЕРИЮ МАКСИМУМА СРЕДНЕГО ЗНАЧЕНИЯ ЕГО ОТКЛИКА

Ковалевская Анастасия Андреевна,
студент МТУСИ, Москва, Россия,
kanan88747@mail.ru

Санников Владимир Григорьевич,
доцент кафедры ОТС, к.т.н., МТУСИ, Москва, Россия,
v.g.sannikov@mtuci.ru

Аннотация

Рассматривается проблема передачи цифровых данных по линейному каналу связи без межсимвольной интерференции. Для решения данной проблемы осуществляется согласование формы передаваемых данных с временными и частотными характеристиками линейного канала связи. Для линейного канала связи, с известными импульсной реакцией и комплексным коэффициентом передачи, проводится синтез новых сигналов по критерию максимума среднего значения его отклика. Получены новые общие соотношения для оптимальных сигналов и новых множителей Лагранжа применительно к модели канала в виде фильтра Баттерворта. Приведены результаты моделирования синтезированных сигналов для фильтра Баттерворта второго порядка.

Ключевые слова: линейный канал связи, фильтр Баттерворта, оптимальные финитные сигналы, синтез по максимуму среднего, нивелирование межсимвольной интерференции.

В высокоскоростных телекоммуникационных системах (ТКС) основной причиной снижения их помехоустойчивости является межсимвольная интерференция (МСИ) [1-3]. Для борьбы с явлением МСИ, вызываемой, в основном, нелинейностью АЧХ канала связи, в современных ТКС часто применяются выравнители каналов или эквалайзеры. Последние, как известно [1-6], вносят значительную задержку в принимаемые решения, и используют достаточно сложный приём в целом, основанный на том или ином варианте алгоритма Витерби [4-6]. Следует отметить, что помехоустойчивость такого приема оценивается лишь приближенно.

В работе рассматривается новый подход при организации ТКС. Основу его составляет синтез таких сигналов на передаче, согласованных с АЧХ канала связи, которые принципиально не вызывают МСИ на входе приемника ТКС. Синтез таких сигналов проводится путем решения новой вариационной задачи на условный экстремум. При этом в приемнике ТКС обеспечивается простой поэлементный прием на согласованный фильтр с импульсной реакцией, представляющей зеркальное отображение сигнала на выходе модулятора ТКС. Помехоустойчивость такого приема оценивается не приближенно, но точно. Этому направлению посвящены работы [7-9].

Синтез новых сигналов по критерию максимума среднего значения отклика линейного канала связи

Рассмотрим следующую вариационную задачу на условный экстремум.

Дан линейный канал связи (КС) с известным комплексным коэффициентом передачи $K(j\omega)$.

На вход КС воздействует финитный сигнал $x(t)$, $0 \leq t \leq T$, с фиксированной энергией

$$E_x = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x^2(\omega) d\omega = \int_0^T x^2(t) dt = const, \quad (1)$$

где $S_x(j\omega)$ – комплексный спектр сигнала.

Требуется определить форму сигнала $x(t)$, $0 \leq t \leq T$, таким образом, чтобы отклик КС $y(t)$ на остался финитным, т.е. был задан на интервале времени $0 \leq t \leq T$, а его среднее значение на этом интервале было бы максимальным по всему множеству спектров $S_x(j\omega)$.

Синтезированные по такому критерию сигналы обозначим как – ОФС2.1.

В отличие от аналогичной задачи, рассмотренной в [10], когда требовалось полное отсутствие МСИ на двух интервалах, а именно при $t \geq 0$ и $t \geq T$, в данной работе синтезируются ОФС2.1, обеспечивающие отсутствие МСИ только на одном интервале $t \geq T$.

Выражения для отклика КС и его k -й производной при $k = 0, 1, \dots, (n-1) = \overline{0, (n-1)}$, равны:

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega)K(j\omega)e^{j\omega t} d\omega, \quad \frac{d^k y(t)}{dt^k} = \frac{1}{2\pi} \int_{-\infty}^{\infty} (j\omega)^k S_x(j\omega)K(j\omega)e^{j\omega t} d\omega. \quad (2)$$

Учет этих соотношений позволяет предложить метод приёма дискретных сигналов, а именно, прием по максимуму среднего значения нового сигнала $y(t)$, вида ОФС 2.1.

Этот метод по реализации значительно проще интегрального метода приёма [7] и незначительно уступает по сложности методу однократного отсчета, являясь, очевидно, более помехоустойчивым последнего, т.к. основан на методе накопления полезного сигнала и шума.

Найдем среднее значение сигнала $y(t)$

$$\begin{aligned} \overline{y(t)} &= \frac{1}{T} \int_0^T y(t) dt = \frac{1}{T} \int_0^T \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega)K(j\omega)e^{j\omega t} d\omega dt = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega)K(j\omega) \frac{1}{T} \int_0^T e^{j\omega t} dt d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega)K(j\omega) \frac{e^{j\omega T} - 1}{j\omega T} d\omega = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega)K(j\omega) \frac{e^{j\omega T} - 1}{j\omega T} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega)k(j\omega) \cdot \frac{\sin \omega T / 2}{\omega T / 2} e^{j\omega \frac{T}{2}} d\omega. \end{aligned} \quad (3)$$

Учитывая (1) и равенства нулю в соотношениях (2) при $t \geq T$, рассмотрим функционал

$$J = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega)K(j\omega) \cdot \frac{\sin \omega T / 2}{\omega T / 2} e^{j\omega \frac{T}{2}} d\omega - \frac{1}{2\pi} \left\{ \sum_{k=0}^{n-1} \lambda_k \int_{-\infty}^{\infty} (j\omega)^k S_x(j\omega)K(j\omega) e^{j\omega T} d\omega \right\}.$$

Приведем его к виду

$$J = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega)K(j\omega) \left\{ \frac{\sin \omega T / 2}{\omega T / 2} e^{j\omega \frac{T}{2}} - \sum_{k=0}^{n-1} (j\omega)^k \lambda_k e^{j\omega T} \right\} d\omega, \quad (4)$$

где λ_k – новые, по сравнению с [10], множители Лагранжа, $k = 0, 1, \dots, (n-1) = \overline{0, (n-1)}$.

На основании неравенства Гельдера [8], максимум функционала (4) $\max_{S_x(j\omega)} J$ достигается при следующем спектре

сигнала на входе КС

$$S_x(j\omega) = \varepsilon \cdot K(-j\omega) \left\{ \frac{\sin \omega T / 2}{\omega T / 2} e^{-j\omega \frac{T}{2}} - \sum_{k=0}^{n-1} (-j\omega)^k \lambda_k e^{-j\omega T} \right\} = \varepsilon \cdot K(-j\omega) \Phi(-j\omega, T), \quad (5)$$

где $\varepsilon = E_x / y(T/2)$ – постоянная, определяемая из ограничения 1, и сигнальная функция вида

$$\Phi(-j\omega, T) = \frac{\sin \omega T / 2}{\omega T / 2} e^{-j\omega \frac{T}{2}} - \sum_{k=0}^{n-1} \lambda_k (-j\omega)^k e^{-j\omega T}. \quad (6)$$

Тогда спектр финитного сигнала $y(t)$, $0 \leq t \leq T$, на выходе КС будет равен

$$S_y(j\omega) = S_x(j\omega)K(j\omega) = \varepsilon K^2(\omega) \Phi(-j\omega, T), \quad (7)$$

При этом спектрам (5) и (7) соответствуют новые функции времени

$$x(t) = \varepsilon \left\{ \frac{1}{T} \int_0^T q(\tau - t) d\tau - \sum_{k=0}^{n-1} \lambda_k \frac{d^k q(T-t)}{dt^k} \right\}, \quad -\infty \leq t \leq T, \quad (8)$$

$$y(t) = \varepsilon \left\{ \frac{1}{T} \int_0^T G(\tau - t) d\tau - \sum_{k=0}^{n-1} \lambda_k \frac{d^k G(T-t)}{dt^k} \right\}, \quad -\infty \leq t \leq \infty, \quad (9)$$

где $q(t)$, $t \geq 0$ – импульсная реакция и $G(t)$ – канальная характеристика КС вида

$$G(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} K^2(\omega) e^{j\omega t} d\omega = \int_0^{\infty} q(\tau) q(\tau + |t|) d\tau, \quad -\infty \leq t \leq \infty. \quad (10)$$

Сигналы в (8) и (9) реализуют максимум функционала (4) при известных характеристиках линейного КС и найденных значений множителей Лагранжа.

**Новые финитные сигналы ОФС2.1,
согласованные с характеристиками фильтра Баттерворта**

Полагаем, что в модуляторе ТКС используется линейный фильтр Баттерворта с АЧХ вида:

$$K(\omega, p) = 1 / \sqrt{1 + \nu[\omega / \omega_\gamma]^2 p}, \quad (11)$$

где p - порядок фильтра, $F_\gamma = \omega_\gamma / 2\pi$ - граничная частота по уровню ослабления его АЧХ на L_γ дБ, ν - величина, зависящая от L_γ ; так при $L_\gamma = -10 \lg 2 \approx -3$ дБ, $\nu = 1$.

Из работы [8] для фильтра Баттерворта с АЧХ (11), имеем

$$q^{(k)}(t) = \frac{\omega_\gamma^{k+p/2}}{p} \sum_{l=1}^{p/2} q_l e^{-s_{F,l} t} \sin[c_{F,l} t + (k+1-p/2)(\pi/2 + \varphi_l)], \quad t \geq 0. \quad (12)$$

$$Q^{(k)}(t) = \frac{\omega_\gamma^{k+1}}{2p} \sum_{l=0}^7 e^{-s_{F,l} t} \sin[c_{F,l} t + \varphi_l + k(\pi/2 + \varphi_l)], \quad \omega_\gamma = 2\pi F_\gamma,$$

где $s_l = \omega_\gamma \sin \varphi_l$, $c_l = \omega_\gamma \cos \varphi_l$, $q_l = 1 / \left[c_l \prod_{i=1}^4 (s_i - s_l)_{i \neq l} \right]$, $\varphi_l = (2l-1)\pi / 2p$.

Для фильтра Баттерворта с импульсной реакцией (12) новые ОФС2.1 в (8) и (9) будут финитными на интервале $0 \leq t \leq T$, если множители Лагранжа λ_m , $m = 0, 1, \dots, (p-1)$, удовлетворяют новой, по сравнению с [10], системе линейных алгебраических уравнений:

$$\begin{cases} \frac{1}{\omega_\gamma^2 T} \left[\alpha_l \cos(\beta_l \frac{T}{2}) (e^{a_l \frac{T}{2}} - e^{-a_l \frac{T}{2}}) + \beta_l \sin(\beta_l \frac{T}{2}) (e^{a_l \frac{T}{2}} + e^{-a_l \frac{T}{2}}) \right] - \sum_{k=0}^{p-1} \lambda_k \omega_\gamma^k e^{-a_l \frac{T}{2}} \cos(\beta_l \frac{T}{2} - k(\pi/2 - \varphi_l)) = 0, \\ \frac{1}{\omega_\gamma^2 T} \left[\beta_l \cos(\beta_l \frac{T}{2}) (e^{a_l \frac{T}{2}} - e^{-a_l \frac{T}{2}}) - \alpha_l \sin(\beta_l \frac{T}{2}) (e^{a_l \frac{T}{2}} + e^{-a_l \frac{T}{2}}) \right] - \sum_{k=0}^{p-1} \lambda_k \omega_\gamma^k e^{-a_l \frac{T}{2}} \sin(\beta_l \frac{T}{2} - k(\pi/2 - \varphi_l)) = 0, \end{cases} \quad (13)$$

$$l = \overline{1, p/2}.$$

В выражении (13): $\omega_\gamma^2 = \alpha_l^2 + \beta_l^2$, $\alpha_l = \omega_\gamma \sin \varphi_l$, $\beta_l = \omega_\gamma \cos \varphi_l$, $\varphi_l = (2l-1)\pi / 2p$.

Итак, задача синтеза финитного сигнала $x(t)$, сохраняющего финитным отклик фильтра Баттерворта $y(t)$, и максимизирующего его среднее значение, решена.

Откуда следует, что при последовательной передаче ОФС2.1 по линейному частотно-ограниченному КС с АЧХ (11), отсутствует явление межсимвольной интерференции, а помехоустойчивость ТКС повышается и определяется только наличием шума наблюдения.

Пример расчета форм ОФС 2.1 и их спектров

Численные значения для множителей Лагранжа, при решении поставленной выше вариационной задачи на условный экстремум, можно получить, если в качестве модели линейного канала связи выбрать фильтр Баттерворта второго порядка.

Решая систему линейных уравнений второго порядка на основе (13), после достаточно сложных вычислений приходим к следующим новым формулам для множителей Лагранжа при синтезе ОФС 2.1.

$$\begin{aligned} \lambda_{2,0} &= [\exp(2b_T) \cdot \cos(2b_T) - 1] / 2b_T, \quad b_T = \omega_\gamma T / 2\sqrt{2}, \\ \lambda_{2,1} &= \frac{1}{\omega_\gamma^2 T} [\exp(2b_T) \cdot (\sin(2b_T) - \cos(2b_T)) + 1]. \end{aligned} \quad (14)$$

На основе полученных соотношений (5) – (10), (12) и (14), проведен машинный эксперимент в среде MatLab

(рис. 1). В расчетах принято $F_{\gamma} = 1,55$ кГц.

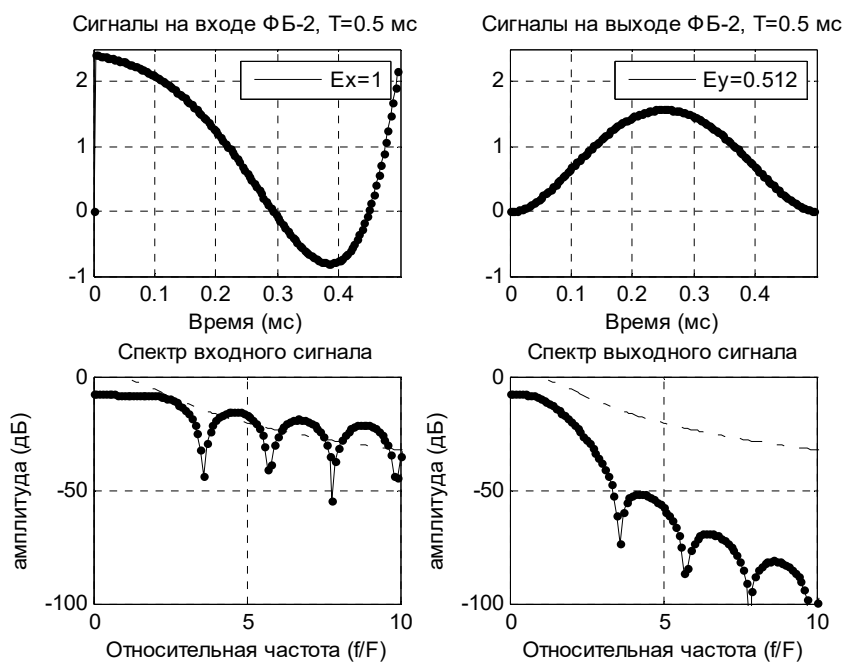


Рис. 1. Временные и частотные характеристики ОФС 2.1 для ФБ-2

Литература

1. Макаров С.Б., Цикин И.А. Передача дискретных сообщений по радиоканалам с ограниченной полосой частот. М.: Радио и связь, 1988. 304 с.
2. Николаев Б.И. Последовательная передача дискретных сообщений по непрерывным каналам с памятью. М: Радио и связь, 1988. 204 с.
3. Зяблов В.В., Коробков Д.Л., Портной С.Л. Высокоскоростная передача сообщений в реальных каналах. М.: Радио и связь, 1991. 288 с.
4. Скляр Бернанд. Цифровая связь. Теоретические основы и практическое применение, 2-е изд.: Пер с англ. М: Издательский дом «Вильямс», 2003. 1104 с.
5. Феер К. Беспроводная цифровая связь. Методы модуляции и расширения спектра: Пер. с англ. / Под ред. В.И. Журавлева. М: Радио и связь, 2000. 520 с.
6. Прокис Дж. Цифровая связь: Пер. с англ./ Под ред. Д.Д. Кловского. М.: Радио и связь, 2000. 800 с.
7. Ключев В.И. Частотно-временные преобразования и прием дискретных сигналов в системах связи. М: Радио и связь, 1990. 208 с.
8. Ямпольский Э.М. Вариационные принципы согласования сигналов с каналом связи. М: Радио и связь, 1987. 136 с.
9. Санников В.Г. Синтез финитных сигналов Найквиста, согласованных с телефонным каналом связи // Электро-связь, № 5, 2012. С. 9-12.
10. Санников В.Г., Алёшинцев А.В. Синтез финитных сигналов, согласованных с характеристиками фильтра Баттерворта, по критерию максимума среднего значения его отклика. Материалы 71-й Международной конференции: Радио-электронные устройства и системы для инфокоммуникационных технологий - «REDS-2016», посвященная Дню радио 25-27 мая 2016 г.

**SYNTHESIS OF NEW SIGNALS FOR THE BUTTERWORTH FILTER ACCORDING
TO THE CRITERION OF THE MAXIMUM AVERAGE VALUE OF ITS RESPONSE**

Anastasiya A. Kovalevskaya,
Student MTUCI, Moscow, Russia,
kanan88747@mail.ru

Vladimir G. Sannikov,
Associate Professor of the Department of GCT, Ph.D., MTUCI, Moscow, Russia,
v.g.sannikov@mtuci.ru

Abstract

The problem of digital data transmission via a linear communication channel without inter-character interference is considered. To solve this problem, the form of the transmitted data is coordinated with the time and frequency characteristics of the linear communication channel. For a linear communication channel, with a known pulse reaction and a complex transmission coefficient, the synthesis of new signals is carried out according to the criterion of the maximum average value of its response. New general ratios for optimal signals and new Lagrange multipliers applied to the channel model in the form of a Butterworth filter are obtained. The results of modeling synthesized signals for the second-order Butterworth filter are given.

Keywords: *linear communication channel, Butterworth filter, optimal finite signals, synthesis at the maximum of the average, leveling of intersymbolic interference.*

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ИЗОБРАЖЕНИЙ В ЗАДАЧАХ НЕФТЕГАЗОВОЙ ЛИТОЛОГИИ

Китаев Александр Павлович,
студент МГУСИ, Москва, Россия,
makavity231@gmail.com

Синева Ирина Сергеевна
доцент кафедры ТВиПМ, к.ф.-м.н., МГУСИ, Москва, Россия,
iss@mtuci.ru

Аннотация

Целью работы является применение нейронной сети в качестве инструмента для анализа высокоточных цифровых изображений петрографических прокрашенных шлифов, определение структуры пустотного пространства и получения характеристик породы. Рассмотрены сверточные нейронные сети с различными параметрами, предложена оптимальная архитектура нейронной сети для поставленной задачи, а также произведена оценка точности в используемых методах. Для обработки выходных данных нейронной сети разработаны алгоритмы на высокоуровневых языках программирования, которые позволяют определить границы произвольной фигуры и получить численную характеристику излома фигуры, которая потребуется для определения вида пористости.

Ключевые слова: анализ изображений, нейронные сети, метод стохастического градиентного спуска, марковские случайные поля, глубокое обучение, пустотное пространство, породы-коллекторы

Введение

Несмотря на то, что выявлению закономерностей распределения порового пространства в карбонатных породах-коллекторах посвящено большое количество работ, имеется ряд проблем, связанных с автоматизацией определения порового пространства для получения интегрированной цифровой модели пород-коллекторов. В настоящее время на территории Восточной Сибири открыто большое количество месторождений нефти и газа, залежи которых находятся в карбонатных породах-коллекторах [1]. Неоднородность распределения в них пустотного пространства снижает эффективность геологоразведческих работ. В связи с этим, предложены методы автоматизации процесса Image-анализа, которые направлены на ускорение анализа и минимизацию влияния человеческого фактора.

Image-анализ – сравнительно новая методика исследования горной породы. Основная цель image-анализа – изучение структуры пористого пространства для прогноза коллекторских свойств изучаемой породы. Для удобства, разобьем этот процесс на несколько стадий:

1. Микрофотосъемка шлифов - шлифы с изучаемой породой располагают под микроскопом с фотокамерой и делают снимок. Полученное изображение сохраняют, для дальнейшей обработки.
2. Обработка изображения - на полученном изображении, с помощью графического редактора выделяют пористое пространство горной породы, маркируя поры ярким цветом.
3. Статистический анализ изображения - подсчитываются числовые параметры обработанного изображения, площадь пор, диаметр и др. На их основании получают вторичную информацию о горной породе – среднюю площадь пор, средний диаметр и др.
4. Прогноз коллекторских свойств горной породы - обобщаются статистические и геологические данные об изучаемой породе на данном месторождении, делается заключение о ее коллекторских свойствах.

Постановка задачи

Рассмотрим цифровое высокоточное изображение петрографического прокрашенного шлифа (рис. 1).

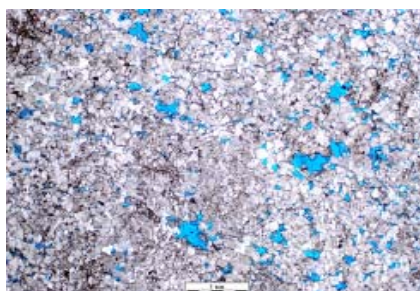


Рис. 1. Цифровое высокоточное изображение петрографического прокрашенного шлифа

На представленном изображении, можно наблюдать участки светло-голубого цвета. Они являются поровым пространством, в которое, в специальном оборудовании, под давлением, была залита прокрашенная смола. Данный снимок сделан с помощью микроскопа со встроенной камерой. Изображение представляет из себя петрографический прокрашенный шлиф, который является тонким спилом породы приклеенным на стекло (рисунок 2). Толщина такого спила исчисляется в сотых долях миллиметров (примерно 0.02-0.03 мм).

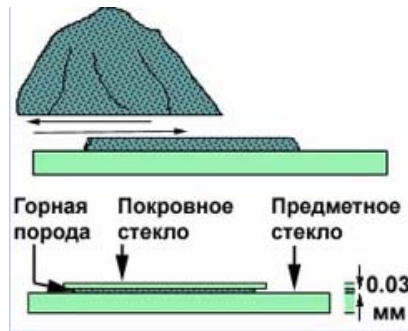


Рис. 2. Изготовление петрографического шлифа

Выделяют несколько основных видов пор, на основе которых будет строиться дальнейшая классификация. Здесь, пористости разделены на два типа: связанные со структурой (рис. 3) и не связанные со структурой (рис. 4).

Для изображений, схожих с представленным на рисунке 1, требуется:

1. Произвести разбиение изображения на два множества: пустотное пространство и порода. Для этого требуется точно выделить границы поры и присвоить пикселям, принадлежащим данной области, метку пустотного пространства. Для реализации данного пункта, предложено использование сверточной нейронной сети.
2. Для каждой поры получить численную характеристику излома ее границы. Для реализации этого предложен алгоритм, основанный на кликах второго порядка для Марковского случайного поля.



Рис. 3. Типы пористости связанные со структурой



Рис. 4. Типы пористости как связанные так и не связанные со структурой

В дальнейшем, на основе полученных данных будет построен более детальный анализ неоднородности параметров в задачах нефтегазовой литологии.

Подготовка датасета

Изначальным подходом в обработке изображений предполагался спектральный анализ с предварительным применением фильтра Гаусса, однако полученная точность (примерно 60% выделенных пор) и сложности с анализом изображения в его спектральном представлении, изменили выбор со спектрального анализа в пользу других алгоритмов. Например, в следующей итерации был выбран подход с увеличением контрастности для более четкого обозначения границ, имеющих другой цвет, и классификации пикселей по цвету. Однако, такой подход не учитывает размытые границы, что также не позволяет достигнуть высокой точности в решении поставленной задачи [2].

В связи с тем, что линейные алгоритмы не позволили достичь требуемой точности, выбор был сделан в пользу глубокого обучения, а именно - сверточной нейронной сети. Выбор сверточной нейронной сети обусловлен тем, что они нацелены на наиболее эффективное распознавание образов. В данном случае, распознаются образы порового пространства, а всё остальное обозначено как порода.

Для построения нейронной сети требуется описать входные данные и выходные данные и внутреннюю структуру сети. Вначале остановимся на входных данных.

Рассмотрим рисунок 1. Данное цифровое изображение имеет размерность 2560x1920 пикселей и содержит довольно большое количество требуемых объектов. Однако, для категоризации этих объектов, первоначально, требуется обучить нейронную сеть. Если в качестве входных данных для обучения использовать исходные изображения, то размер работы по подготовке набора данных будет весьма велик и потребует большое количество человеческих ресурсов. В связи с этим, было принято решение о фрагментации исходного изображения для формирования датасета. При таком подходе, выделение требуемых фрагментов на изображении не вызывает затруднений, а количество входных данных существенно возрастает.

Так как исходные изображения имеют одинаковый формат и одинаковую размерность, произведем фрагментацию изображений фрагментами, размерность которых кратна соответствующему размеру исходного изображения. В данном случае был выбран размер участков 64x64 пикселей и получено 1200 фрагментов

Для обучения нейронной сети требуется выделить области, которые будут определять нейронная сеть. В процессе изучения возможностей решения данной проблемы, был выбран ресурс [3], с помощью которого происходит выделение областей и дальнейшая генерация требуемых данных. Пример выделения требуемых областей на фрагменте изображения представлен на рисунке 5.

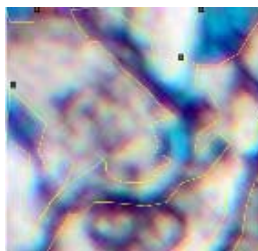


Рис. 5. Выделение областей на фрагменте изображения

На основе выделенных областей на фрагментах изображений, генерируются аннотации в формате JSON, которые содержат информацию о названии файла и координаты выделенных полигонов. Однако, форматом данных для нейронной сети являются изображения, в связи с чем, из полученных аннотаций требуется сгенерировать датасет.

В данной работе в качестве датасета было использовано 40 фрагментов изображений, на каждом из которых были выделены области, отвечающие пустотному пространству. Далее, с помощью программы, написанной на языке *Python* версии 3.9, с применением библиотеки компьютерного зрения *OpenCV* и математической библиотеки *NumPy* был сгенерирован датасет в пропорциях 80/20, состоящий из обучающих и проверочных данных соответственно. Данный датасет содержит исходный фрагмент изображения и соответствующую ему маску в градациях серого цвета, которая является выделенной порой на изображении.

Пример готового датасета представлен на рисунке 6. Получив готовый датасет для обучения, можно приступать к проектированию нейронной сети. Изучив инструменты для реализации нейронных сетей и алгоритмы машинного обучения, выбор был сделан в пользу математического пакета *Matlab* с дополнением *Deep Learning Toolbox*. И хотя библиотека для машинного обучения *Tensorflow*, из-за возможностей использовать несколько ядер процессора или использования нескольких ядер графического процессора является более производительным решением, выбор всё же был сделан в пользу простоты реализации и удобства пользования.



Рис. 6. Пример готового изображения и соответствующей ему маски

Архитектура нейронной сети

Рассмотрим входной слой нейронной сети. На входе должно обрабатываться цветное изображение размером 64x64 пикселей, следовательно, имеем 3 цветовых канала RGB, а значит размерность входного слоя составляет 64x64x3. Также, к входным значениям применяется нормализация. Используется метод 'zerocenter', который из всех изображений вычитает математическое ожидание выборки [4,5].

Следующими слоями будут выступать слои, отвечающие за даунсэмплинг. Здесь у нас будет 6 слоев со следующими характеристиками: первый слой - слой свертки с размером фильтров 2x2 и суммарным количеством в 50 единиц (слой свертки применяет расширение границ в 1 пиксель ко всем граням, иначе – паддинг); второй слой - слой *ReLU*, который будет обнулять отрицательные значения; третий слой - слой пулинга, или слой, выполняющий максимальное объединение, размером 2x2 и со страйдом, имеющим вертикальный и горизонтальный шаг равный трем. Эти слои повторяются дважды, в той же последовательности, и следовательно, имеем всего 6 слоев.

В связи с тем, что размерность входных данных была уменьшена в 4 раза, на следующем шаге потребуется восстановить исходную размерность. Этим будет заниматься группа слоев апсэмплинга, который является обратным процессом к даунсэмплингу. В качестве таких слоев рассмотрим транспонированный сверточный слой со следующими параметрами: размер фильтра возьмем равный 4x4, а количество фильтров, как и на первом шаге, будет равно 50. Вместе с тем, вертикальные и горизонтальные шаги страйда равны 2. И также, потребуется убрать данные, добавленные паддингом. Зададим параметр отсечения по границам, равный 1.

В качестве активационной функции будет выступать слой с *ReLU* функцией. Таким образом, в результате двойного применения данных слоёв, в качестве слоев апсэмплинга будет выступать последовательность транспонированного слоя свертки, слой *ReLU*, еще один транспонированный слой свертки, и также еще один слой *ReLU*. Итого – 4 слоя в группе апсэмплинг слоев.

И наконец, группой выходных слоев будут являться сверточный слой, с двумя фильтрами 1x1. Это требуется для линейной трансформации входных каналов с последующей нелинейностью. Далее, для получения вероятностей используется слой с функцией *Softmax*. Этот слой преобразует вектор в вектор той же размерности. Элементы полученного вектора являются вещественным числом, лежащим на отрезке 0 -1, и трактуются как вероятности принадлежности пикселя определенному классу.

Следующим шагом на основе вероятностей требуется промаркировать пиксель, в соответствие с максимальной вероятностью принадлежности определенному классу. Этим будет заниматься слой классификации пикселя.

Учитывая, что нейронная сеть оперирует пакетами данных, а именно - в выборке участвует не один элемент, а несколько, то для повышения производительности и стабилизации работы, требуется нормализовать данные: на вход слоя поступает вектор заданной размерности $n : x = (x^{(1)}, \dots, x^{(n)})$. Нормализуем исходный вектор по каждой размерности j , формула (1).

$$\hat{x}^{(j)} = \frac{x^{(j)} - Ex^{(j)}}{\sqrt{Dx^{(j)}}}, \quad (1)$$

где: $x^{(j)}$ – исходное значение вектора, $Ex^{(j)}$ – математическое ожидание всей обучающей выборки, $Dx^{(j)}$ – дисперсия всей обучающей выборки.

Учитывая, что нормализация выполнена по всей обучающей выборке, она может изменить представление данных внутри слоя. Во избежание данной проблемы, вводятся два параметра сдвига и сжатия нормализованной величины для каждого $x^{(k)} : \beta^{(k)}, \gamma^{(k)}$, их действие представлено формулой (2).

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)} \quad (2)$$

Параметры сдвига и сжатия подбираются в результате процесса обучения.

Однако, в случае обучения модели с помощью пакетов используется алгоритм пакетной нормализации, который можно описать следующим образом:

Пусть обучение некоторой модели выполняется с помощью пакетов B заданной размерности $m : B = \{x_1, \dots, x_m\}$. В этом случае, нормализация будет применена к каждому входному элементу с номером k по отдельности. Обозначим функцию нормализации формулой (3).

$$BN_{\gamma, \beta} : \{x_1, \dots, x_m\} \longrightarrow \{y_1, \dots, y_m\}, \quad (3)$$

На рисунке 7 представлен алгоритм пакетной нормализации.

На вход алгоритма подаются значения пакета, параметры сдвига и сжатия γ, β , а также константа ϵ , используемая для вычислительной устойчивости.

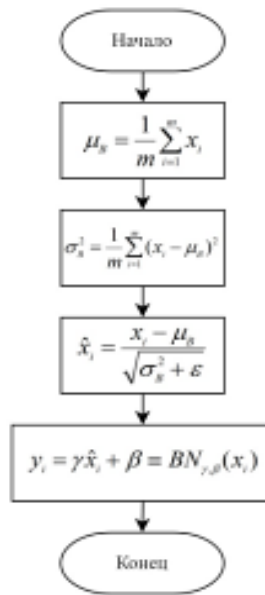


Рис. 7. Алгоритм пакетной нормализации

С учётом описанных параметров используемых слоев, архитектура слоев нейронной сети представлена в таблице 1.

Таблица 1

Архитектура слоев нейронной сети

Номер слоя	Назначение	Параметры
1	Входной слой	Размерность 64x64x3 с нормализацией по математическому ожиданию
2	Сверточный слой	50 фильтров размерности 2x2 с расширением всех граней на 1 пиксель
3	Активационная функция	<i>ReLU</i>
4	Пулинг	Пулинг 2x2 методом максимального объединения, со страйдом (2,2).
5	Сверточный слой	50 фильтров размерности 2x2 и расширением всех границ на 1 пиксель
6	Активационная функция	<i>ReLU</i>
7	Пулинг	Пулинг 2x2 методом максимального объединения, со страйдом (2,2).
8	Пакетная нормализация	См. рисунок 2.5
9	Активационная функция	<i>ReLU</i>
10	Транспонированный сверточный слой	50 фильтров размерности 4x4 со страйдом 2x2 и усечением всех границ на 1 пиксель.
11	Активационная функция	<i>ReLU</i>
12	Транспонированный сверточный слой	50 фильтров размерности 4x4 со страйдом 2x2 и усечением всех границ на 1 пиксель.
13	Активационная функция	<i>ReLU</i>
14	Активационная функция	<i>ReLU</i>
15	Сверточный слой	2 фильтра размерности 1x1 со страйдом (2,2).
16	<i>SoftMax</i>	Слой определения вероятности принадлежности классу
17	Выходной слой	Слой маркировки принадлежности пикселя определенному классу.

Следующим шагом требуется определить параметры обучения для нейронной сети. В процессе проверки работы нейронной сети, эмпирическим путем были подобраны следующие параметры: в качестве функции оптимизации выбран метод стохастического градиентного спуска с моментом, который описывается с помощью формулы (4).

$$\theta_{\ell+1} = \theta_{\ell} - \alpha \nabla E(\theta_{\ell}) + \gamma(\theta_{\ell} - \theta_{\ell-1}), \tag{4}$$

где: ℓ – номер итерации, α – параметр момента (здесь равен 0.9), γ – вклад предыдущего шага градиентного спуска в текущей итерации.

Добавление коэффициента α в метод стохастического градиентного спуска позволяет уменьшить колебания в процессе обучения.

Остальные параметры приведены в таблице 2.

Параметры обучения нейронной сети

Наименование параметра	Значение
Начальная скорость обучения	0.5
Коэффициент регуляризации L_2	0.0001
Количество эпох	150
Число итераций, между которым происходит проверка значений	30
Режим обучения	Обучение согласно плану с коэффициентом «Кусочно-заданной функции»
Перемешивание входных и валидационных данных	Каждую эпоху
Метод усечения градиента	По норме L_2
Граница градиента для усечения	0.05
Размер пакета для обучения и валидации	8

В таблице 2 приведены параметры, которые отвечают за усечение градиента и регуляризацию, где в качестве нормы используется норма L_2 , формула (5).

$$|x| = \sum_{i=1}^n \sqrt{|x_i|^2}, \tag{5}$$

где x – вектор размерности n : $x = [x_1 \ x_2 \ \dots \ x_n]^T$.

Усечение и нормализация градиента требуется для того, чтобы избежать большого увеличения в норме градиента в процессе обучения [6].

Пример такого выброса приведен на рисунке 8.

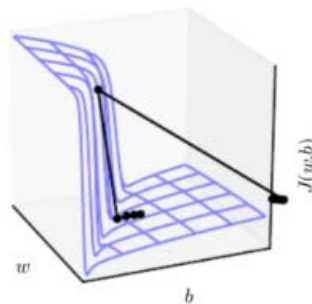


Рис. 8. Выброс при градиентном спуске

Как сказано выше, путем добавления отсечения, а именно – если норма градиента больше заданного значения, то считаем полученную норму равную верхней заданной границей, выбросов можно избежать.

Пример, приведенный на рисунке 8, с добавлением отсечения представлен на рисунке 9.

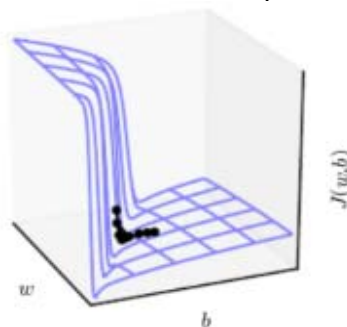


Рис. 9. Градиентный спуск с отсечением по норме

После описания всех параметров нейронной сети можно приступить к программной реализации. Она выполнена с помощью *Deep Learning Toolbox* дополнения для среды *Matlab*.

Обучение нейронной сети

Для анализа изображений после построения нейронной сети требуется обучение на заранее подготовленном датасете. Рассмотрим несколько вариантов обучения с различными параметрами. Изменяемыми параметрами будут вы-

ступать размер пакета, начальная скорость обучения и активационная функция [6,7]. Результаты обучения нейронной сети приведены на рисунках 10-14.

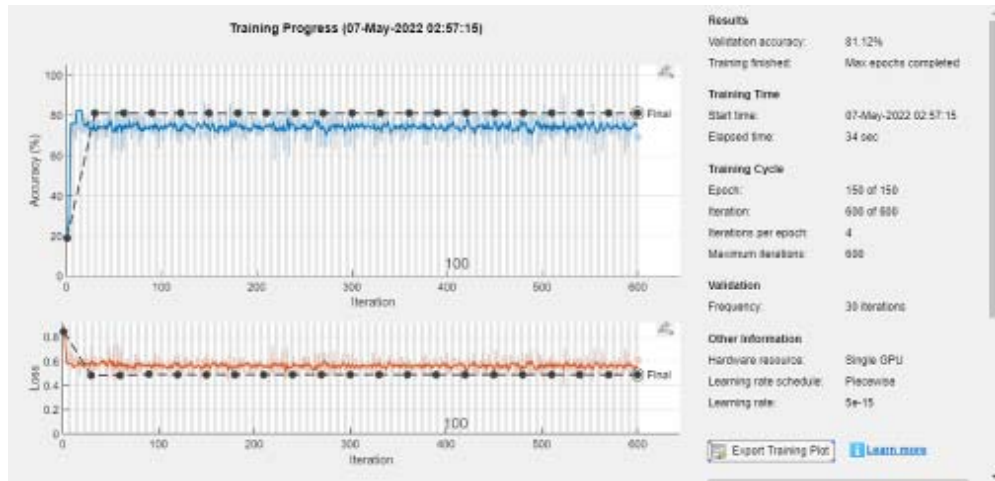


Рис. 10. Процесс обучения нейронной сети с сигмоидной активационной функцией

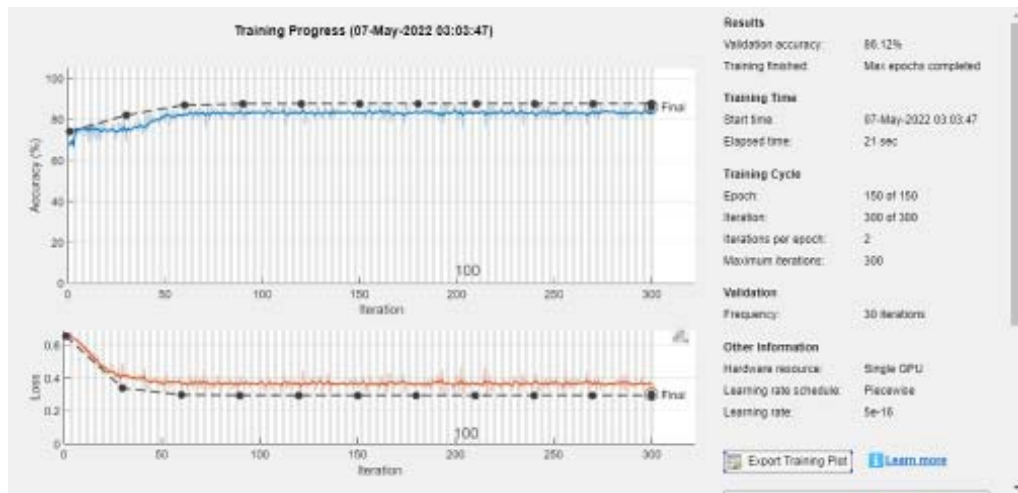


Рис. 11. Процесс обучения нейронной сети с размером пакета в 16 элементов

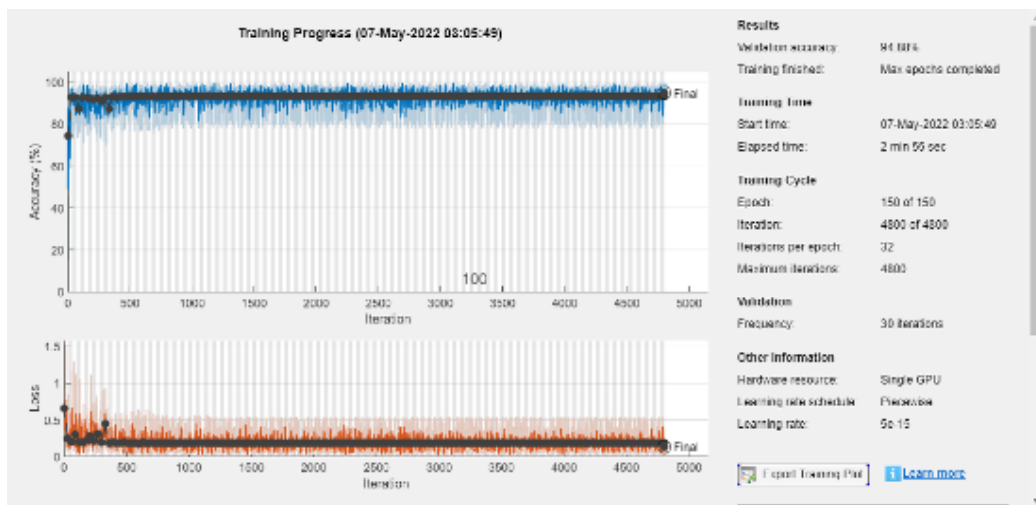


Рис. 12. Процесс обучения нейронной сети с размером пакета в 1 элемент

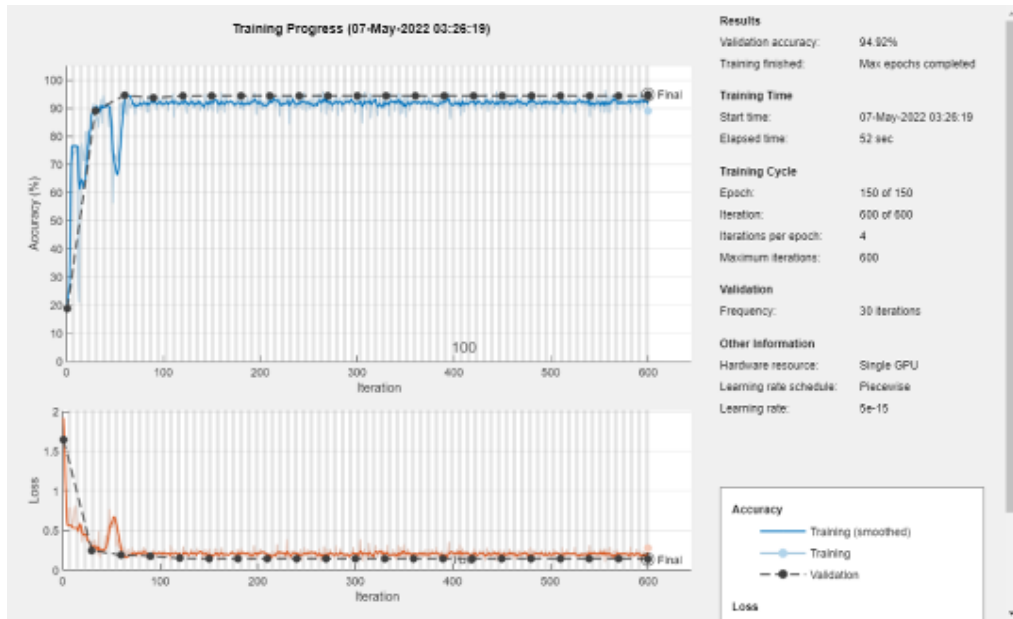


Рис. 13. Процесс обучения нейронной сети с активационной функцией *softplus*

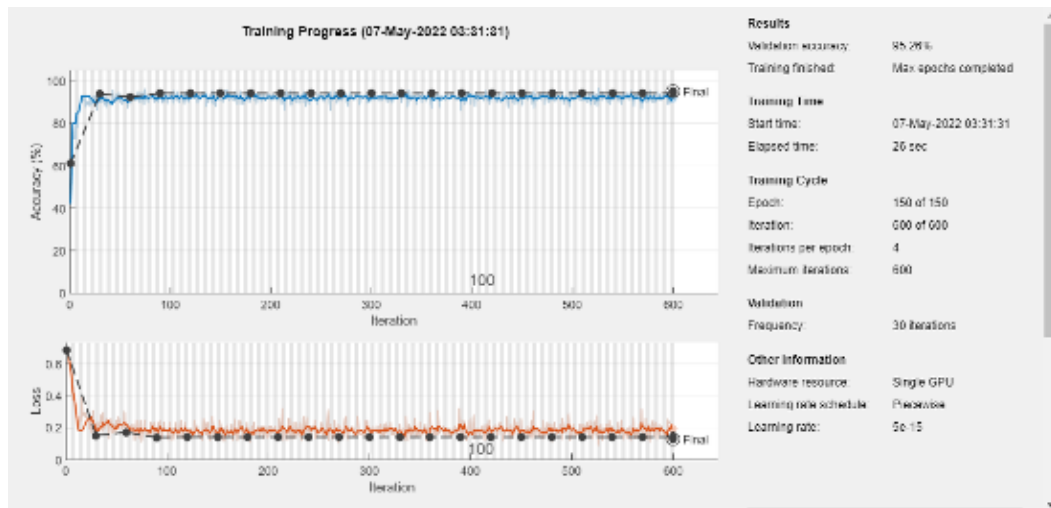


Рис. 14. Процесс обучения нейронной сети с параметрами, приведенными в таблице 2

Исходя из рисунков 10,11,12,13 и 14, максимальным полученным результатом является точность определения объектов равная 95.26%, которая получена для данной задачи с наилучшими подобранными параметрами, которые приведены в таблице 2. Пример этого процесса представлен на рисунке 14.

Рисунок 12 наглядно демонстрирует необходимость объединения данных в пакеты, вследствие чего существенно уменьшается число итераций, а, следовательно, и время, затрачиваемое на процесс обучения, а также объем вычислительных ресурсов. При этом, при отсутствии объединения данных в группы, точность, в среднем за 4800 итераций не превосходит точности, полученной при 600 итерациях.

Результаты классификации

После того как нейронная сеть обучена, на ее вход можно подать исходное изображение, представленное на рисунке 1. Результатом работы нейронной сети будет массив категорических переменных: пиксели соответствующие породы будут отмечены как “stone”, а пиксели, соответствующие поровому пространству, “void”, соответственно. Размерность выходного массива соответствует размерности исходного изображения в пикселях. Несколько примеров классификации изображения приведены на рисунках 15-17.

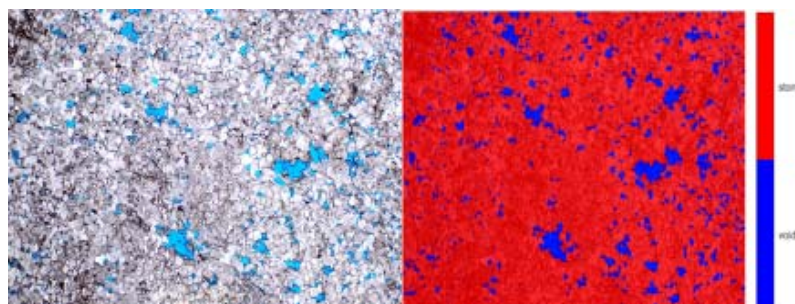


Рис. 15. Цветовая маска классификации для рисунка 1., коэффициент пористости равен 10.73%

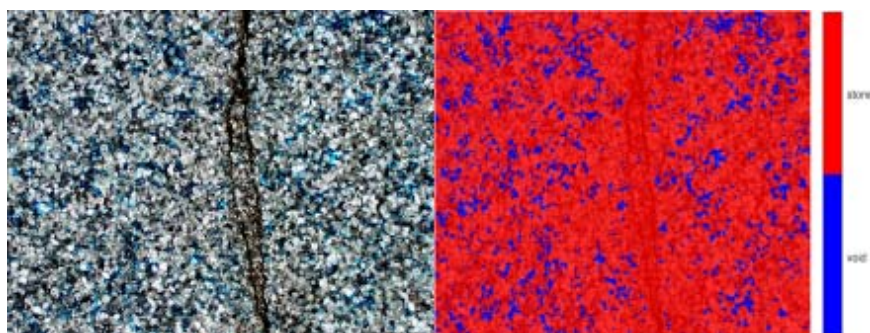


Рис. 16. Цифровой снимок шлифа и его цветовая маска, коэффициент пористости равен 14.66%

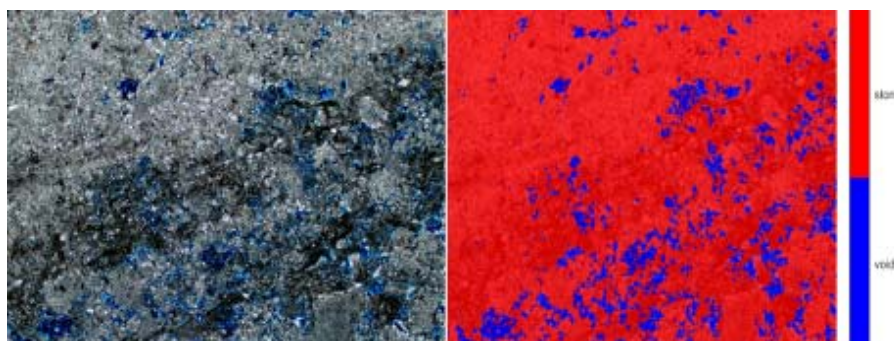


Рис. 17. Цифровой снимок шлифа и его цветовая маска, коэффициент пористости равен 13%

На данном этапе работы, с помощью сверточной нейронной сети была получена классификация цифрового высокоточного снимка петрографического прокрашенного шлифа с крайне высокой точностью. На рисунках можно наблюдать, что выделенное поровое пространство почти точно соответствует сопоставленному изображению. Для увеличения точности классификации, требуется рассмотреть обучающую выборку большей размерности, а также более детально выделить границы пустотного пространства. Ещё одним вариантом модификации может являться рассмотрение других цветовых диапазонов для обучающей выборки. Также, в связи с тем, что выделение границ для датасета на данном этапе разработки выполняется вручную, следует учитывать человеческий фактор.

На будущих стадиях разработки следует выделять границы на датасете с помощью машинных методов с последующей корректировкой человеком. При этом, возможно достижение точность близкой к 100%.

Анализ цветовой маски

Представленная на рисунках 15-17 цветовая маска связана с двумерным массивом из категорических переменных, которые характеризуют два возможных состояния пикселя: "void" – поровое пространство или "stone" – камень. Для дальнейшего анализа и классификации порового пространства, разобьём всё множество пор на отдельные участки и поставим им в соответствие числовую характеристику. Этой числовой характеристикой будет являться излом границ участка пустоты.

Для определения излома границы в двумерном массиве следует выделить границы каждого участка, сч использованием предложенного следующего алгоритма:

1. Найти координаты первого встретившегося пикселя с меткой "void".
2. Найти все соседние пиксели.
3. Если соседний пиксель является соседним к пикселю с меткой "stone", то шаг 4., иначе - переход к следующему пикселю.

4. Записать пиксель в список уже пройденных, добавить в маршрут.
5. Если пиксель следующий пиксель является началом маршрута – конец, иначе – переход к шагу 2 для текущего пикселя.

После отработки данного алгоритма, в рамках одного участка пустотного пространства получаем маршрут, состоящий из кортежа двух элементов: координаты пикселя, из которого был совершен переход, и координаты пикселя в который был совершен переход.

Дальнейшей задачей анализа является получение численной характеристики излома границ фигуры. Для этого, переход из одного пикселя в другой будем считать изломом, если он не соответствует переходу по следующим заданным направлениям, представленным на рисунке 18.

(x-1, y-1)	(x, y-1)	(x+1, y-1)		(x-1, y-1)	(x, y-1)	(x+1, y-1)
(x-1, y)	(x, y)	(x+1, y)		(x-1, y)	(x, y)	(x+1, y)
(x-1, y+1)	(x, y+1)	(x+1, y+1)		(x-1, y+1)	(x, y+1)	(x+1, y+1)
(x-1, y-1)	(x, y-1)	(x+1, y-1)		(x-1, y-1)	(x, y-1)	(x+1, y-1)
(x-1, y)	(x, y)	(x+1, y)		(x-1, y)	(x, y)	(x+1, y)
(x-1, y+1)	(x, y+1)	(x+1, y+1)		(x-1, y+1)	(x, y+1)	(x+1, y+1)

Рис. 18. Возможные переходы, не являющиеся изломом

С использованием вышеописанного подхода, выполняется проход по всему маршруту, а коэффициент считается следующим образом:

- Если переход выполнен в рамках возможных направлений – $k = k+1$
- Иначе - $k = k-1$

Так, например, для прямоугольника этот коэффициент равен 11, а для фигуры, у которой границы представлены в виде случайных переходов в форме "лесенки" – минус 4. Вопросом дальнейшего исследования является классификация типов пустот на изображении в соответствие с данной числовой характеристикой на основе статистических методов.

Заключение

По результатам данной работы для решения актуальной проблемы в сфере литологии были применены алгоритмы машинного обучения, в частности сверточные нейронные сети. Были проверены различные параметры для обучения и выбран их оптимальный набор. Данное исследование обеспечивает дальнейшую автоматизацию процесса *Image*-анализа, а именно – статистических исследований, и с учётом этого обеспечивает снижение временных затрат человека.

Литература

1. *Алексеев В.П.* Литология: учебное пособие. Екатеринбург: Изд-во УГГА, 2001. 249 с.
2. *Пушкарев А.В., Орлов В.Г.* Идентификационные параметры цифровых камер // Телекоммуникации и информационные технологии. 2014. Т. 1. № 1. С. 19-22
3. *Кузнецов В.Г.* Литология карбонатных пород-коллекторов: учебное пособие. М.: МИНГ, 1986. 81 с.
4. *Burtsev S.V., Kuzmin Ye.P.* An efficient flood-filling algorithm // Computers & Graphics. 1993. №5. С. 549-561.
5. *Хайкин С.* Нейронные сети: полный курс. - 2-е изд. - М.: ИД "Вильямс", 2006. 1104 с.
6. *Kindermann R., Snell J. Laurie* Markov Random Fields and Their Applications. Providence, USA: AMS, 2000. 147 с.
7. *Sineva I.S., Louniakova N.N.* Split method for assessing real images as Markov random field // Proceedings of SPIE - The International Society for Optical Engineering, 3457. San Diego: 1998. С. 64-81.

MACHINE LEARNING USAGE FOR IMAGE ANALYSIS IN OIL AND GAS LITHOLOGY PROBLEMS

Alexandr P. Kitaev,
Student MTUCI, Moscow, Russia,
makavity231@gmail.com

Irina S. Sineva,
Associate Professor of the Department of PT&AM, Ph.D. in Physics and Mathematics,
MTUCI, Moscow, Russia,
iss@mtuci.ru

Abstract

The aim of the work is to use a neural network as a tool for analyzing high-precision digital images of petrographic stained thin sections, determining the structure of void space and obtaining rock characteristics. Convolutional neural networks with different parameters are considered, the optimal architecture of the neural network for the task is proposed, and the accuracy of the methods used is estimated. To process the output data of the neural network, algorithms have been developed in high-level programming languages that allow you to determine the boundaries of an arbitrary figure and obtain a numerical characteristic of the figure break, which will be required to determine the type of porosity.

Keywords: *image analysis, neural networks, stochastic gradient descent, Markov random fields, deep learning, void space, reservoir rock*

РАЗРАБОТКА ПРОГРАММНЫХ МОДУЛЕЙ ФОРМИРОВАНИЯ И ОБРАБОТКИ СИГНАЛОВ С ЧАСТОТНОЙ МАНИПУЛЯЦИЕЙ ДЛЯ ДЕКАМЕТРОВОЙ РАДИОЛИНИИ

Демин Антон Димитриевич,
студент МТУСИ, Москва, Россия,
ademin42@mail.ru

Лобов Евгений Михайлович
доцент кафедры РТС, к.т.н., МТУСИ, Россия
e.m.lobov@mtuci.ru

Аннотация

Рассматривается принцип построения, а также измерения основных характеристик модема, использующего ЧМ-2 модуляцию без разрыва фазы, предназначенного для работы в декаметровом диапазоне радиоволн. Представленная в работе экспериментальная модель включает в себя реализованные блоки модулятора и демодулятора ЧМ-2, а также модель канала связи, имитирующая многолучевое распространение радиосигнала. Модель реализована с помощью системы имитационного моделирования радиотехнических систем «Спектр-2».

Ключевые слова: частотная манипуляция, ионосферный канал связи, «Спектр-2», частотная манипуляция без разрыва фазы, частотная манипуляция с непрерывной фазой, декаметровая радиолиния.

Введение

Декаметровая радиолиния – это радиолиния для передачи радиосигналов в диапазоне от 3 до 30 МГц, которые отражаются от ионосферы Земли и могут распространяться на расстояния, измеряемые тысячами километров. Декаметровые радиолинии активно используются военными, МЧС, в радиовещании и радиолюбителями. Основное достоинство данного частотного диапазона - минимум инфраструктуры, а значит дешевизна и простота организации. К недостатком этих радиолиний можно отнести узкие полосы (единицы - сотни кГц), малую скорость передачи информации (от бит/с до сотен кбит/с), высокую загруженность помехами, многолучевое распространение радиоволн, ухудшающее качество передачи информации [6-18].

Техническое задание

Цель работы – реализовать программные модули модулятора и демодулятора с частотной манипуляцией, которые будут представлять собой классы, реализованные на языке программирования C++. Блок модулятора должен формировать отсчеты комплексной огибающей (IQ-квадратуры) по заданной передаваемой информации для последующего переноса её в высокочастотный диапазон, преобразования в аналоговый вид, усиления и излучения. Блок демодулятора должен обрабатывать поступающие от радиоприемного устройства отсчеты комплексной огибающей принятого сигнала и осуществлять её демодуляцию в условиях наличия помех и многолучевости распространения. Для проверки работоспособности реализованных классов было решено использовать систему имитационного моделирования «Спектр-2». Для выполнения моделирования системы связи будут реализованы отдельные «блоки» модулятора и демодулятора, работающие с комплексными сигналами, и собрана испытательная схема, включающая в себя модель канала связи и различные вспомогательные блоки для получения необходимых характеристик работы реализованной системы связи.

Модулятор и демодулятор должны обеспечивать работу в двух режимах, различающихся значением символьной скорости SR и девиации частоты f_d .

Значения для первого режима («Режим 1»):

– $SR = 100$ симв/с;

– $f_d = 250$ Гц.

Параметры для второго режима («Режим 2»):

– $SR = 200$ симв/с;

– $f_d = 500$ Гц.

Для реализации предлагается использовать частотную манипуляцию без разрыва фазы: данное решение позволит улучшить спектральные характеристики сигнала при незначительных изменениях схемы работы модулятора.

Частотная манипуляция без разрыва фазы

Частотная манипуляция без разрыва фазы (англ. CPFSK) или частотная манипуляция с непрерывной фазой (ЧМНФ) – это частный случай частотной манипуляции, при котором фаза сигнала не имеет скачков на границе единичных импульсов, что качественно отражается на спектральных характеристиках: при одинаковых параметрах сигнала, амплитудный спектр ЧМНФ сигнала обладает меньшим уровнем боковых лепестков, а также большей скоростью их убывания. Амплитудный спектр сигналов ЧМ и ЧМНФ приведен на рисунке 1 [4].

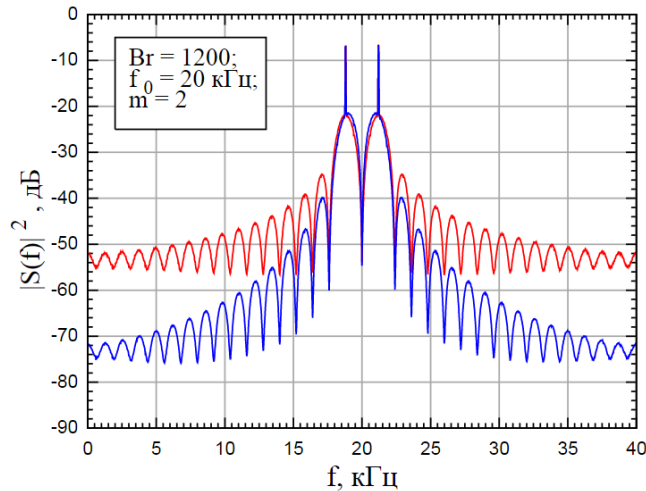


Рис. 1. Амплитудный спектр сигнала (с ЧМ- красный цвет; с ЧМНФ - синий цвет)

При данном виде манипуляции значение фазы сигнала пропорционально интегралу информационного сообщения. Тогда значение фазы сигнала можно определить, как:

$$b_{\phi}(t) = \omega_d \int_0^T b_{\text{норм}}(t) dt, \tag{1}$$

здесь: ω_d – девиация частоты в рад/с; $b_{\text{норм}}(t)$ – нормированный цифровой сигнал; T – длительность информационного сообщения. Величина изменения фазы на одном информационном символе называется набегом фазы и определяется, как:

$$\Delta\phi = \omega_d \int_0^T 1 \cdot dt = \omega_d \cdot \tau \tag{2}$$

Структура модулятора

Рассмотрим структурную схему модулятора ЧМНФ, формирующего отсчеты комплексной огибающей (рис. 2). На этом схеме: $b(t)$ – информационное сообщение; $b_{\text{норм}}(t)$ – нормированное информационное сообщение; $b_{\text{инт}}(t)$ – интеграл нормированного информационного сообщения. Усилитель с коэффициентом ω_d задает девиацию частоты сигнала, элементы $\cos()$ и $\sin()$ формируют синфазную ($I(t)$) и квадратурную ($Q(t)$) составляющую соответственно. Полученные отсчеты формируют комплексную огибающую $Z(t)$ и передаются для дальнейшей обработки. В системе «Спектр-2» полученные отсчеты комплексной огибающей сразу поступают на вход канала связи.

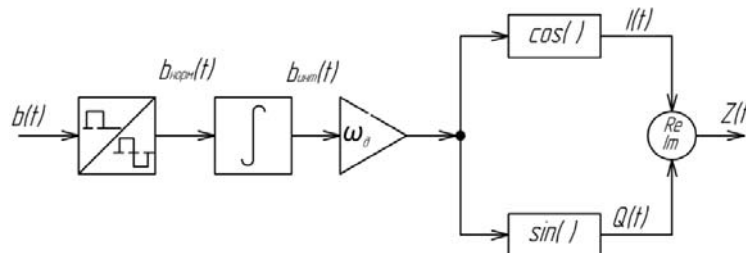


Рис. 2. Структурная схема модулятора ЧМНФ

Алгоритм работы элементов обработки информационного сообщения поясняется временными диаграммами на рисунке 3. На них продемонстрировано: поступление информационной последовательности на вход модулятора, её нормировка и интегрирование (здесь $A = \tau$).

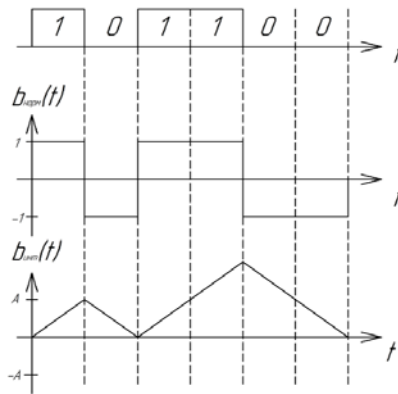


Рис. 3. Временные диаграммы работы модулятора

Структура демодулятора

Демодулятор ЧМНФ выполнен по схеме некогерентного приемника [2, 3], работающего с комплексными отсчетами. Его структурная схема приведена на рисунке 4.

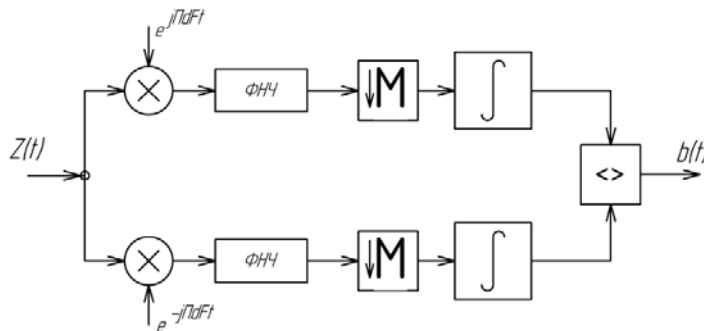


Рис. 4. Структурная схема демодулятора

Обработка полученных отсчетов комплексной огибающей производится параллельно в двух ветвях по следующему алгоритму: перенос комплексного сигнала на нулевую частоту; фильтрация сигнала; децимация; накопление отсчетов сигналов по длительности одного радиоимпульса и сравнение накопленных значений для принятия решения о принятом символе.

Так как демодулятор может работать в двух режимах, параметры ФНЧ и фильтра дециматора для них различаются. В качестве ФНЧ используется фильтр Чебышёва II рода с полосой пропускания, определяемой:

$$\Delta f = 1,2 \cdot SR \tag{3}$$

Тогда:

$$- SR = 100 \text{ симв/с, } \Delta f = 120 \text{ Гц;}$$

$$- SR = 200 \text{ симв/с, } \Delta f = 240 \text{ Гц.}$$

Коэффициент децимации определяется:

$$M = \frac{Fs}{40 \cdot SR} \tag{4}$$

Измеряемые характеристики

Основной метрикой систем связи является характеристика помехоустойчивости. Она определяет вероятность ошибки при передаче сообщения от значения отношения сигнал/шум (ОСШ, англ. SNR) в канале связи. Для получения данной характеристики необходимо измерять коэффициент битовой ошибки (BER) при различных значениях ОСШ. BER определяется, как:

$$BER = \frac{N_{\text{ош}}}{N_{\text{общ}}}, \tag{5}$$

здесь $N_{\text{ош}}$ – число ошибочно принятых бит, $N_{\text{общ}}$ – общее число переданных бит.

Для измерения величины ОСШ необходимо найти отношение мощности в полосе смеси сигнала с шумом, и чистого шума в канале связи, при условии, что полосы одинаковой ширины. Тогда:

$$SNR = \frac{C + \text{Ш}}{\text{Ш}} = \frac{C}{\text{Ш}} + 1 \tag{6}$$

Для определения величин SNR и BER в системе «Спектр-2» имеются соответственно готовые блоки: «Вычислитель ОСШ» и «BER автосинхронизирующийся».

Модели каналов связи

Создание точной математической модели реального канала связи – сложная задача. Поэтому применяются упрощенные модели, которые наиболее полно отражают физику распространения радиосигнала с учетом особенностей окружения. Существуют однолучевые и многолучевые каналы связи [3]. Разрабатываемые программные модули предназначены для применения в оборудовании КВ диапазона, что означает, что распространение радиоволн будет проходить по ионосферному каналу.

Ионосфера – это верхний слой атмосферы, ионизирующийся под воздействием на него солнечной радиации. Она представляет собой сильно разреженный газ, содержащий свободные электроны и ионы. В результате падения на ионосферу радиоволн, наблюдается преломление или отражение (частичное или полное) сигнала, сопровождаемое энергетическими потерями, что вызвано частичным поглощением сигнала. Данные процессы называются прохождением радиосигнала и зависят от окружающих условий.

Для моделирования ионосферного канала связи применяется рекомендация Международного союза электросвязи (МСЭ, англ. ITU) ITU-R F.1487. Данная рекомендация описывает стационарную модель канала Ваттерсона и её параметры в зависимости от широты распространения сигнала. Значения параметров данного канал (взаимной задержки лучей и доплеровского сдвига частоты) в зависимости от географической широты и состояния канала приведены в таблице 1 [1, 5].

Таблица 1

Параметры модели канала Ваттерсона

Широта	Условия	Взаимная задержка лучей, мс	Доплеровский сдвиг частоты, Гц
Низкие широты	Спокойные	0,5	0,5
	Умеренные	2	1,5
	Возбужденные	6	10
Средние широты	Спокойные	0,5	0,1
	Умеренные	1	0,5
	Возбужденные	2	1
	Возбужденные при почти вертикальном распространении сигнала	7	1
Высокие широты	Спокойные	1	0,5
	Умеренные	3	10
	Возбужденные	7	30

Функциональная схема модели канала Ваттерсона приведена на рисунке 5. Входной сигнал, представляющий собой комплексную огибающую, поступает на линию задержки, длина которой определяется количеством требуемых принимаемых лучей, после чего задержанные лучи умножаются на весовые коэффициенты и суммируются в выходной сигнал.

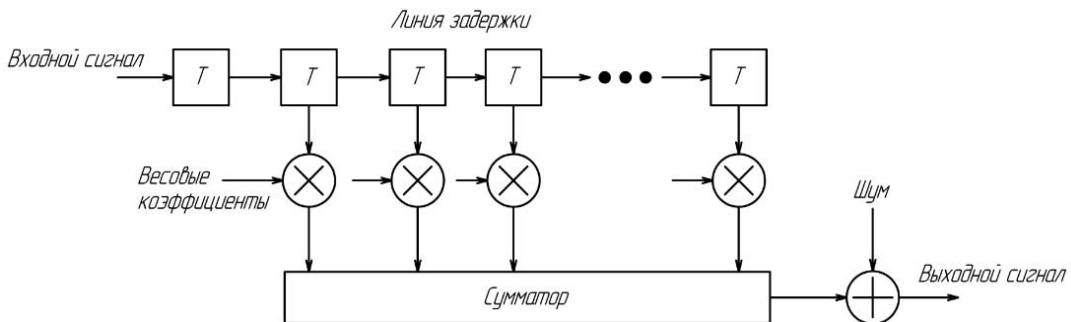


Рис. 5. Функциональная схема модели ионосферного канала связи

Имитационное моделирование

Модель системы связи с ЧМНФ, используемая при моделировании, приведена на рисунке 6. Она содержит:

- модулятор («ЧМ-2 мод») и демодулятор («ЧМ-2 демод») ЧМНФ;
- генератор белого гауссовского шума («БГШ») с усилителем и сумматором, имитирующий канал связи с аддитивным гауссовским шумом;
- модель многолучевого канала связи («РАДИОКАНАЛ»);
- блоки вывода («ТЕКСТ») переданного и принятого сообщения;
- вычислитель ОСШ («SNR ОСШ») с мультиметром для отображения значений;
- вычислитель BER («BER») с мультиметром для отображения значений;
- анализаторы спектра до и после канала связи.

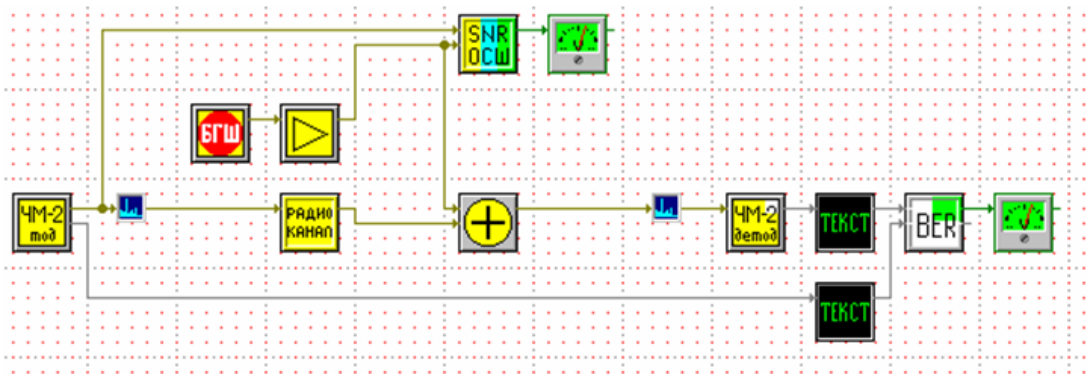


Рис. 6. Исследуемая модель системы связи с ЧМНФ

Диалоговые окна настройки блоков модулятора и демодулятор приведены на рисунке 7 и рисунке 8 соответственно. Блок модулятора позволяет пользователю выбрать используемый режим и источник входного сигнала, демодулятор позволяет выбрать только рабочий режим. При выборе режимов в диалоговом окне отображаются значения их параметров.

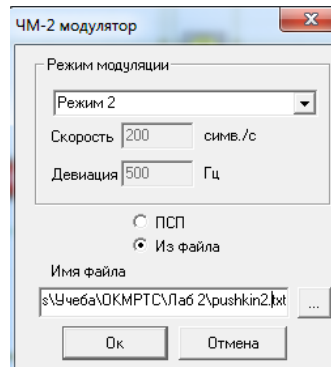


Рис. 7. Диалоговое окно настройки модулятора

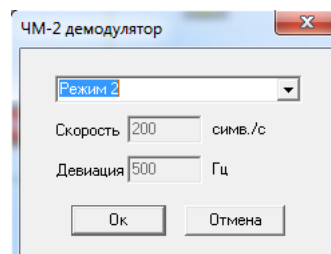


Рис. 8. Диалоговое окно настройки демодулятора

В качестве результатов работы модели на рисунке 9 приведены спектрограммы сигнала до и после прохождения канала связи, а на рисунке 12 сравниваются переданное и принятое информационное сообщение. На рисунках 10 и 11 приведено измерение девиации частоты и символьной скорости соответственно (символьная скорость численно определяется как половина ширины основного лепестка амплитудного спектра ЧМНФ сигнала). Измерения проводились в режиме работы «Режим 2», измеренные значения указаны в поле «Полосы».

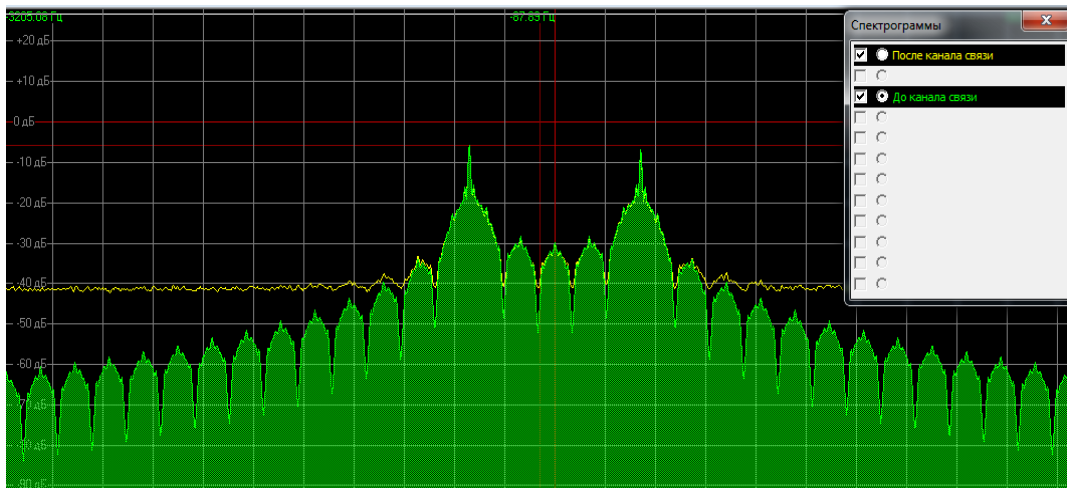


Рис. 9. Спектрограммы сигнала до и после прохождения канала связи

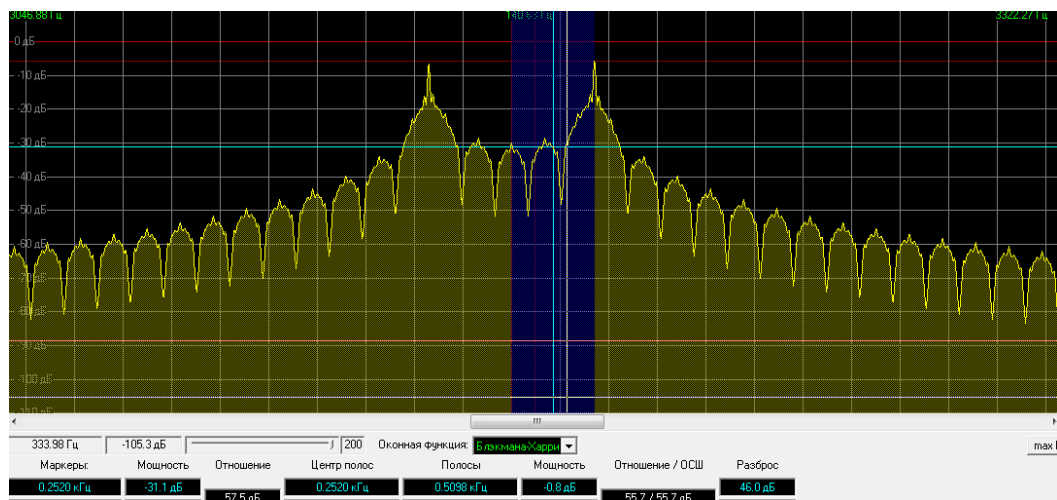


Рис. 10. Измерение величины девиации фазы по спектрограмме

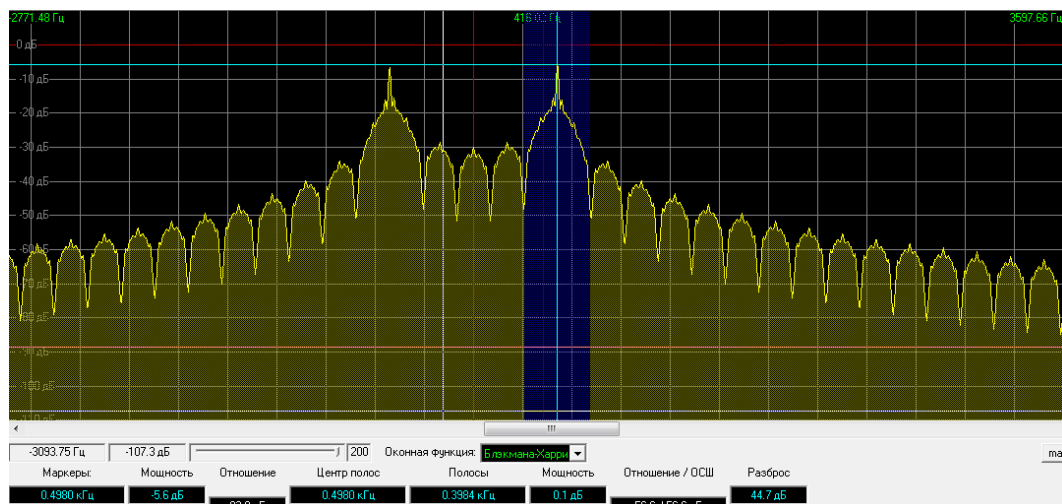


Рис. 11. Измерение ширины основного лепестка сигнала по спектрограмме

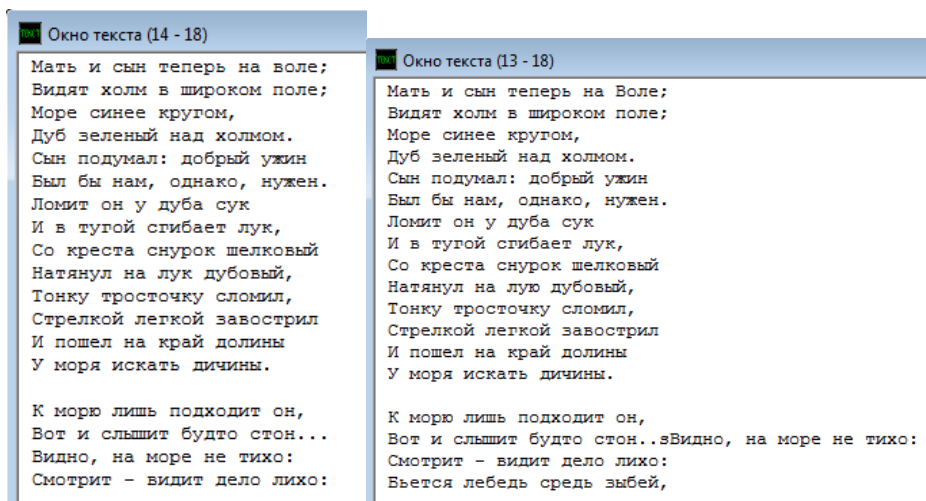


Рис. 12. Сравнение переданного (слева) и принятого (справа) сообщения

Заключение

В работе были рассмотрены основные принципы построения и работы системы связи, использующей ЧМНФ. Получен практический результат в виде реализованных блоков связи с помощью имитационного моделирования и системы «Спектр-2». Результаты моделирования подтвердили работоспособность и соответствие техническим требованиям программно реализованных модулей модулятора и демодулятора с частотной манипуляцией.

Литература

1. *ITU-R F.1487*. Testing of HF modems with bandwidths of up to about 12 kHz using ionospheric channel simulators, 2000, 13 p.
2. *Скляр Б.* Цифровая связь. Теоретические основы и практическое применение. Перевод с англ. М.: Издательский дом "Вильямс", 2003. 1104 с.
3. *Прокис Дж.* Цифровая связь: Пер. с англ. М.: Радио и связь, 2000. 800 с.
4. Частотно-манипулированные сигналы FSK (frequency shift key) и FSK сигналы с непрерывной фазой CPFSK (continuous phase FSK) // DSPLib.org URL: https://ru.dsplib.org/content/signal_fsk/signal_fsk.html (дата обращения: 26.04.2022).
5. *Головкин И.В.* Разработка и исследование сигнально-кодовых конструкций на основе турбо-подобных кодов и дискретных частотных сигналов для систем коротковолновой связи: дис. канд. Радиотех. наук: 05.12.04. М., 2016. 152 с.
6. *Lobov E.M., Priputin V.S., Lobova E.O., Liberman E.A.* Dispersion distortion tracking compensator based on the sigma-point kalman // T-Comm. 2021. Т. 15. № 1. С. 41-47.
7. *Аджемов С.С., Лобов Е.М., Кандауров Н.А., Лобова Е.О., Липаткин В.И.* Алгоритмы оценки и компенсации дисперсионных искажений широкополосных сигналов ионосферных радиолиний связи // Научные технологии в космических исследованиях Земли. 2021. Т. 13. № 5. С. 57-74.
8. *Кучумов А.А., Лобов Е.М., Варламов В.О.* Обзор технологии модуляции сигнала в ортогональном частотно-временном пространстве // Системы синхронизации, формирования и обработки сигналов. 2021. Т. 12. № 6. С. 41-49.
9. *Lobov E.M., Shubin D.N., Kandaurov N.A., Lobova E.O.* A new class of polyphase pseudo-random sequences with a nonlinear generation algorithm for secure radio communication link between unmanned vehicles and dispatch centers of the "smart city" // T-Comm. 2020. Т. 14. № 8. С. 42-46.
10. *Платонов О.И., Лобов Е.М.* Исследование гибридных банков фильтров в задачах цифровой обработки сверхширокополосных сигналов // Телекоммуникации и информационные технологии. 2019. Т. 6. № 1. С. 18-23.
11. *Голубев Е.А., Лобов Е.М., Кандауров Н.А., Шубин Д.Н.* Новый класс двоичных псевдослучайных последовательностей с нелинейным алгоритмом формирования для систем связи с кодовым разделением абонентов // T-Comm: Телекоммуникации и транспорт. 2018. Т. 2. С. 76-80.
12. *Лобов Е.М., Лобова Е.О., Курочкин А.А.* Вычислительно упрощенная реализация компенсатора дисперсионных искажений на базе банка цифровых фильтров // Системы синхронизации, формирования и обработки сигналов. 2018. Т. 9. № 1. С. 106-110.
13. *Лобов Е.М., Лобова Е.О., Елсуков Б.А.* Устройство компенсации дисперсионных искажений широкополосных сигналов на базе банка цифровых фильтров // DSPA: Вопросы применения цифровой обработки сигналов. 2018. Т. 8. № 3. С. 60-65.
14. *Чиров Д.С., Лобов Е.М.* Выбор сигнально-кодовой конструкции для командно-телеметрической линии радиосвязи с беспилотными летательными аппаратами средней и большой дальности // T-Comm: Телекоммуникации и транспорт. 2017. Т. 11. № 10. С. 21-28.

15. Лобов Е.М., Смердова Е.О. Оценка импульсной характеристики широкополосного ионосферного канала с помощью метода обратной фильтрации // Т-Comm: Телекоммуникации и транспорт. 2017. Т. 11. № 3. С. 24-27.
16. Лобов Е.М., Смердова Е.О. Алгоритмы оценки и фильтрации наклона дисперсионной характеристики широкополосного ионосферного канала // Системы синхронизации, формирования и обработки сигналов. 2017. Т. 8. № 1. С. 68-72.
17. Толкачев М.М., Лобов Е.М., Кандауров Н.А. Разработка и оценка помехоустойчивости алгоритмов итеративной обработки М-ЧТ сигналов в ионосферном канале с узкополосными помехами // Системы синхронизации, формирования и обработки сигналов. 2017. Т. 8. № 2. С. 68-73.
18. Лобов Е.М., Смердова Е.О., Терешонок М.В. Комплекс лабораторных работ по дисциплине "цифровая обработка сигналов" для студентов направления 11.03.01 и 11.03.02 на базе системы имитационного моделирования "Спектр-2" // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2017. Т. 6. № 2. С. 9-15.

**DEVELOPMENT OF SOFTWARE MODULES FOR GENERATING AND PROCESSING SIGNALS
WITH FREQUENCY MANIPULATION FOR DECAMETER RADIO LINES**

Anton D. Demin,
Student MTUCI, Moscow, Russia,
ademin42@mail.ru

Evgeny M. Lobov,
Associate Professor of the Department of RTS, Ph.D., MTUCI, Moscow, Russia,
e.m.lobov@mtuci.ru

Abstract

The principle of construction and measurement of the main characteristics of a modem using FM-2 modulation without phase discontinuity, designed to operate in the decameter range of radio waves, is considered. The experimental model considered in this paper includes the implemented blocks of the FM-2 modulator and demodulator, as well as a communication channel model simulating multipath propagation of a radio signal. The model is implemented using the system of simulation of radio engineering systems "Spectr-2".

Keywords: *Frequency manipulation, ionospheric communication channel, "Spectr-2", frequency manipulation without phase discontinuity, frequency manipulation with continuous phase, decameter radio line.*

МОДЕЛИРОВАНИЕ ВОСХОДЯЩЕЙ ЛИНИИ СВЯЗИ СИСТЕМЫ WSMA

*Литвяков Вячеслав Сергеевич,
магистрант МТУСИ, Москва, Россия,
slavalit32@mail.ru*

*Доронина Анжела Дмитриевна,
студент МТУСИ, Москва, Россия
doronina.angela@mail.ru*

*Панкратов Денис Юрьевич
доцент кафедры СиСРТ, к.т.н., МТУСИ, Россия
dpankr@mail.ru*

Аннотация

Ключевой технологией будущих систем мобильной связи 6G является неортогональный множественный доступ (NOMA, Non-Orthogonal Multiple Access). Одной из распространенных разновидностей NOMA является схема множественного доступа на основе последовательностей Велча (WSMA, Welch-bound equality spread multiple access), которая позволяет обеспечить качественные услуги беспроводной связи одновременно большому количеству абонентов. В данной статье рассматривается методика моделирования восходящей линии связи системы WSMA, получены характеристики помехоустойчивости для разного числа абонентов, а также рассмотрены особенности кодовых последовательностей Велча.

Ключевые слова: системы мобильной связи, 5G, 6G, NOMA, WSMA, последовательности Велча, моделирование систем радиосвязи.

Введение

В системах мобильной связи предыдущих поколений использовались схемы ортогонального множественного доступа (англ. OMA): в первом поколении (1G) применялся множественный доступ с частотным разделением каналов (англ. FDMA), во втором поколении (2G) – множественный доступ с временным разделением каналов (англ. TDMA), в третьем поколении (3G) – множественный доступ с кодовым разделением каналов (англ. CDMA), и в четвертом и пятом поколении (4G и 5G) – множественный доступ с ортогональным частотным разделением (англ. OFDMA) [1-3, 22-27]. В технологии NOMA (Non-Orthogonal Multiple Access) неортогональные ресурсы используются несколькими пользователями одновременно (системы после 5G и 6G), что обеспечивает высокую спектральную эффективность, а также возможность одновременной работы большого числа абонентов по сравнению с традиционными системами OMA [4, 5, 20].

Главным образом, первоначально предложенные схемы NOMA делятся на два типа: с кодовым разделением и с разделением по мощности. При разделении абонентов по мощности для того, чтобы достигнуть высокую производительность системы, разным пользователям в зависимости от характеристик их каналов назначаются разные коэффициенты распределения мощности [1, 6]. Преимуществом этой разновидности NOMA является гибкость при совмещении с такими технологиями, как MIMO (Multiple Input Multiple Output) и Beamforming [7, 8].

В рамках стандартизации 3GPP было предложено множество разновидностей NOMA с кодовым разделением [9, 10, 11]. При мультиплексировании с кодовым разделением, разным пользователям назначаются разные кодовые последовательности. С помощью кодовых последовательностей пользователи мультиплексируются по одним и тем же частотно-временным ресурсам, например с помощью схем NOMA кодами низкой плотности (англ. LDS), с разреженными кодами (англ. SCMA) и схемы многопользовательского совместного доступа с комплексными кодовыми последовательностями (англ. MUSA) [1, 2, 9]. Следует также отметить, что кроме мультиплексирования с кодовым разделением и мультиплексирования с разделением по мощности, существуют такие схемы NOMA, как множественный доступ с разделением по шаблонам (англ. PDMA) и мультиплексирование с битовым разделением (англ. BDM) [1, 2, 13]. Одной из наиболее перспективных схем NOMA по количеству упоминаний в публикациях считается схема множественного с расширением на основе последовательностей Велча (англ. WSMA) [2, 9, 10, 11, 12].

Для схем OMA основной проблемой является низкая спектральная эффективность, когда пользователям с плохими условиями канала выделяются ресурсы систем мобильной связи (код/частота/время) [1, 2, 5]. В схемах NOMA один и тот же ресурсный блок совместно используется пользователями с разными условиями канала. Технология NOMA может в значительной степени улучшить спектральную эффективность и плотность пользователей будущих сетей мобильной связи и планируется для использования в системах связи после 5G [4, 5, 11]. Следует отметить, что NOMA

в системах 6G рассматривается как одна из технологий, позволяющих удовлетворить требования низкой задержки, высокой скорости, плотному покрытию и широким возможностям подключения [14].

Основными целями представленной работы, наряду с изучением методов обработки сигналов в системы WSMA с учетом потенциальных возможностей NOMA, являются: разработка методики моделирования системы NOMA с кодовым разделением на восходящей линии системы мобильной связи и получение характеристик помехоустойчивости системы WSMA для разного числа абонентов и разных значений отношения сигнал/шум (англ. SNR).

Структурная схема и модель системы WSMA

На рисунке 1 изображена структурная схема системы WSMA, каждый функциональный блок осуществляет соответствующее преобразование данных, которое можно описать с помощью математических соотношений, т. е. формального описания [9, 11, 12]. На этом рисунке иллюстрируется сценарий восходящей линии связи (от абонентских станций к базовой станции), когда данные передаются одновременно четырьмя абонентами.

Рассмотрим схему рисунка 1 системы WSMA для случая 4-х абонентов и восходящей линии связи и запишем с помощью математических формул и обозначений преобразования, которые осуществляются в каждом функциональном блоке. Для каждого абонента генерируются биты b_1, b_2, b_3, b_4 со значениями из множества $\{0,1\}$, которые поступают на вход модуляторов, где они отображаются в информационные символы s_1, s_2, s_3, s_4 со значениями из множества $\{-1,1\}$. Далее символы подаются на вход модуляторов WSMA. В модуляторах WSMA символы умножаются на векторы кодовых последовательностей Велча $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4$. После этого символы передаются по радиоканалу с аддитивным белым гауссовским шумом \mathbf{n} , каждый отсчет которого имеет нулевое среднее и дисперсию σ_n^2 . Релеевские замирания учитываются с помощью комплексных коэффициентов h_1, h_2, h_3, h_4 .

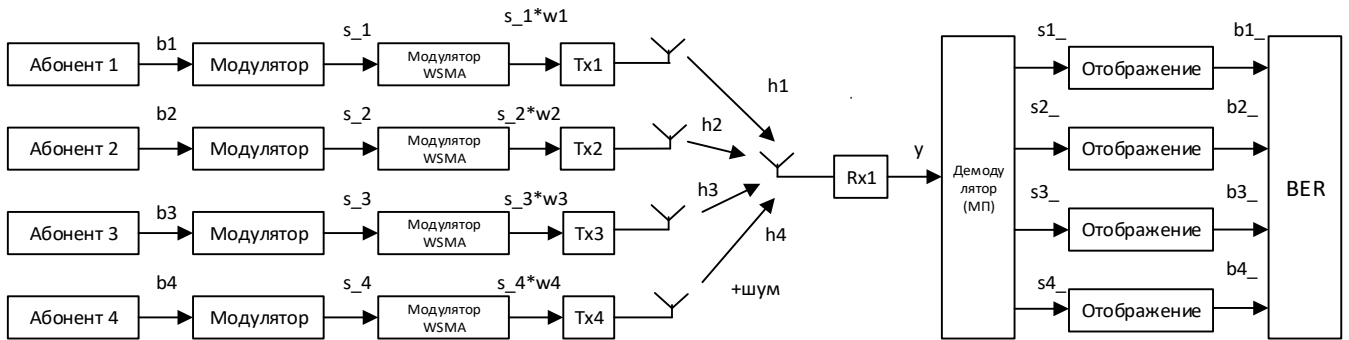


Рис. 1. Структурная схема системы WSMA

Рассмотрим модель системы NOMA с технологией WSMA на восходящей линии связи. На вход демодулятора базовой станции поступает сигнал:

$$\mathbf{y} = h_1 \mathbf{w}_1 s_1 + h_2 \mathbf{w}_2 s_2 + h_3 \mathbf{w}_3 s_3 + h_4 \mathbf{w}_4 s_4 + \mathbf{n}, \quad (1)$$

где $\mathbf{y} = [y_1 \ y_2 \ y_3 \ y_4]^T$ – вектор отсчетов принимаемого сигнала, h_k – комплексный коэффициент передачи k-го абонента, \mathbf{S}_k – информационный символ k-го абонента, $\mathbf{w}_k = [w_{1,k} \ w_{2,k} \ w_{3,k} \ w_{4,k}]^T$ – вектор-столбец с элементами кодовой последовательности k-го абонента, которые задаются согласно таблице кодовых последовательностей WSMA для базы 4 (всего таких последовательностей 8 – см. таблицу из документа 3GPP далее), $\mathbf{n} = [n_1 \ n_2 \ n_3 \ n_4]^T$ – вектор-столбец отсчетов гауссовского шума в канале связи [9, 11].

Коды WSMA для различного числа абонентов K можно записать в матрицу \mathbf{W}_{WSMA} размера $4 \times K$, ее подробная структура задается согласно документу 3GPP [9]. Для 4 абонентов матрица \mathbf{W}_{WSMA} будет иметь следующий вид:

$$\mathbf{W}_{WSMA} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \mathbf{w}_3 \ \mathbf{w}_4] = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \end{bmatrix}^T. \quad (2)$$

Демодуляция сигнала WSMA осуществляется по алгоритму, оптимальному по критерию максимального правдоподобия (англ. ML):

1. Начало цикла перебора комбинаций информационных символов;
2. Вычисление реплики и нормы для данной комбинации;
3. Завершение цикла перебора;
4. Нахождение минимальной нормы и ее индекса;
5. Выбор из матрицы комбинаций, комбинации с найденным индексом – это и будет искомый вектор оценок ML переданных информационных символов.

После демодуляции оценки информационных символов отображаются в биты и поступают в блок вычисления коэффициента битовых ошибок (англ. BER).

Для удобства запишем модель принимаемого сигнала WSMA в матричном виде:

$$\mathbf{y} = \mathbf{H} \odot \mathbf{W}_{WSMA} \mathbf{s} + \mathbf{n}, \quad (3)$$

где $\mathbf{y} = [y_1 \ y_2 \ y_3 \ y_4]^T$ – принимаемый вектор размерностью 4×1 . $\mathbf{W}_{WSMA} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \mathbf{w}_3 \ \mathbf{w}_4]$ – матрица последовательностей Велча размерностью 4×4 (приведены в таблице 1), $\mathbf{s} = [s_1 \ s_2 \ s_3 \ s_4]^T$ вектор информационных символов абонентов размерностью 4×1 [11,12]. При этом матрица \mathbf{H} коэффициентов передачи канала для 4 абонентов размерностью 4×4 имеет вид:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \\ h_{41} & h_{42} & h_{43} & h_{44} \end{bmatrix}. \quad (4)$$

Кодовые последовательности системы WSMA

В таблице 1 приведен пример 8 кодовых последовательностей системы WSMA с базой 4, при этом максимальное число активных абонентов составляет 8 [9, 11]. Эти последовательности основаны на выполнении так называемого равенства границе Велча и называются WBE-последовательностями [12].

Таблица 1

Кодовые последовательностей WSMA для 8 абонентов (база сигналов – 4)

	Последовательность	1	2	3	4
Номер элемента (чипа)	1	-0.6617 + 0.1004i	-0.0912 + 0.4191i	0.4151 - 0.3329i	0.2736 - 0.4366i
	2	0.0953 + 0.4784i	-0.4246 - 0.0859i	0.2554 - 0.3140i	0.5452 + 0.2068i
	3	-0.4233 - 0.1399i	-0.4782 + 0.3752i	-0.3808 - 0.1569i	-0.4690 - 0.2225i
	4	-0.1265 + 0.3153i	0.4936 + 0.1233i	0.6130 - 0.0873i	-0.3399 + 0.0974i
	Последовательность	5	6	7	8
Номер элемента (чипа)	1	-0.4727 - 0.1234i	-0.3413 + 0.1257i	0.4216 + 0.1187i	0.4603 + 0.2142i
	2	0.0592 - 0.6432i	0.3671 - 0.1430i	-0.0241 - 0.5620i	0.0048 - 0.4244i
	3	0.3493 - 0.1988i	0.6514 - 0.0660i	-0.4507 + 0.0958i	0.4047 + 0.1601i
	4	-0.0975 - 0.4161i	0.2174 + 0.4864i	-0.5167 + 0.1116i	-0.4908 + 0.3629i

Следует отметить, что преимущества системы WSMA обусловлены особыми свойствами кодовых последовательностей – так называемых последовательностей Велча, примеры которых приведены в таблице 1. Формирование этих последовательностей осуществляется с помощью специального алгоритма на основе параметра TSC (Total Squared Correlation), учитывающего общую корреляцию последовательностей, а также на основе границы Велча (англ. WBE) [11, 12]. Алгоритм формирования WBE-последовательностей описан в приложении документа 3GPP [9].

Значение параметра TSC ограничено снизу границей Велча и для оптимальной производительности системы WSMA в определенных условиях требуется, чтобы выполнялось равенство WBE, при этом одновременно оптимизируются такие показатели, как емкость и взаимные помехи в системе. Также может применяться оптимизация для получения низкой корреляции или между WBE-последовательностями, чтобы уменьшить сложность приемника, не снижая производительности системы WSMA. Для этого применяется подход, который состоит в том, чтобы обеспечить ортогональность между определенным подмножеством векторов в данном наборе и выполнение равенства границе Велча. Говорят, что такой набор векторов формирует кодовую книгу WSMA на основе упаковки подпространств, а эти векторы также удовлетворяют границе WBE [9]. Однако не всегда удается уменьшить сложность приемника за счет структуры кодовых последовательностей NOMA (например, как в схеме SCMA), тогда необходимо использовать алгоритмы демодуляции с приемлемой вычислительной сложностью и искать компромисс между характеристиками BER и сложностью обработки групповых сигналов NOMA [17, 18, 19, 21].

Алгоритм и результаты моделирования системы WSMA

Рассмотрим алгоритм моделирования системы WSMA по шагам, по которому будет составлена программа и проведено моделирование в среде Matlab для случая K абонентов с использованием теории компьютерного моделирования систем радиосвязи [3, 15, 16].

1. Ввод исходных данных.
2. Циклы по отношению сигнал/шум (англ. SNR) и по числу испытаний:
 - количество ресурсных блоков $N=4$;
 - количество пользователей K (максимально 8);
 - диапазон SNR от 0 дБ до 20 дБ с шагом 2 дБ;
 - число испытаний 10 000.
3. Формирование матрицы кодовых последовательностей (берем столбцы из таблицы 1) для разного числа абонентов.
4. Формирование случайных битовых последовательностей.
5. Модуляция BPSK (QPSK).
6. Умножение символов абонентов на кодовые последовательности и формирование эквивалентной матрицы коэффициентов передачи канала с элементами $w_{lk}h_k$.
7. Формирование наблюдаемого вектора на входе демодулятора.
8. Вычисление оценки вектора переданных символов с использованием алгоритма ML.
9. Отображение оценки вектора символов в биты.
10. Завершение цикла по числу испытаний
11. Вычисление коэффициента ошибок (англ. BER).
12. Завершение цикла по ОСШ
13. Вывод кривых помехоустойчивости $BER = f(SNR)$ для разного числа абонентов.

В результате моделирования по указанному алгоритму в среде Matlab был получен график помехоустойчивости для восходящей линии системы WSMA для разного числа абонентов (1, 2, 4, 6, 8) – зависимости BER от параметра SNR в дБ.

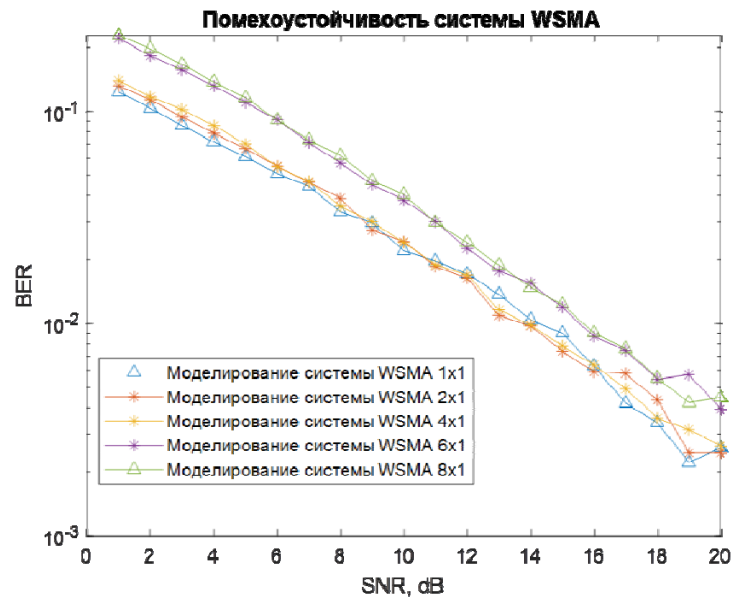


Рис. 2. Помехоустойчивость системы WSMA с разным числом абонентов

Анализируя рисунок 2, можно сделать вывод о том, что при увеличении числа абонентов системы WSMA, помехоустойчивость снижается незначительно. Для числа абонентов от 1 до 4, характеристики BER почти не отличаются, а для случаев 6 и 8 абонентов характеристики BER смещаются приблизительно на 2 дБ. Случай 6 и 8 абонентов соответствуют загрузке системы WSMA 150% и 200%, соответственно. Следует отметить, что преимущество системы WSMA по сравнению с традиционной системой с кодовым разделением сигналов (англ. CDMA) состоит в том, что она предназначена для работы в перегруженном режиме, когда число абонентов существенно превышает базу сигналов.

Заключение

В данной работе представлена подробная методика моделирования восходящей линии системы NOMA с технологией WSMA. Также были проиллюстрированы концепции и проблемы систем NOMA на основе анализа зарубежных работ. Полученные в результате моделирования характеристики помехоустойчивости системы WSMA свидетельствуют о том, что данная схема NOMA может использоваться на восходящей линии в перегруженном режиме работы. При этом потери в помехоустойчивости составляют примерно 2 дБ (при максимальном числе абонентов – 8). Это достигается за счет применения комплексных кодовых последовательностей на основе последовательностей Велча. Однако применение таких сигналов усложняет обработку на приемной стороне.

Литература

1. *Fa-Long Luo et al.* Signal Processing for 5G: Algorithms and Implementations. Wiley-IEEE Press, 2016.
2. *L. Dai, B. Wang, Y. Yuan, S. Han, C. I. I and Z. Wang.* "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," IEEE Commun. Mag., vol. 53, no. 9, pp. 74-81, Sep. 2015.
3. *Бакулин М.Г., Крейнделин В.Б., Панкратов Д.Ю.* Технологии в системах радиосвязи на пути к 5G. М.: Горячая линия – Телеком, 2018.
4. *M.Kamel, W.Hamouda, and A.Youssef,* "Ultra-dense networks: a survey," IEEE Communications Surveys & Tutorials, vol.18, no. 4, pp. 2522-2545, 2016.
5. *Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu.* "5G network capacity: Key elements and technologies," IEEE Vehicular Technology Magazine, vol. 9, no.1, pp. 71-78, 2014.
6. *S. M. R. Islam, N. Avazov, O. A. Dobre and K. Kwak.* "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges," in IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 721-742.
7. *J. Choi.* "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," IEEE Trans. Commun., vol. 63, no. 3, pp. 791-800, Mar. 2015.
8. *B. Wang, L. Dai, Z. Wang, N. Ge and S. Zhou.* "Spectrum and energy efficient beamspace MIMO-NOMA for millimeter-wave communications using lens antenna array," IEEE J. Sel. Areas Commun., vol. 35, no. 10, pp. 2370-2382, Oct. 2017.
9. 3GPP TR 38.821 V16.0.0, Published by Alliance for Telecommunications Industry Solutions, 2019.
10. *Yifei Yuan, Zhifeng Yuan, Li Tian.* 5G Non-Orthogonal Multiple Access Study in 3GPP // IEEE Communications Magazine, Vol. 58, № 7, pp. 90-96, July 2020.
11. *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? // Т-Comm: Телекоммуникации и транспорт. 2022. Т. 16. № 1. С. 4-14.
12. R1-1805003, "Signature Design for NoMA", 3GPP TSG RAN WG1 Meeting, Ericsson, Sanya, China, 16-20 April, 2018.
13. *S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun and K. Niu.* "Pattern division multiple access a novel nonorthogonal multiple access for fifth-generation radio networks," IEEE Trans. Veh. Technol., vol. 66, no. 4, pp. 3185-3196, Apr. 2017.
14. *Alsharif, M.H.; Kelechi, A.H.; Albreem, M.A.; Chaudhry, S.A.; Zia, M.S.; Kim, S.* Sixth Generation (6G) Wireless Networks: Vision, Research Activities, Challenges and Potential Solutions. Symmetry 2020, 12, 676.
15. *Борисов Ю.П., Цветнов В.В.* Математическое моделирование радиотехнических систем и устройств М.: Радио и связь, 1985.
16. *Панкратов Д.Ю., Степанова А.Г.* Компьютерное моделирование технологии MIMO для систем радиосвязи // Т-Comm: Телекоммуникации и транспорт. 2018. Т. 12. № 12. С. 33-37.
17. *H. Nikopour and H. Baligh.* "Sparse code multiple access," 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2013, pp. 332-336.
18. *K. Higuchi and Y. Kishiyama.* "Non-orthogonal access with successive interference cancellation for future radio access," in Proceedings of APWCS, 2012, Kyoto, Japan, 2012.
19. *S. Verdu, Multiuser Detection.* Cambridge University Press, New York, NY, USA, 1st edition, 1998.
20. *Литвяков В.С., Орлов В.Г.* Эволюция и тенденции совершенствования радиоинтерфейсов в сетях GERAN/UTRAN/E/UTRAN // Телекоммуникации и информационные технологии. 2020. Т. 7. № 1. С. 26-32.
21. *Бакулин М.Г., Крейнделин В.Б., Панкратов Д.Ю., Степанова А.Г.* Новый подход к задачам MIMO-детектирования и многопользовательской демодуляции. Информационные процессы, 2021. Т. 21. № 2. С. 93-107.
22. *Крейнделин В.Б., Смирнов А.Э., Бен Режеб Т.Б.К.* Эффективность методов обработки сигналов в системах MU-MIMO высоких порядков // Т-Comm: Телекоммуникации и транспорт. 2016. Т. 10. № 12. С. 24-30.
23. *Крейнделин В.Б., Старовойтов М.Ю.* Повышение помехоустойчивости системы связи MIMO с пространственным мультиплексированием методом додетекторного сложения // Т-Comm: Телекоммуникации и транспорт. 2017. Т. 11. № 4. С. 4-13.
24. *Крейнделин В.Б., Усачев В.А.* LTE-advanced pro как основа для новых сценариев M2M // Т-Comm: Телекоммуникации и транспорт. 2017. Т. 11. № 3. С. 28-32.
25. *Бакулин М.Г., Крейнделин В.Б.* Проблема повышения спектральной эффективности и емкости в перспективных системах связи 6G // Т-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 2. С. 25-31.
26. *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.
27. *Крейнделин В.Б., Панкратов Д.Ю.* Вероятностная модель радиоканала MIMO с учетом взаимной корреляции передающей и приемной сторон // REDS: Телекоммуникационные устройства и системы. 2016. Т. 6. № 1. С. 103-107.

SIMULATION OF THE UPLINK OF WSMA SYSTEM

Vyacheslav S. Litvyakov,
Graduate, MTUCI, Moscow, Russia,
slavalit32@mail.ru

Angela D. Doronina,
Student MTUCI, Moscow, Russia,
doronina.angela@mail.ru

Denis Y. Pankratov,
Associate Professor of the Department of SiSRT, Ph.D., MTUCI, Russia,
dpankr@mail.ru

Abstract

The key technology of future mobile communication systems 6G is Non-Orthogonal Multiple Access (NOMA). One of the most popular varieties of code domain NOMA is the Welch-bound equality spread multiple access (WSMA) scheme, which allows to provide high-quality wireless communication services simultaneously to a large number of subscribers. This article discusses a method for uplink simulation of WSMA system, obtains noise immunity characteristics for a different number of subscribers, and also considers the features of Welch code sequences.

Keywords: mobile communication systems, 5G, 6G, NOMA, WSMA, Welch sequences, computer simulation of radio communication systems.

«Сетевые технологии и системы телекоммуникаций»

ПЕРСПЕКТИВЫ РАЗВИТИЯ СИСТЕМ ЛАЗЕРНОЙ СВЯЗИ В АТМОСФЕРЕ И ГИДРОСФЕРЕ

*Андреев Степан Андреевич,
магистрант МТУСИ, Москва, Россия,
stefan_4@mail.ru*

*Казанцев Сергей Юрьевич,
профессор кафедры НТС, д.ф.-м.н., МТУСИ, Москва, Россия,
s.i.kazantsev@mtuci.ru*

*Свистунова Анастасия Игоревна,
магистрант МТУСИ, Москва, Россия,
nastya.svistunova.99@bk.ru*

*Титовец Павел Александрович,
доцент кафедры ТЭДиА, к.т.н., МТУСИ, Москва, Россия,
paveltitovec@yandex.ru*

Аннотация

Рассмотрены основные особенности передачи информации по лазерному лучу в оптике свободного пространства. Представлен анализ научно-технической литературы в области связи беспроводных оптических коммуникаций в атмосфере и под водой. Рассмотрены основные направления развития систем лазерной связи в воздухе и на воде. Определены точки роста технологий беспроводной лазерной связи в атмосфере и гидросфере. Определены группы стран, наиболее активно развивающие технологии беспроводной лазерной связи в оптике свободного пространства и основные источники финансирования работ в этом направлении. Выявлены точки роста этих технологий, определен список организаций-лидеров в области исследований по оптической связи в открытом пространстве и организаций, осуществляющих финансирование этого направления. Составлен список изданий, в которых публикуются результаты исследования по тематике АОЛС и БОПС.

***Ключевые слова:** атмосферные оптические линии связи, наукометрический анализ, беспроводная оптическая подводная связь, квантовое распределение ключей, коммуникации в свободном пространстве.*

Передача информации в открытом пространстве

В настоящее время наибольшие объем и скорость передачи информации в воздушной или водной среде реализуется при использовании оптических систем связи [1-3]. Высокие частоты электромагнитных волн оптического и инфракрасного диапазона спектра определяют огромную емкость этого канала связи. В отличие от акустических волн скорость электромагнитных волн много выше, а использование лазерных передатчиков с высокой концентрацией энергии в малом угле позволяет осуществлять максимально адресную передачу данных от источника к приемнику. В англоязычной литературе для оптической связи в открытом пространстве используется термин Free-Space optics (FSO) - оптика свободного пространства, при этом под открытым пространством подразумевается: атмосфера, подводная среда и космическое пространство [1].

Целью настоящей работы являлся анализ информации, публикуемой в области оптической связи в атмосфере Земли и гидросфере, а также оценка уровня технологической зрелости этих технологий в разных странах.

Для создания эффективного канала связи необходимо, чтобы длина волны источника излучения попадала в окна прозрачности соответствующей среды, кроме того, окружающая среда предъявляет и другие ограничения на применяемое оборудование [2]. Поэтому, развитие атмосферных линий связи (АОЛС) и беспроводной оптической связи (БОПС) анализировались раздельно.

1. Атмосферная оптическая связь

Окна прозрачности, представленные на рисунке 1, обусловлены взаимодействием электромагнитного излучения разных длин волн с разными слоями атмосферы [4]. Рентгеновские и гамма-лучи достигают только 30-40 км, а основная часть УФ-диапазона поглощается озоновым слоем на высоте около 30 км. Первое узкое окно прозрачности пропускает видимый свет с примыкающими к нему УФ и ближним ИК-диапазоном, при котором атмосфера имеет несколько узких окон благодаря механизмам поглощения излучения молекулами воды и углекислого газа. Большая часть ИК-диапазона поглощается водяным паром, углекислым газом и кислородом нижних слоев атмосферы. Затем на

шкале длин волн от 1 см до 30 м возникает широкое радиопрозрачное окно. Более длинные радиоволны отражаются ионосферой. Именно на этих окнах прозрачности ведется передача информации [4].

Сегодня, скорость передачи данных, обеспечиваемая современными терминалами АОЛС, выпускаемыми российскими компаниями достигает 40 Гбит/с [5,6]. Кроме того, атмосферная оптическая связь позволяет использовать большое количество каналов за счет отсутствия помех между различными каналами передачи данных. Поэтому интерес к технологии АОЛС растет, чему способствует совершенствование технологий передачи, обработки и приёма оптического сигнала [2,7]. Вызывает интерес оценка уровня развития этой технологии, который можно получить из наукометрического анализа публикационной активности [8]. Для определения перспектив развития систем АОЛС и точек роста данной технологии связи использовалась также методика, описанная в работах [9,10]. С этой целью проводился поиск в БД SCOPUS по ключевым словам **атмосферная оптическая связь («atmospheric optical communication»)**, в результате которого была получена выборка научных публикаций, индексируемых в этой базе за временной период с 2001 года по 2021 год.

1.1 Динамика мирового потока научно-технической литературы по тематике атмосферно оптической связи за период 2001-2021 гг.

На рисунке 2 показана зависимость числа публикаций, удовлетворяющих поисковому запросу: TITLE-ABS-KEY (atmospheric AND optical AND communication) AND PUBYEAR > 2000 AND PUBYEAR < 2022. Всего в выборку попало 5494 публикации. Как видно из рисунка 2 на временном отрезке с 2001 по 2021 годы наблюдается заметный рост интереса к данной области исследований, при этом, несмотря на то, что на графике видны небольшие колебания, общая тенденция отражает рост числа научных публикаций. Следует заметить, что на этом графике учитывались только публикации в научных журналах и трудах конференций, патенты не учитывались.

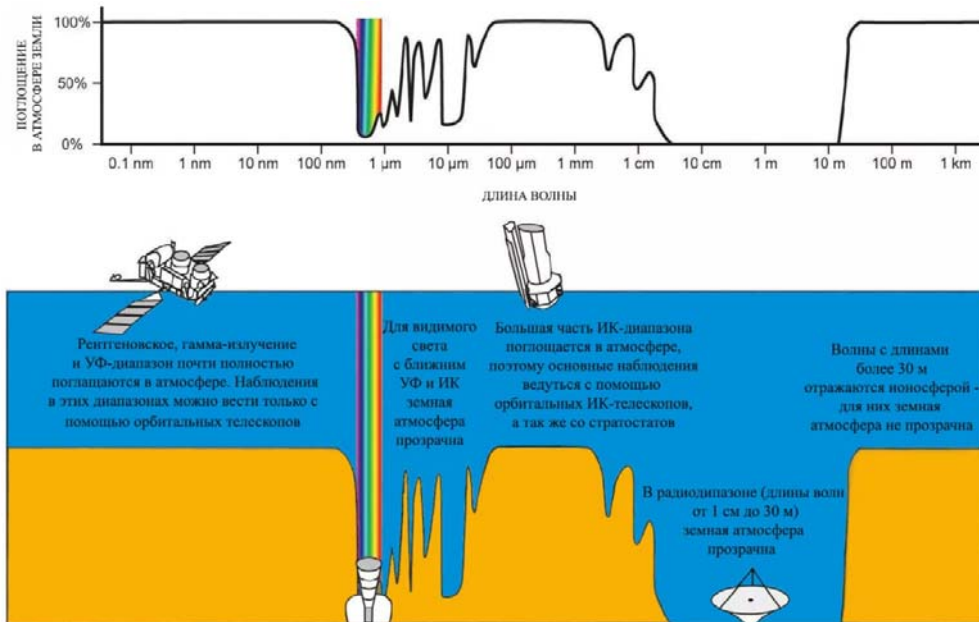


Рис. 8. Окна прозрачности атмосферы

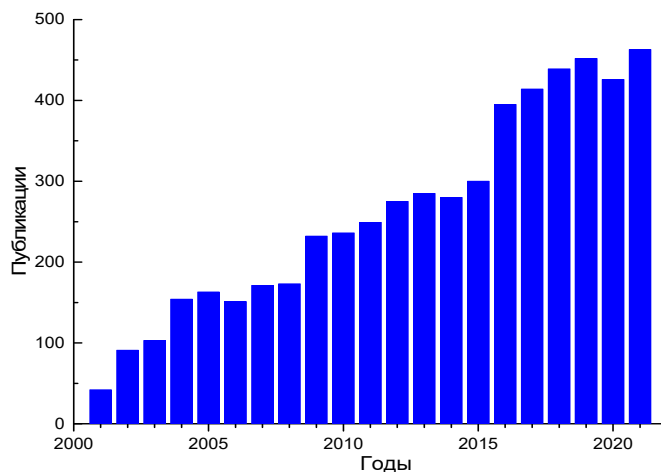


Рис. 9. Динамика публикационной активности по тематике АОЛС

1.2. Основные источники, в которых опубликованы результаты исследований атмосферно оптической связи

В таблице 1 представлен список 10 изданий, в которых было больше всего опубликовано статей по атмосферной оптической связи за 2017-2021 годы. В этой таблице указано число публикаций в журнале и его квартиль по БД SCOPUS и WoS по состоянию на 2020 год.

Таблица 3

Журналы, в которых за 2017–2021 годы опубликовано наибольшее число статей по атмосферной оптической связи

Название журнала	Число статей за 2017–2021 годы	Квартиль журнала (2020 г.) Scopus/WoS
Proceedings of SPIE	270	Без квартиля
Optics Communications	131	Q2/Q3
Optics Express	64	Q1/Q1
Journal of Optical Communications	57	Q3/не входит
Applied Optics	45	Q1/Q3
IEEE Photonics Journal	44	Q1/Q2
Optik	43	Q2/Q2
IEEE Access	38	Q1/Q2
Optical Engineering	36	Q2/Q4
Optical And Quantum Electronics	35	Q2/Q3

Из данной таблицы можно увидеть основные журналы, в которых публикуются исследования по атмосферной оптической связи, а также их рейтинг по наукометрическим базам SCOPUS и WoS.

1.3. Распределения публикаций в области АОЛС по отраслям знаний

В таблице 2 представлены данные о числе публикаций отнесенные в различные предметные области за последние 5 лет (период 2017–2021 годы).

Таблица 4

Распределение публикаций по отраслям знаний за 2017–2021 года

Отрасль знаний	Количество публикаций		Доля, %	
	2017–2021	2021	2017–2021	2021
Технические науки (Engineering)	1553	321	27,4%	27,1%
Физика и астрономия	1408	304	24,8%	25,6%
Вычислительные науки	1025	187	18,1%	15,8%
Материаловедение (Materials Science)	818	173	14,4%	14,6%
Математика (Mathematics)	403	80	7,1%	6,7%
Химия (Chemistry)	161	32	2,8%	2,7%
Науки о Земле и планетах	80	12	1,4%	1%
Науки о принятии решений	48	15	0,8%	1,3%
Энергетика (Energy)	47	12	0,8%	1%
Медицина (Medicine)	36	15	0,6%	1,3%

Из таблицы 2 следует, что наибольшее число публикаций относится к категории «Технические науки», однако близкое число статей приходится на раздел «Физика и астрономия», в котором часто публикуют результаты фундаментальных исследований, что свидетельствует о значительном числе проблем фундаментального характера, которые еще требуют своего решения в области атмосферной оптической связи.

1.4. Распределение мирового потока НТЛ в области АОЛС по странам

На рисунках 3 и 4 приведены первые 10 стран с наибольшим количеством публикаций в области атмосферной оптической связи, которые были проанализированы за периоды с 2001-2010 и за 2021 годы.

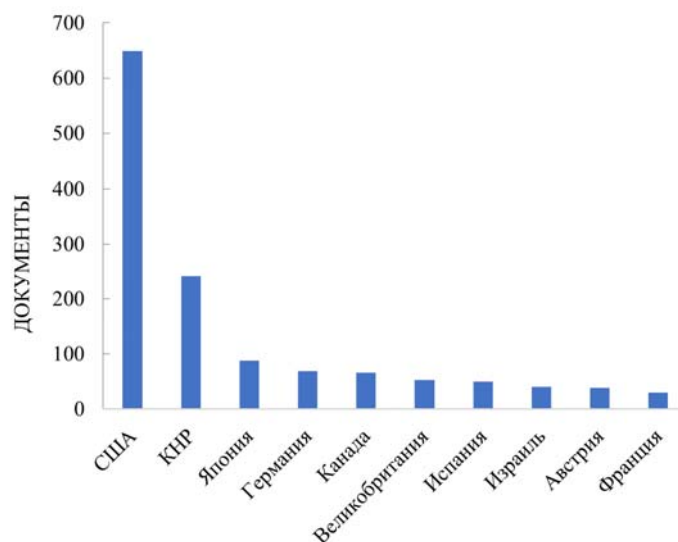


Рис. 10. Распределение публикаций в области атмосферной оптической связи по странам, полученное за период 2001-2010 годы.

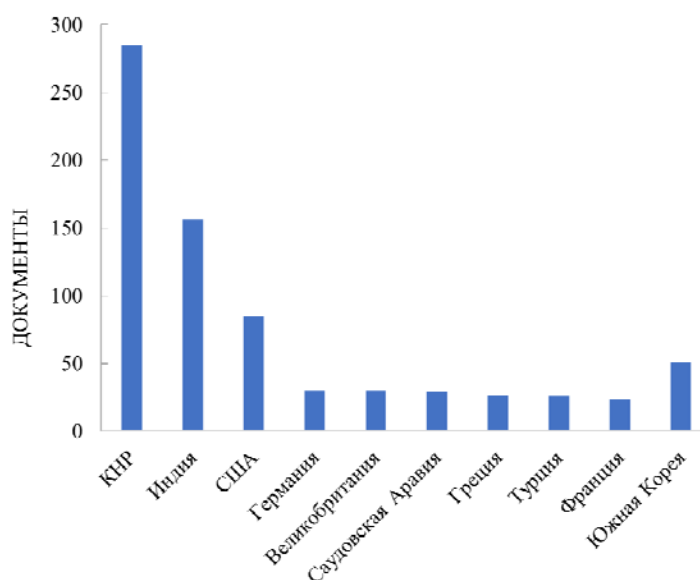


Рис. 11. Рейтинг стран по публикационной активности в области атмосферной оптической связи за 2021 год

Как видно из приведенных диаграмм на рисунках 3 и 4, если раньше явными лидерами в области атмосферной оптической связи являлись исследовательские группы США, то с течением времени лидером в данной области стал Китай. Следует заметить, что в настоящее время в списке стран с наибольшей публикационной активностью в области АОЛС, за исключением Саудовской Аравии, фигурируют страны, территория которых имеет сложный рельеф, затрудняющий прокладку ВОЛС. Поэтому, кроме государств с сильной научной школой по тематике АОЛС, сегодня активно публикуются научные группы из стран, для которых эта технология является жизненно необходимой.

1.5. Организации с высоким уровнем достижений в области атмосферной оптической связи

Для определения точек роста технологии АОЛС, в соответствии с методикой [10], в таблицах 3 и 4 приведены, соответственно, организации с наибольшим числом работ за 2017-2021 годы и организации, финансирующие эти исследования.

Таблица 3

**Распределение организаций, сотрудники которых опубликовали наибольшее число статей
в период с 2017-2021 годы**

Название организаций	Количество опубликованных статей
Chinese Academy of Sciences (CN) https://english.cas.cn/	117
Changchun University of Science and Technology (CN) https://www.cust.edu.cn/	79
University of Chinese Academy of Sciences (CN) https://english.cas.cn/	49
Indian Institute of Technology Delhi (IN) https://iitd.irins.org/	45
Deutsches Zentrum für Luft- Und Raumfahrt (DE) https://www.dlr.de	44
Xidian University (CN) https://en.xidian.edu.cn/	41
Xi'an University of Technology (CN) http://www.xaut.edu.cn/	41

Таблица 4

**Распределение финансирующих организаций по числу исследований, выполненных
при их содействии в период с 2017-2021 годы**

Название организаций	Число исследований, выполненных при их содействии
National Natural Science Foundation of China (CN) https://www.nsf.gov.cn/english/site_1/index.html	333
Fundamental Research Funds for the Central Universities (CN)	56
Ministry of Education of the People's Republic of China (CN)	34
Office of Naval Research (USA) www.onr.navy.mil	34
European Commission (EC) https://ec.europa.eu/	28
National Science Foundation (USA) https://www.nsf.gov/	27
National Key Research and Development Program of China (CN)	26
U.S. Department of Defense (USA) https://www.defense.gov/	24
National Aeronautics and Space Administration (USA) https://www.nasa.gov/	23

В таблице 3 указаны сайты организаций, что позволяет ускорить поиск информации о научных группах специализирующихся в области АОЛС.

Как видно из таблицы 4 наибольшее число публикаций было поддержано различными фондами КНР, что и обуславливает высокую публикационную активность научных организаций Китая. Также видно, что работы в области АОЛС активно финансируются военными ведомствами США, в частности, Управлением военно-морских исследований, Министерством обороны США и Национальным управлением по авиации и исследованию космического пространства США.

1.6. Анализ динамики зарегистрированных патентов на изобретения в области атмосферно оптической связи

Кроме научных статей большое значение для анализа развития технологии в области создания АОЛС можно почерпнуть из патентов на изобретения и устройства, поэтому нами проводился анализ динамики патентов, индексируемых в базе данных SCOPUS. Следует заметить, что данные о числе публикаций и, особенно, патентов, могут меняться со временем, что связано с задержкой индексации публикаций в БД SCOPUS. На рисунке 5 показана зависимость числа выданных патентов на устройства АОЛС от года регистрации патента. Данные о патентах на устройства АОЛС получены на основе поиска в базах следующих патентных бюро: United States Patent & Trademark Office, Japan Patent Office, World Intellectual Property Organization, European Patent Office, United Kingdom Intellectual Property Office. Заявки на патенты из указанных выше патентных бюро индексируются в наукометрической базе данных SCOPUS и поэтому доступны для анализа инструментами БД SCOPUS. Примечательно, что даже в условиях пандемии, когда в 2020 году по многим направлениям наблюдалось снижение научной активности, число зарегистрированных патентов на устройства АОЛС росло, что свидетельствует о безусловной актуальности развития технологии атмосферной оптической связи.

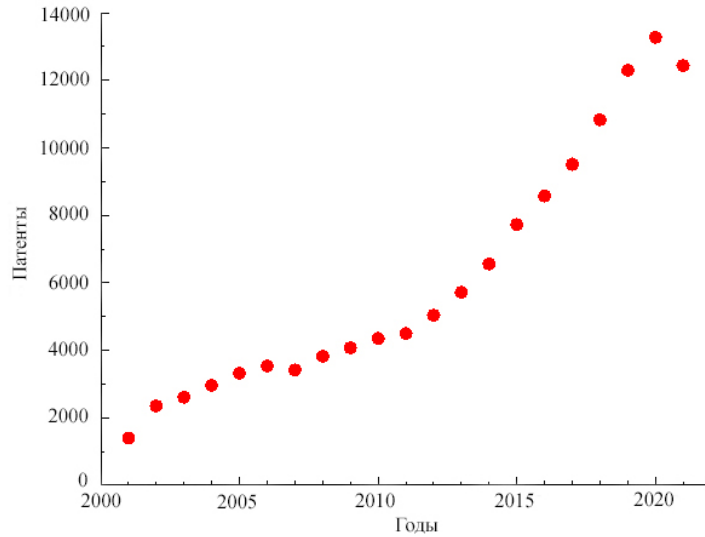


Рис. 12. Зависимость числа выданных патентов, в которых в ключевых словах указана фраза «atmospheric optical communication» от года регистрации патента

Таким образом, анализ публикационной активности в области АОЛС показывает устойчивый рост интереса к данной технологии с 2001 по настоящее время. Рост интереса научных групп заметен, как в общем увеличении числа статей в трудах конференции, так и в высоком рейтинге журналов, в которых публикуются данные исследований. Было выявлено, что за последние 10 лет лидерами в области атмосферной оптической связи стал Китай, хотя раньше явными лидерами являлись исследовательские группы США. Установлено, что даже в условиях пандемии, когда в 2020 году по многим направлениям наблюдалось снижение научной активности, число зарегистрированных патентов на устройства АОЛС в мире продолжало увеличиваться.

2. Подводная оптическая связь

Для определения перспектив развития систем беспроводной оптической подводной связи (БОПС) и точек роста данной технологии также использовалась методика [10]. Системы беспроводной подводной оптической связи имеют большие перспективы на небольших расстояниях, когда использование ВОЛС невозможно, либо не целесообразно [11]. Сегодня это практически единственная технология, которая может быть использована для создания высокоскоростного беспроводного канала связи в водной среде [12-13].

К основным преимуществам этого способа связи можно отнести высокую скорость передачи информации, скрытность и мобильность, что вызывает большой интерес у потенциальных заказчиков. Поэтому активно разрабатываются и совершенствуются численные модели для анализа оптического канала связи в гидросфере и возможности реализации квантовой связи [14, 15]. В БОПС используется область видимого спектрального диапазона, которая попадает в окно прозрачности воды. На рисунке 6 приведен график оптического пропускания воды, из которого видно, что под водой для связи оптимально использовать излучения в спектральной области 450-530 нм [12]. Для оценки актуальности и уровня развития технологий БОПС проводился поиск в БД SCOPUS по ключевым словам: **беспроводная подводная оптическая связь («wireless underwater optical communication»)**, в результате которого была получена выборка научных публикаций, индексируемых в этой базе за временной период с 2001 года по 2021 год.

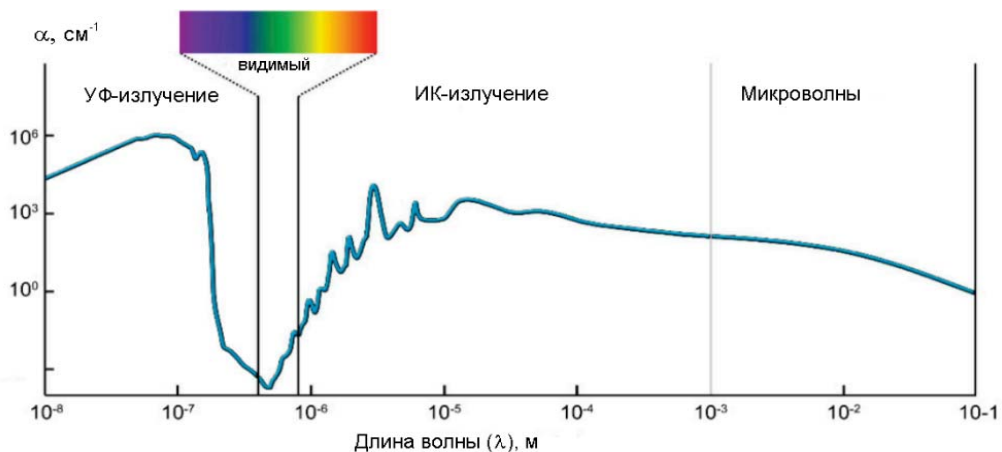


Рис. 13. Зависимость коэффициента поглощения оптического излучения от длины волны света [13]

2.1 Динамика публикационной активности в области БОПС за период 2001–2021 гг.

Рисунок 7 иллюстрирует зависимость числа публикаций от года публикации. Всего в выборку попало 1072 публикации. На временном отрезке с 2005 по 2021 годы заметен рост интереса к данной области. На графике учитывались только публикации в научных журналах и трудах конференций, патенты не учитывались.

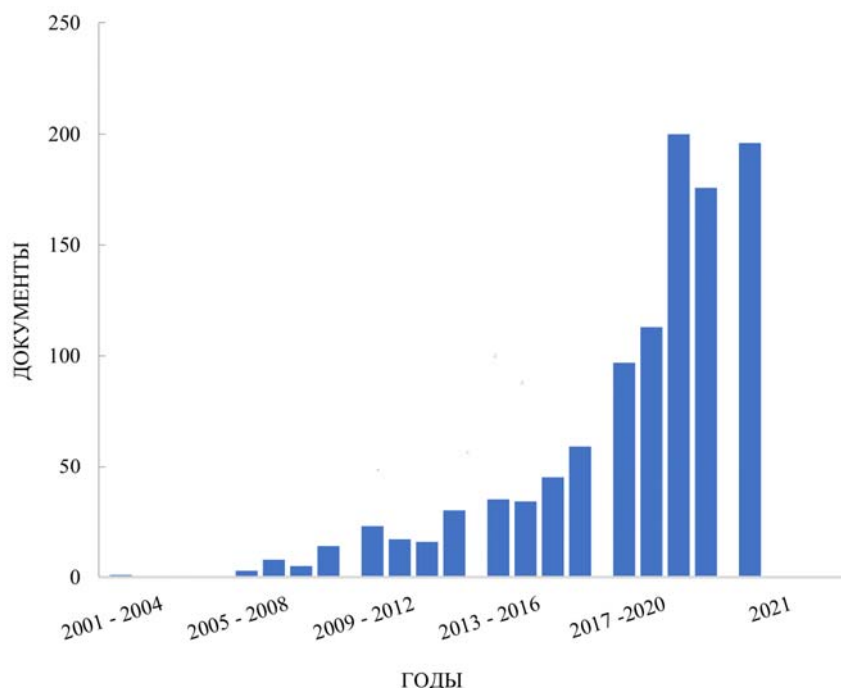


Рис. 14. Динамика публикационной активности по тематике беспроводной подводной оптической связи

2.2 Основные издания, в которых опубликованы результаты исследований по БОПС

В таблице 6 представлен Топ 10 журналов, в которых было больше всего опубликовано статей по БОПС за последние 5 лет в период с 2017–2021 годы.

Таблица 5

Журналы, в которых за 2017–2021 годы опубликовано наибольшее число статей по беспроводной подводной оптической связи

Название журнала	Число статей за 2017-2021 годы	Квартиль журнала (2020 г.)
Optics Express	42	Q1/Q1
Optics Communications	37	Q2/ Q3
Proceedings of SPIE	34	Без квартиля
IEEE Photonics Journal	28	Q1/Q2
IEEE Access	27	Q1/Q2
Applied Optics	19	Q1/Q3
Journal Of Lightwave Technology	14	Q1/Q1
Chinese Optics Letters	13	Q2/Q2
IEEE Transactions On Communications	11	Q1/Q1
Journal of The Optical Society of America A Optics and Image Science and Vision	11	Q1/Q2

Основываясь на данных таблицы 6, можно сделать вывод: рейтинг журналов, в которых публикуются данные исследований по БОПС высокий, что свидетельствует об актуальности данной тематики.

2.3. Распределения публикаций в области БОПС по отраслям знаний

В таблице 7 представлены данные о числе публикаций отнесенные в различные предметные области за последние 20 лет (период 2017–2021 годы).

Таблица 6

Распределение публикаций по отраслям знаний за 2017–2021 года

Отрасль знаний	Количество публикаций		Доля, %	
	2017–2021	2021	2017–2021	2021
Технические науки	562	144	29%	27,7%
Физика и астрономия	442	120	22,8%	23,1%
Вычислительные науки	391	106	20,2%	20,4%
Материаловедение	236	65	12,2%	12,5%
Математика	83	31	4,3%	6%
Химия	56	10	2,9%	1,9%
Науки о Земле и планетах	29	4	1,5%	<1%
Науки о принятии решений	37	11	1,9%	2,1%
Энергетика	24	7	1,2%	1,3%
Медицина	19	6	1%	1,2%

Из таблицы 7 видно, что наибольшее число публикаций относится к категории «Технические науки», в которую часто включаются работы по беспроводным подводным оптическим исследованиям. Однако суммарное число публикаций, в таких технических категориях, как «Вычислительные науки» и «Физика и астрономия», так же имеют большое значение, что свидетельствует о значительном интересе, как к прикладным, так и к фундаментальным аспектам технологий БОПС [10].

2.4. Распределение мирового потока НТЛ в области беспроводной подводной оптической связи по странам

На рисунках 8 и 9 приведены данные для десяти стран-лидеров с наибольшим количеством публикаций по теме исследований в области беспроводной подводной оптической связи, которые были проанализированы в разные промежутки времени.

Как видно из приведенных диаграмм, если раньше явными лидерами в области атмосферно оптической связи являлись исследовательские группы США, то с течением времени лидером в данной области стал Китай. Также стоит отметить, что в топ стран лидеров по публикационной активности вошла Индия и другие страны, которые ведут активную деятельность в морских акваториях, например, связанную с добычей полезных ископаемых.

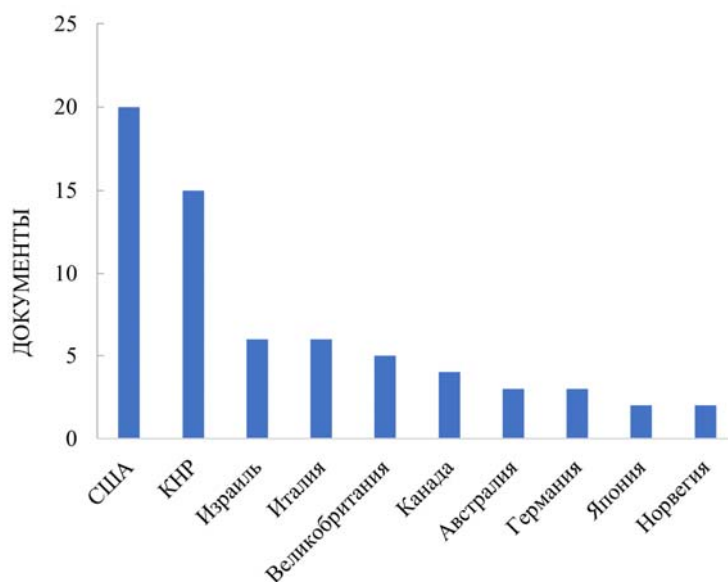


Рис. 15. Распределение публикаций в области беспроводной подводной оптической связи по странам, полученное за период 2001–2010 годы.

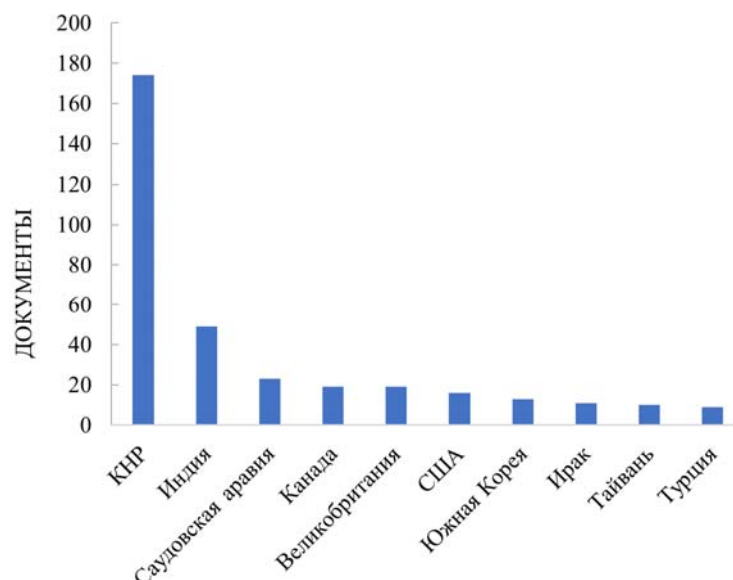


Рис. 16. Рейтинг стран по публикационной активности в области беспроводной подводной оптической связи за 2021 год

2.5. Организации с высоким уровнем достижений в области БОПС

В таблицах 8, 9 приведены, соответственно, организации и финансирующие организации с наибольшим числом работ за 2001-2021 годы.

Таблица 8

Распределение организаций, сотрудники которых опубликовали наибольшее число статей в период с 2001–2021 годы

Название организаций	Количество опубликованных статей
King Abdullah University of Science and Technology (Saudi Arabia) https://www.kaust.edu.sa/en	65
Zhejiang University (China) https://www.zju.edu.cn/	54
Ocean College, Zhejiang University (China) http://www.zjou.edu.cn/	33
Chinese Academy of Sciences (China) https://english.cas.cn/	30
Tsinghua University (China) https://www.tsinghua.edu.cn/	29
Fudan University (China) https://www.fudan.edu.cn/	23
Xidian University (China) https://www.xidian.edu.cn/	21
Huazhong University of Science and Technology (China) https://www.hust.edu.cn/	18
University of Science and Technology of China (China) https://www.ustc.edu.cn/	17
National Taipei University of Technology (Taiwan) https://www.ntut.edu.tw/bin/home.php	16

Таблица 7

Распределение финансирующих организаций по числу исследований, выполненных при их содействии в период с 2001-2021 годы

Название организаций	Число исследований, выполненных при их содействии
National Natural Science Foundation of China (China) https://www.nsf.gov.cn/english/site_1/index.html	184
Fundamental Research Funds for the Central Universities	40
National Key Research and Development Program of China (China)	33
Ministry of Science and Technology of the People's Republic of China (China) http://www.most.gov.cn/index.html	32
King Abdullah University of Science and Technology (Saudi Arabia) https://www.kaust.edu.sa/en	32

National Basic Research Program of China (973 Program) (China) https://www.mfa.gov.cn/ce/cgny/eng/kjsw/std/t1345404.htm	29
Ministry of Education of the People's Republic of China (China) http://en.moe.gov.cn/	24
Ministry of Finance	20
National Science Foundation (USA) https://www.nsf.gov/	14
King Abdulaziz City for Science and Technology (Saudi Arabia) http://www.kacst.edu.sa/	14

2.6 Анализ динамики выдачи патентов на изобретения в области атмосферно оптической связи

На рисунке 10 показана зависимость числа выданных патентов в области БОПС от времени.

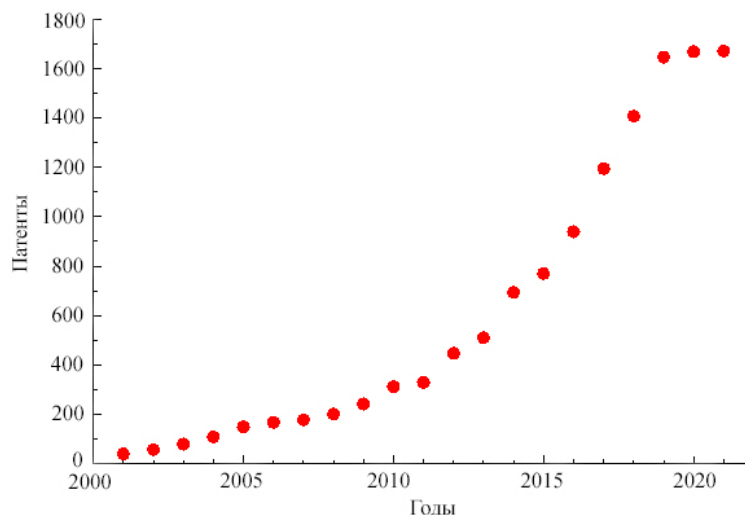


Рис. 17. Число выданных патентов на устройства и изобретения в области беспроводной подводной оптической связи

Следует заметить, что база данных Scopus пополняется непрерывно, поэтому данные о числе публикаций могут меняться и количество публикаций за последние 2 года будет со временем увеличиваться. Примечательно, что, как видно из рисунка 10, даже в условиях пандемии, когда в 2020 году по многим направлениям наблюдалось снижение научной активности, число зарегистрированных патентов на устройства для исследований в области беспроводной подводной оптической связи в мире продолжало увеличиваться, что свидетельствует о высоком прикладном потенциале данного вида связи. В публикациях за последние 10 лет постоянно сообщается об увеличении, скорости и дальности передачи данных в БОПС, в связи с чем, так же как в атмосфере сегодня стали интенсивно проводиться исследования квантового распределения ключа (КРК) в гидросфере. Известно, что технология КРК позволяет на физическом уровне обеспечить безопасность передачи криптографических ключей [16, 17] и поэтому, вслед за ВОЛС технология КРК начинает интенсивно исследоваться применительно к открытому пространству. С учетом этого представляет интерес оценка степени зрелости данного направления, однако это требует отдельного рассмотрения, которое будет проведено в следующих работах авторов.

Заключение

Установлено, что публикационная активность научных групп разных стран в области оптической связи в атмосфере и гидросфере коррелирует с географическим положением страны и ее заинтересованностью в развитии этих технологий. При этом научные журналы, публикующие эти материалы, имеют высокий рейтинг. В развитии оптической связи в гидросфере в настоящее время повышенную активность проявляют научные группы из стран, которые активно добывают или транспортируют природные ресурсы в своих акваториях.

Анализ результатов опубликованных исследований, показал, что страны со сложным географическим рельефом активно развивают АОЛС, а страны, для которых ресурсы моря и океана являются жизненно необходимыми - активно развивают БОПС. При этом Россия не вошла в топ 10 стран-лидеров в данных разделах, что может свидетельствовать о не достаточно высоком уровне проработки этих технологий, а также недостаточном финансировании работ в области оптической связи в открытом пространстве. Поэтому требуется специальная финансовая поддержка работ направленных на развитие технологий оптической связи в атмосфере и гидросфере, чтобы в полной мере использовать на практике преимущества, которые дает беспроводная оптическая связь.

Литература

1. *Езекелян Э. Э., Джалалов И. К.* Перспективное применение технологии FSO // Технологии информационного общества: Материалы XIII Международной отраслевой научно-технической конференции, Москва, 20-21 марта 2019 года. М.: Издательский дом Медиа паблицер, 2019. С. 19-22.
2. *Zafar S., Khalid H.* Free Space Optical Networks: Applications, Challenges and Research Directions // Wireless Personal Communications. 2021. Т. 121. №. 1. С. 429-457.
3. *Кузнецов С., Огнев Б., Поляков С.* FSO на "последней (и не только) миле". Практические результаты // Первая миля. 2009. Т. 10. №. 1. С. 30-33.
4. *Зуев В.Е.* Распространение видимых и инфракрасных волн в атмосфере. М.: Сов. Радио, 1970. 325 с.
5. <http://www.mostkom.ru/optical-ground-stations/>
6. *Боев А. А., Кернос М. Ю., Кузнецов С. Н. и др.* Беспроводной канал передачи информации со скоростью 40 Гбит/с // Вестник Рязанского государственного радиотехнического университета. 2017. № 62. С. 44-48. DOI 10.21667/1995-4565-2017-62-4-44-48.
7. *Чуляева Е. Г., Кузнецов С. Н., Огнев Б. И.* Концепция построения сети беспроводной оптической связи между надводными кораблями // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Физико-математические науки. 2018. Т. 11. № 1. С. 66-72. DOI 10.18721/JPM.11107.
8. *Дымкова С.С.* Научометрические инструменты в оценке и управлении публикационной активностью научных организаций // Культура: теория и практика. 2021. № 5-6(44-45).
9. *Мионов Ю.Б., Казанцев С. Ю., Шаховой Р. А. и др.* Анализ перспектив развития источников одиночных фотонов в системах квантового распределения ключей // Научометрические технологии в космических исследованиях Земли. 2021. Т. 13. № 6. С. 22-33.
10. *Машковцева Л. С., Болотов Д. В., Казанцев С. Ю. и др.* Научометрический анализ публикаций по источникам одиночных фотонов для систем связи с квантовым распределением ключей // Научно-техническая информация. Серия 1: Организация и методика информационной работы. 2022. № 1. С. 22-31. DOI 10.36535/0548-0019-2022-01-3.
11. *Душин С. В. и др.* // DSPA: Вопросы применения цифровой обработки сигналов. 2020. Т. 10. №. 2. С. 11-18.
12. *Shherbakov A. V., Petruhin G. D., Miroshnikova N. E., Titovets P. A.* Estimation of underwater optical communication link operating distance // T-Comm. 2020. Vol. 14. No 3. P. 54-60. DOI: 10.36724/2072-8735-2020-14-3-54-60.
13. *Schirripa Spagnolo G., Cozzella L., Leccese F.* // Sensors. 2020. Vol. 20. №. 8. P. 2261. DOI: 10.3390/s20082261
14. *Мирошникова Н. Е., Петрухин Г. Д., Щербаков А. А.* Проблемы моделирования гидросферных оптических линий связи // Системы синхронизации, формирования и обработки сигналов. 2019. Т. 10. № 5. С. 29-39.
15. *Казанцев С. Ю., Мирошникова Н. Е., Титовец П. А., Подгорный А. А.* Квантовая подводная связь // Новые технологии высшей школы. Наука, техника, педагогика: Материалы Всероссийской научно-практической конференции, Москва, 26 марта 2021 года. М.: Московский Политех, 2021. С. 23-28.
16. *Корольков А. В.* // Вопросы кибербезопасности. 2015. №. 1 (9).
17. *Быковский А. Ю., Компанец И. Н.* // Квантовая электроника. 2018. Т. 48. №. 9. С. 777-801.

PROSPECTS FOR THE DEVELOPMENT OF LASER COMMUNICATION SYSTEMS IN THE ATMOSPHERE AND HYDROSPHERE

Stepan A. Andreev,
Graduate MTUCI, Moscow, Russia,
stefan_4@mail.ru

Sergey Y. Kazantsev,
Professor of the Department of the DTS, Doctor of Physical and Mathematical Sciences, MTUCI, Moscow, Russia,
es.i.kazantsev@mtuci.ru

Anastasia I. Svistunova,
Graduate MTUCI, Moscow, Russia,
nastya.svistunova.99@bk.ru

Pavel A. Titovets,
Associate Professor of the Department of the TE&A, Ph.D., MTUCI, Moscow, Russia,
paveltitovec@yandex.ru

Abstract

The key features of information transmission along a laser beam in FSO. An analysis of the flow of scientific and technical literature in the field of wireless optical communications in the atmosphere and under water is presented. The main trends in the development of laser communication systems in air and water are considered. The points of growth of wireless laser communication technologies in the atmosphere and hydrosphere have been identified. The groups of countries that are most actively developing wireless laser communication technologies in FSO and the main sources of funding for work in this direction are identified. It has been established that the publication activity of scientific groups in different countries in the field of optical communications in the atmosphere and hydrosphere correlates with the geographical position of the country and its interest in the development of these technologies. In the development of optical communications in the hydrosphere, scientific groups from countries that are actively developing natural resources in the coastal waters are currently showing increased activity.

Keywords: *atmospheric optical communication lines, scientometric analysis, wireless optical underwater communication, quantum key distribution, communications in free space, FSO communications.*

ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК ЛАЗЕРНЫХ МОДУЛЕЙ, ВХОДЯЩИХ В МНОГОФУНКЦИОНАЛЬНУЮ ЛАЗЕРНУЮ УСТАНОВКУ (МЛУ-3)

Кузьмин Михаил Сергеевич,
магистрант МТУСИ, Москва, Россия,
restlogin@yandex.ru

Насараиа Арчил Петрович,
магистрант МТУСИ, Москва, Россия,
archil.111@mail.ru

Машковцева Любовь Сергеевна,
доцент кафедры НТС, к.х.н., МТУСИ, Москва, Россия,
liubov.mash@gmail.com

Подлесных Сергей Владимирович,
научный сотрудник, МТУСИ, Москва, Россия,
winrad@yandex.ru

Аннотация

Представлено описание многофункциональной лазерной установки МЛУ-3, разработанной для проведения лабораторных работ в курсах «Оптика» и «Квантовая оптика». Созданная лазерная установка МЛУ-3 позволяет формировать лазерное излучение на трёх длинах волн: 532 нм (зеленый), 405 нм (фиолетовый) и 660 нм (красный). Приведены результаты детального исследования энергетических и поляризационных характеристик лазерного излучения на всех трех длинах волн.

Ключевые слова: круговая поляризация, степень поляризации, лазерная установка, демонстрационный эксперимент.

Введение

В настоящее время при проведении лабораторных и учебно-исследовательских работ в области оптики, квантовой физики и квантовой криптографии широко применяются лазерные установки. При чтении теоретических курсов по этим предметам также большое значение имеет демонстрационные эксперименты, для которых целесообразно применять лазеры видимого диапазона [1,7]. Лазер стал мощным исследовательским инструментом изучения геометрической и физической оптики. В частности, такие фундаментальные явления как интерференция, дифракция и поляризация хорошо демонстрируются с помощью лазерной установки [2]. В системах квантовой связи вместо применения идеальных источников одиночных фотонов используют ослабленные лазерные импульсы [3,4]. Целью данной статьи является представление результатов создания многофункциональной установки для выполнения лабораторных и учебно-исследовательских работ по курсам «Оптика», «Квантовая оптика».

1. Составные части МЛУ-3

Блок схема многофункциональной лазерной установки представлена на рисунке 1. Установка включает в себя три независимых ШИМ-генератора с цифровой системой управления модели Zk -PP1, имеющих два режима управления: режим непрерывной ШИМ-генерации с включением/выключением выходного сигнала и режим счета импульсов с ручной регулировкой длительности импульса/пауза, количества выходных импульсов и времени задержки перед включением. Генераторы имеют следующие характеристики: напряжение питания может варьироваться от 3,3 В до 30 В; генераторы оснащены защитой выхода от короткого замыкания и защитой от ВЧ помех; выходная частота следования импульсов может изменяться от 1 Гц до 150 кГц с точностью 2%, а в диапазоне 0-100 кГц с точностью 1%; количество импульсов может варьироваться в диапазоне 1-9999-бесконечность, обозначаемой на дисплее символом «----»; время задержки запуска может изменяться в диапазоне 0,000s-9999s; минимальное значение задержки запуска=1 мс; длительность выходного импульса/паузы также может изменяться в диапазоне 0,000s-9999s; минимальное значение длительности равно 1 мс; выходной ток не превышает 30мА; амплитуда выходного сигнала равна амплитуде питающего напряжения.

Блок-схема (рис. 1) также содержит три драйвера управления лазерными модулями с обратной связью по току, проходящему через лазерный диод модуля и выходом на вентилятор охлаждения с напряжением питания 12 В и име-

ющим вход управления TTL импульсами с частотой следования 0-150 кГц; зеленый лазерный модуль марки APTLHG модели G-200-005 с максимальной выходной мощностью излучения 200 мВт, с длиной волны 532 нм, напряжением питания 12 В и принудительным охлаждением; фиолетовый лазерный модуль марки Laserlands, модели TEMOO с максимальной выходной мощностью излучения 200 мВт, с длиной волны 405 нм, напряжением питания 12В и принудительным охлаждением. Красный лазерный диод запитан через второй (запасной) драйвер от модуля зеленого лазера с максимальной выходной мощностью 1 мВт и длиной волны 650-670 нм.

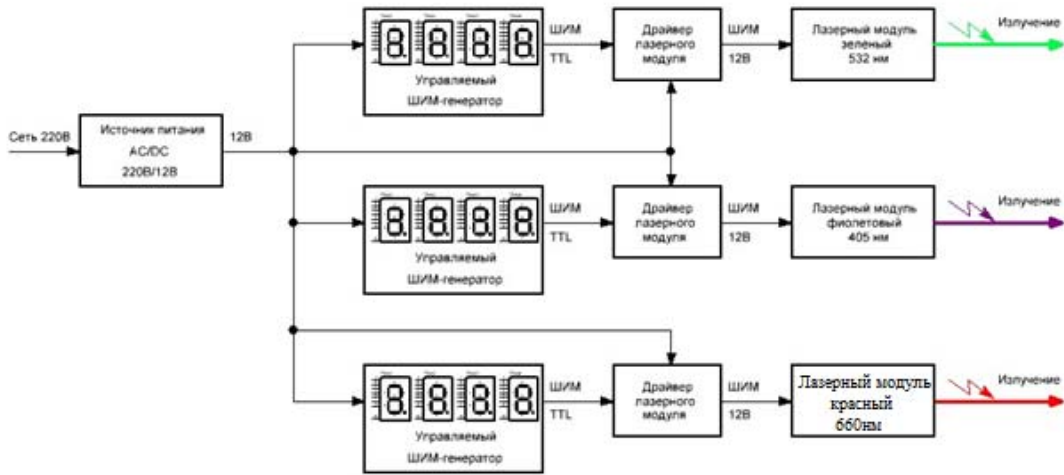


Рис. 1. Блок схема МЛУ-3

2. Основные параметры многофункциональной лазерной установки МЛУ-3

МЛУ-3 питается от сети переменного тока 220В, потребляя общую мощность не более 5 Вт. Установка имеет 3 выходных канала лазерного излучения. Выходная длина волны лазерного излучения: 1 канал – 660 нм (красный), 2 канал – 532 нм (зеленый), 3 канал – 405 нм (фиолетовый). Выходная максимальная мощность излучения 200 мВт по каналам 2 и 3, 1 мВт – по каналу 1. Реальные значения выходной мощности по фиолетовому каналу полностью соответствуют описанию, по зеленому каналу значение завышено практически в 2 раза (при 200 мВт реальное значение мощности составляет 114 мВт). Диапазон регулировки выходной частоты следования лазерных импульсов 1-150 кГц с точностью до 1 Гц по всем каналам не зависимо друг от друга. Диапазон регулирования выходной мощности 0-100% с точностью 1% по всем каналам не зависимо друг от друга. Регулировка выходной мощности осуществляется методом изменения соотношения длительностей импульс/пауза (ШИМ-регулирование мощности). Данный метод позволяет менять частоту излучения без изменения средней мощности излучения.

На рисунке 2 показаны: осциллограмма импульсов зеленого лазерного модуля с частотой 39 Гц и осциллограмма импульсов зеленого лазерного модуля с частотой 100 Гц. При этом на данных осциллограммах одинаковая амплитуда равная 3.56 В. При установке 100% выходной мощности на выходе модуля появится непрерывное излучение. Есть возможность ручного включения/выключения выходного излучения по всем каналам не зависимо друг от друга. Охлаждение лазерных модулей – принудительное.

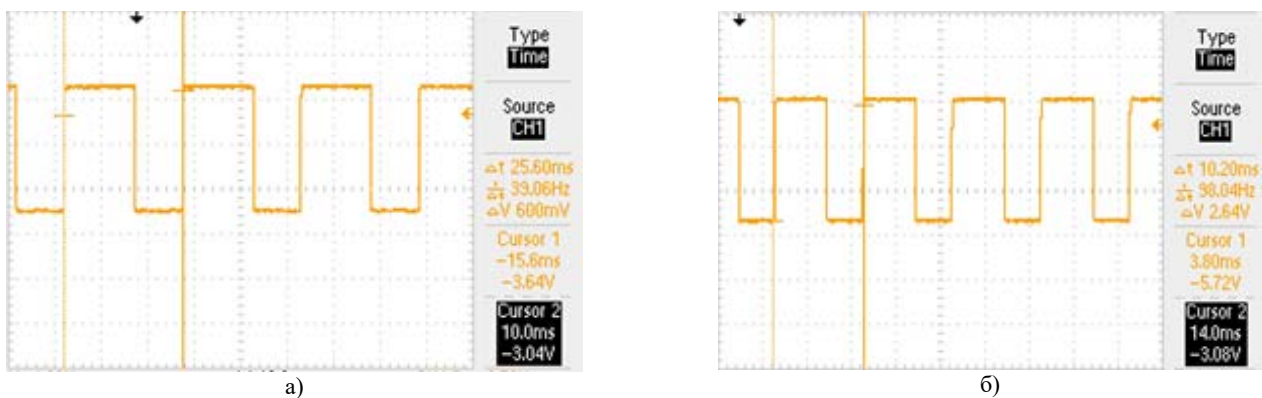


Рис. 2. а) осциллограмма импульсов с частотой 39 Гц; б) осциллограмма импульсов с частотой 100 Гц

3. Экспериментальная установка и процедура измерений.

На рисунке 3 приведена экспериментальная установка для исследования характеристик МЛУ-3.

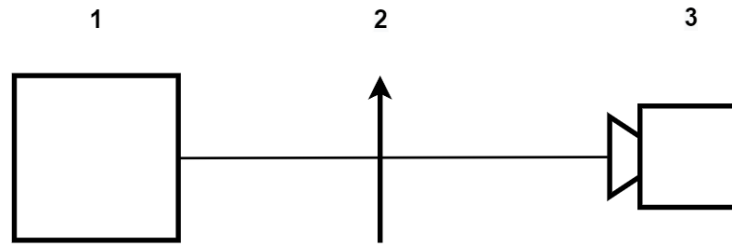


Рис. 3. Обобщённая схема измерительной установки:
1 – многофункциональная лазерная установка (МЛУ-3); 2 – анализатор; 3 – фотоприемник

В качестве анализатора была использована поляризационная пленка, закреплённая в оптическую подвижку. В качестве фотоприёмника использовался фотодетектор, построенный на готовой монолитной микросхеме OPT101.

Фотоприемник и его принципиальная схема представлены на рисунке 4. Фиксирование сигналов с фотоприемника производится на осциллографе Tektronix.

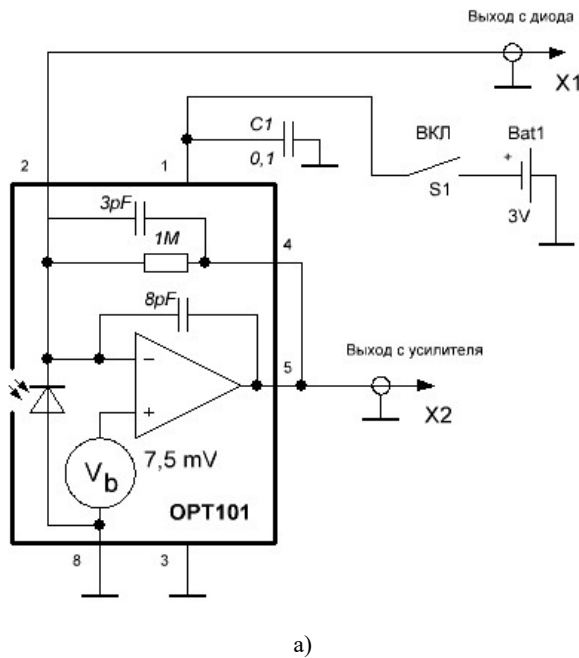


Рис. 4: а) принципиальная схема фотодетектора; б) фотоприемник

Излучение одного из трех лазерных модулей с интенсивностью I_0 , подавалось на фоточувствительный элемент фотоприемника/термического датчика. При этом лазерное излучение проходило через анализатор, состоящий из: поляризационной пленки, зафиксированной в оптической подвижке, которая позволяет изменять угол поляризации относительно оптической оси. Таким образом, исходя из закона Малюса [4,5], фотоприёмник фиксирует излучение с интенсивностью равной $I_0 \cos^2 \varphi$, где φ - угол между плоскостью падающего света и плоскостью поляризатора.

Основные характеристики используемого в экспериментах фотоприемника приведены в таблице 1.

Таблица 1

Основные характеристики фотоприемника

Размер фоточувствительного элемента	5,2 мм ²
Точковая чувствительность на длине волны 650 нм	0,45 А/Вт
Темновой ток при напряжении на диоде 7,5 мВ	2,5 пА (при 7 °С)
Внутренняя емкость перехода диода	1200 пФ

Измерение мощности лазерных модулей производилось термическим датчиком Thorlabs S405C, который подключался к измерителю мощности Thorlabs PM100D (рис. 5).



Рис. 5. Измеритель мощности Thorlabs PM100D и термический датчик Thorlabs S405C

В таблице 2 представлены основные характеристики термического датчика

Таблица 2

Основные характеристики фотоприемника

Диапазон длин волн	190 нм. – 20 мкм.
Диапазон оптической мощности	100 мкВт. – 5 Вт.
Размер апертуры	10 мм.
Размер активной области	10 мм. * 10 мм.
Макс. плотность оптической мощности	1.5 кВт/см ² (средн.)
Линейность	±0.5%
Время отклика	1.1 с.

4. Результаты выполненных исследований

На рисунке 6 приведены осциллограмма импульсов излучения лазерного модуля (а) и осциллограмма импульсов излучения лазерного модуля после прохождения через анализатор (б).

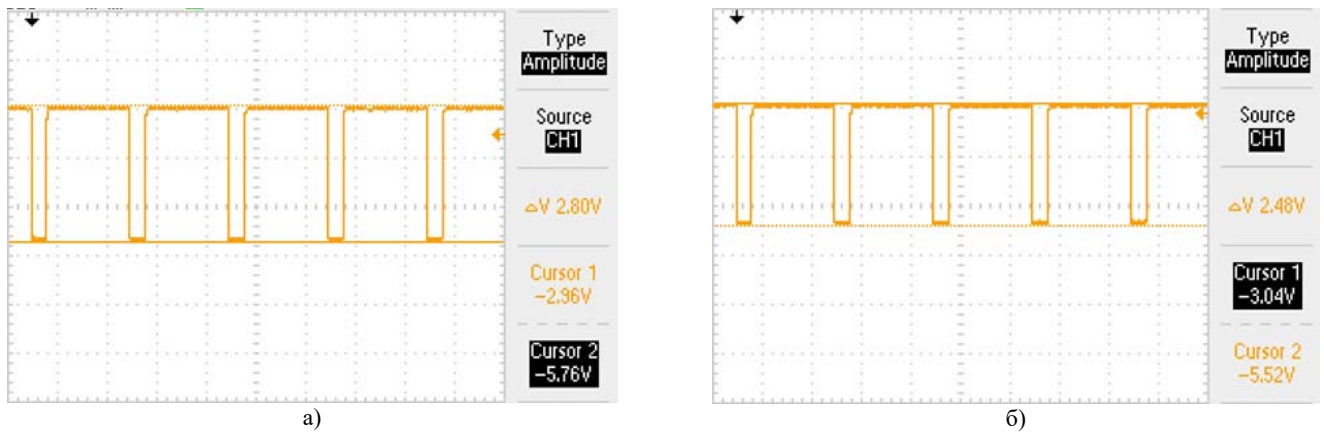


Рис. 6: а) осциллограмма лазерных импульсов с выхода МЛУ-3;
 б) осциллограмма лазерных импульсов после прохождения через анализатор

Как видно из рисунка 6 прохождение через анализатор не вносит значительных колебаний амплитуды.

На рисунке 7 показаны соотношение мощности лазерных импульсов и амплитуды от угла φ для всех лазерных модулей МЛУ-3.

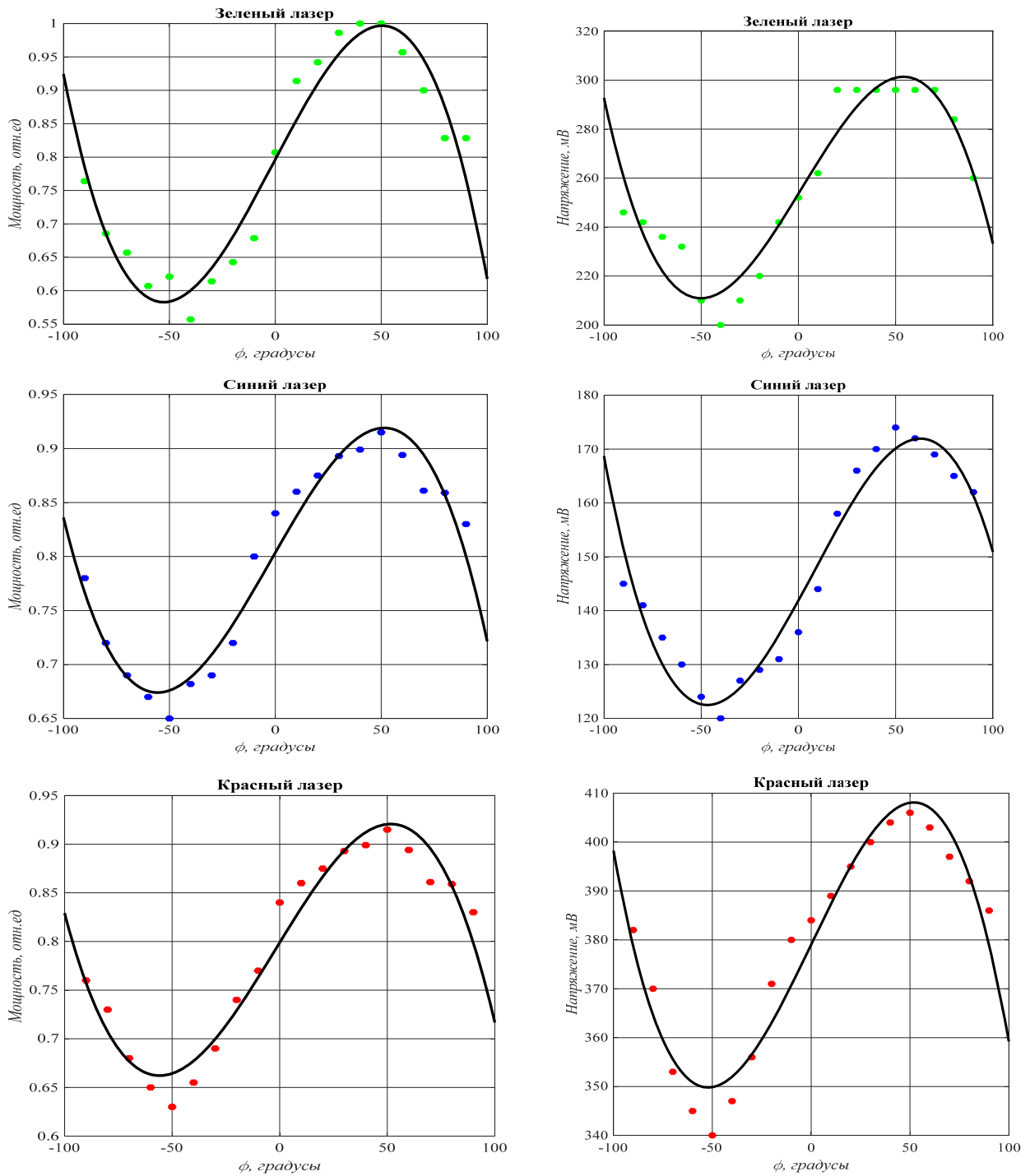


Рис. 7. Соотношение мощности лазерных импульсов и амплитуды от угла φ

Зависимости мощности лазерного излучения от угла поворота анализатора полученные, в результате эксперимента для красного и синего лазера практически совпадают с теоретической зависимостью. В то время как соотношения излучение зеленого модуля имеет слабую степень поляризации.

Для трех лазерных модулей МЛУ-3 была рассчитана степень поляризации P :

$$P = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \quad (1)$$

Где I_{\max}, I_{\min} - максимальная и минимальная интенсивность частично поляризованного света.

В таблице 3 представлены показатели степени поляризации трех лазерных модулей, рассчитанных на основе данных, полученных в ходе эксперимента [6].

Степень поляризации лазерных модулей

Лазерный модуль	Степень поляризации, P
Фиолетовый лазер	0,874
Красный лазер	0,84
Зеленый лазер	0,19

Степень поляризации фиолетового и красного лазерного модуля высока, т.к. она близка к 1, это доказывает, что излучение фиолетового и красного модуля линейно поляризовано. Зеленый лазерный модуль имеет слабую степень поляризации, что может свидетельствовать о его эллиптической поляризации. Это возможно ввиду сложной конструкции зеленого лазерного модуля, состоящего из ИК -диода, который накачивает неодимовый лазер. В качестве активной среды неодимового лазера используется кристалл алюмо-иттриевого граната ($Y_3Al_5O_{12}$), легированного ионами неодима (Nd). Неодимовый лазер генерирует излучения с $\lambda=1064$ нм. Далее это излучение попадает в оптический нелинейный кристалл ($KTiOP_4$), в котором происходит удвоение частоты и на выходе получается излучение с $\lambda=532$ нм [8]. Известно, что линейно поляризованное излучение при прохождении через призму может приобрести эллиптическую поляризацию [9-10].

Заключение

Многофункциональной лазерной установки МЛУ-3 позволяет формировать непрерывное и импульсно-периодическое излучение на трех длинах волн: $\lambda=405$ нм; $\lambda=532$ нм; $\lambda=650$ нм. Диапазон частот для всех трех модулей может изменяться от 0 до 150кГц. Максимальная мощность выходного излучения для зеленого и фиолетового модуля – 200 мВт, для красного модуля - 1мВт. Степень поляризации фиолетового лазерного модуля – 0.874, красного – 0.84, зеленого – 0.19. Характеристики МЛУ-3 позволяют использовать ее в демонстрационных экспериментах по курсам вузовских учебных дисциплин: «Оптика», «Квантовая оптика», а также при выполнении учебно-исследовательских работ.

Литература

1. Агеев Л. А., Кулева М. Г., Яровая Р. Г. Простой демонстрационный эксперимент по нелинейной оптике (Тепловая дефокусировка лазерного излучения) // Успехи физических наук. 1987. Т. 151. №. 3. С. 537-540.
2. Ландсберг Г.С. Оптика. М.: Наука, 1976.
3. Миронов Ю.Б., Казанцев С.Ю., Шаховой Р.А. Анализ перспектив развития источников одиночных фотонов в системах квантового распределения ключей // Научные технологии в космических исследованиях Земли. 2021. Т. 13. № 6. С. 22-33.
4. Чижевский В.Н. Быстрая генерация случайных бит на основе поляризационных шумов в полупроводниковом лазере с вертикальным резонатором // Оптика и спектроскопия. 2011. Т. 111. № 5. С. 724-730.
5. Шаховой Р.А., Тумачек А.С., Андропова Н.М. Анализ зависимости фазовой диффузии от параметров лазера, работающего в режиме переключения усиления // Сборник научных трудов X Международной конференции по фотонике и информационной оптике – Национальный исследовательский ядерный университет "МИФИ", 2021. С. 105-106.
6. Кингсен А.С., Локшин Г.Р., Ольхов О.А. Основы Физики - Т.1 Механика, электричество и магнетизм, колебания и волны, волновая оптика. М.: Физматлит, 2001. Ч. II, гл. 11; ч. III, гл. 10.
7. Еркович С. П., Воробьев С. А., Наумов А. Ф. Изучение явления поляризации света. Методические указания к лабораторной работе 0-22 по курсу общей физики./ Под редакцией А.Г. Андреева. "Ярославль", 2013. С. 2-6.
8. Васенин А.Г., Орлов В.Г. Компьютерная модель измерительного комплекса для ВОЛС // Фундаментальные проблемы радиоэлектронного приборостроения. 2010. Т. 10. № 1-1. С. 168-17.
9. Звельто О. Принципы лазеров. Лань, 2008.
10. Калитеевский, Н. И. Волновая оптика: учебное пособие/ Н. И. Калитеевский. -5-е изд. Санкт-Петербург: Лань, 2008. 480 с. ISBN 978-5-8114-0666-1.

**INVESTIGATION OF THE CHARACTERISTICS OF LASER MODULES INCLUDED
IN A MULTIFUNCTIONAL LASER INSTALLATION (MLU-3).**

Mikhail S. Kuzmin.,
Graduate MTUCI, Moscow, Russia,
restlogin@yandex.ru

Archil P. Nasaraia,
Graduate MTUCI, Moscow, Russia,
archil.111@mail.ru

Liubov S. Mashkovtseva,
Associate Professor of the Department of DTS, Ph.D. in chemistry, MTUCI, Moscow, Russia? liubov.mash@gmail.com

Sergey V. Podlesnykh,
Researcher, MTUCI, Moscow, Russia,
winrad@yandex.ru

Abstract

The description of the multifunctional MLU-3 laser installation developed for laboratory work in the courses "Optics" and "Quantum Optics" is presented. The created MDR-3 laser installation allows to generate laser radiation at 3 wavelengths: 532nm (green), 405 nm (purple) and 660 nm (red). A detailed study of the energy and polarization characteristics of laser radiation at all three wavelengths has been carried out.

Keywords: *circular polarization, degree of polarization, laser installation, demonstration experiment.*

АНАЛИЗ РИСКОВ БЕСПРОВОДНОЙ ОПЛАТЫ С ИСПОЛЬЗОВАНИЕМ NFC

*Тимошук Юлия Сергеевна,
студент МТУСИ, Москва, Россия,
iul.tim2012@yandex.ru*

*Маклачкова Виктория Валентиновна,
директор по науке НО АПОС,
старший преподаватель кафедры СИТиС, МТУСИ, Москва, Россия,
v.v.maklachkova@mtuci.ru*

Аннотация

NFC-платежи становятся все более распространенным способом оплаты повседневных нужд. NFC – это технология RFID ближнего радиуса действия для обеспечения безопасности передачи данных. Однако, технология обладает рядом уязвимостей, которые могут подвергать риску не только денежные средства пользователей, но и их личные данные. В данной статье обсуждаются текущие проблемы NFC-платежей и возможные способы их решения.

Ключевые слова: *риск, уязвимость, радиочастотная идентификация, NFC, RFID, безопасность данных, бесконтактные платежи.*

Введение

В последние годы технологии RFID (radio frequency identification, радиочастотной идентификации) и NFC (near field communication, связи в ближнем поле) продолжают приобретать широкую популярность, не только в привычных сферах логистики и транспорта (для идентификации грузов и товаров), но также и в таких сферах, как здравоохранение и торговля в качестве электронных удостоверений личности и идентификаторов бесконтактных электронных платежей [13, 21, 22, 24-29].

NFC – это усовершенствованная беспроводная технология малого радиуса действия. С помощью NFC потребители могут обмениваться цифровым контентом, подключать другие устройства, использовать беспроводные идентификационные карты и проводить бесконтактные транзакции одним касанием, что делает жизнь пользователя более удобной и простой. Поскольку технология работает на очень коротком расстоянии и не нуждается в поддержке сторонних технологий, она преподносится как очень безопасная [3]. Однако, с ростом числа устройств и карт, поддерживающих технологию NFC, эксперты все чаще высказывают опасения по поводу данного утверждения.

Принцип работы NFC

Технологии RFID делятся по диапазону частот, в котором происходит взаимодействие. Несущая частота определяет физические величины, в числе которых скорость передачи данных, максимальное расстояние между приемником и передатчиком, помехоустойчивость, а также особенности реализации протоколов прикладного уровня и более высокого уровня, эксплуатационные преимущества и ограничения оборудования, работающего в соответствующем диапазоне (низкочастотном, высоком или сверхвысоком) [13].

NFC – это технология малого радиуса действия, основанная на RFID и работающая на частоте 13,56 МГц, с передачей данных до 6780 кбод в секунду. Связь запускается, когда два NFC-совместимых устройства оказываются в непосредственной близости друг от друга на расстоянии около десяти сантиметров. При этом сохраняется полная совместимость со стандартными смарт-картами стандарта ISO / IEC 14443 [1], что впоследствии повлияло на широкое распространение мобильных устройств, поддерживающих данную технологию.

Связь в NFC осуществляется либо в активном, либо в пассивном режиме. Активный режим – это режим, в котором устройство генерирует радиочастотный сигнал и имеет свой собственный источник питания. В пассивном режиме устройство питается от другого активного устройства [18]. Кроме того, существуют два способа связи: двусторонний и односторонний [19]. В случае двусторонней связи устройства способны взаимодействовать друг с другом. К примеру, используя NFC, вы можете прикосновением передать между двумя устройствами такие данные, как контакты, ссылки или фотографии. Для односторонней связи чтение и запись данных на чип NFC осуществляется с помощью устройства с питанием (например, телефоном, считывателем кредитных карт или терминалом пригородных карт). Примером данного взаимодействия может являться установление связи между меткой NFC и устройством с поддержкой NFC (к примеру, считывание смартфоном ссылки с NFC-ярлыка), платежной картой NFC и POS-терминалом [1].

Кроме того, активные устройства, поддерживающие NFC, могут находиться в одном из трех режимов, основанных на стандартах ISO/IEC 18092 NFC IP-1 [14] и ISO/IEC 14443.1 [1]:

- Режим считывателя/записывающего устройства: в режиме считывания/записи устройство с поддержкой NFC считывает NFC-метки, такие как бесконтактные смарт-карты и RFID метки, обнаруженные в непосредственной близости. После обнаружения метки записывающее устройство может считать данные с обнаруженной метки или записать новые данные.

- Режим одноранговой сети: два устройства, на которые подается питание, могут работать в одноранговом режиме. Этот режим позволяет двум устройствам общаться как объединенным в одну сеть [20].

- Режим эмуляции карты: само активное устройство выступает в качестве NFC-метки, переводя себя в режим пассивной связи. В данном режиме устройство не генерирует собственное радиочастотное поле, вместо него это поле создает считыватель NFC. Эмулированная карта NFC карта может быть доступна внешнему считывателю NFC, такому как терминал NFC в точке продаж (POS).

С каждым годом использование устройств с поддержкой NFC для бесконтактной оплаты растет, в частности за время пандемии использование наличных денег уменьшилось, в то время как количество бесконтактных платежей выросло. По данным ЦБ РФ в 2020 году количество операций в России, совершенных эмитированными картами, возросло на 61,08% по сравнению с 2018 годом и на 19,52% с 2019 года, и продолжает расти в настоящее время [2].

Процесс осуществления оплаты

Бесконтактные платежи могут совершаться либо с использованием банковской карты, либо с использованием устройства, эмулирующего банковскую карту. Процесс оплаты виртуальной картой мало отличается от оплаты физической – карта-токен также содержит срок действия, идентификатор запросчика, криптограмму токена и PAN-номер. На рисунке 1 показан процесс транзакции с бесконтактного устройства.

На первом шаге карта или устройство прикладывается к POS терминалу для передачи необходимой информации с карты. Далее полученные данные POS терминал по сети отправляет в банк-эквайер, который предоставляет продавцу услугу оплаты при помощи банковских карт. Полученные данные банк-эквайер отправляет в международную платежную систему для подтверждения возможности платежа у банка-эмитента, выпустившего карту. При подтверждении покупки происходит списание средств с карты [8].

Безопасность платежей с физических карт обеспечивает SE (Secure Element) – устойчивая к взлому платформа (обычно защищенный микроконтроллер), способная безопасно размещать приложения и их конфиденциальные и криптографические данные (например, криптографические ключи) в соответствии с правилами и требованиями безопасности, установленными определенными доверенными органами. SE обеспечивает безопасную область для выполнения приложений и защиты платежных активов (например, платежных данных, ключей, кода платежного приложения) [5]. Помимо использования для платежных приложений, платформа SE также может быть задействована в процессе аутентификации, а также для хранения приложений, не связанных с платежами, которые требуют механизмов безопасности [6].

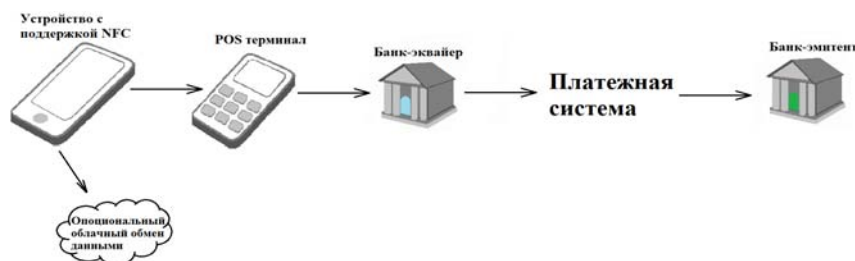


Рис. 1. Процесс бесконтактной транзакции

Для обеспечения безопасности платежей на смартфонах используется технология HCE (host card emulation), позволяющая не встраивать SE внутрь устройства, а использовать виртуальное представление различных электронных идентификационных (пропускных, транзитных и банковских) карт, используя только программное обеспечение.

Уязвимости NFC технологии

Устройство с поддержкой NFC посылает специальный код в платежную систему розничного продавца, которая отправляет пользователю данные о транзакции. Пользователю необходимо ввести PIN-код для одобрения конкретной транзакции или, к примеру, разблокировать смартфон. Чтобы оставаться защищенным от вирусов и хакеров, на устройстве с поддержкой NFC только одна программа способна одобрить транзакцию, к тому же изолировано от остальной операционной системы телефона. Эта особенность делает NFC достаточно безопасным способом оплаты, однако, не лишенным нескольких уязвимостей. Рассмотрим некоторые из них.

- **Подслушивание.** Радиочастотный сигнал принимается антенной для беспроводной передачи данных. Расстояние, с которого злоумышленник может подслушать сигнал, зависит от различных параметров, но обычно составляет несколько метров. Кроме того, на подслушивание сильно влияет режим связи. Пассивное устройство, которое не генерирует собственное радиочастотное поле, подслушать гораздо сложнее, чем активное устройство.

- **Возможность обхода подтверждения личности владельца карты на его устройстве (CDCVM).** Для очень крупных сумм устройство с приложением для оплаты требует CDCVM, но для мелких транзакций устройство просто должно иметь активный экран. Это позволяет совершать электронные кражи из кошелька жертвы до тех пор, пока на устройстве активны NFC и экран. Экран может быть активирован любым способом: запрос соединения Bluetooth, звонок на телефон или нажатие клавиш регулировки громкости. Путем проксирования данных между устройством и терминалом можно провести транзакцию, превышающую лимит CDCVM [12]. Данная уязвимость характерна для карт Visa, для которых характерно использование CTQ (Card Transaction Qualifiers, объект, описывающий возможности карты), через которые определяются способы верификации владельца карты, в том числе и CDCVM.

- **Человек посередине.** Данная форма уязвимости NFC подразумевает, что данные между двумя сторонами могут перехватываться третьей стороной. Третья сторона действует как ретранслятор, используя полученную информацию и изменяя ее при необходимости, чтобы позволить злоумышленнику достичь своих целей. Это должно быть выполнено так, чтобы обе стороны не знали, что между ними находится перехватчик. Однако, атаку типа "человек посередине" достаточно сложно осуществить при использовании технологии NFC, в частности из-за малого расстояния передачи данных.

- **Pre-play атака на бесконтактные банковские карты с чипом.** В области инфокоммуникационной безопасности атака с предварительным воспроизведением - это криптографическая атака, к которой злоумышленник готовится заранее, выполняя имитацию транзакции, притворяясь устройством, которое будет атаковано, а затем повторяет атаку второй раз с реальным устройством в то время, когда оно, скорее всего, выполнит ту же серию операций, что и при имитации. Целостность транзакции обеспечивается UN (Unpredictable Number, генерируемое терминалом случайное число), которое является уникальным, и АТС (Application Transaction Counter, счетчик транзакций), который увеличивается при последующих транзакциях. Но на практике эмитенты не отклоняют транзакции со значением АТС ниже текущего. Это делается для того, чтобы разрешить транзакции, которые могли быть совершены в условиях отсутствия сети. Также нет никаких ограничений, препятствующих тому, чтобы терминал каждый раз посылал на карту один и тот же UN [12].

- **Повреждение данных.** Это, по сути, форма DOS-атаки. Злоумышленник может попытаться прервать связь, передавая данные, которые могут быть недействительными или, блокируя канал таким образом, что передаваемые данные будут повреждены. Прослушивание данных во время их передачи позволит обнаружить любую атаку такой формы, поскольку для успешной атаки мощность, требуемая системой, значительно превышает мощность, которая может быть обнаружена устройством NFC, передающим данные.

Конфиденциальность, целостность и доступность являются основными требованиями к любой эффективной системе. Безопасная система должна гарантировать передачу данных безопасным способом. Таким образом, она предотвращает несанкционированные манипуляции или доступ к данным. Конфиденциальность предотвращает несанкционированный доступ к данным, в то время как целостность предотвращает данные от несанкционированной модификации, а доступность — это время, в течение которого данные доступны. Любое нарушение этих трех требований приведет к нарушению безопасности системы [15-17,22,23].

Риски при использовании NFC-платежей

Технология NFC распространяется все шире в финансовой сфере. Но ввиду наличия уязвимостей, существуют определенные риски при её использовании. Рассмотрим некоторые из них, возникающие при бесконтактной оплате физической или виртуальной картой:

Риск перехвата данных. Если злоумышленник перехватывает сигнал, передаваемый между картой и считывателем, то, теоретически, он получает доступ к информации о карте этого человека. Кроме того, под угрозой оказываются другая личная информация, перехватываемая при передаче [4,7]. Канал NFC полагается на шифрование на уровне приложения для всей криптографической безопасности. Дополнительной мерой по обеспечению безопасности может служить добавление асимметричности [10].

Риск изменения передаваемых данных. Третья сторона перехватывает передаваемый сигнал и повреждает или модифицирует его, после чего отправляет его дальше [4]. Иногда злоумышленнику нужна не информация с карты, а предотвращение передачи правильной информации. Для защиты от модификации данных NFC устройства должны проверять радиочастотное поле во время отправки данных [9].

Риск выхода NFC-устройства или физической карты из строя. В случае выхода из строя SE внутри пластиковой карты (вследствие физического повреждения) или NFC-антенны устройства (например, отсутствие питания) выполнить бесконтактную транзакцию становится невозможно. Кроме того, NFC-считыватель может быть подвергнут DOS-атаке, когда поступает непрерывный запрос, к примеру, с вредоносного приложения в мобильном телефоне, или к нему прикасаются пустой меткой, что влечет за собой появление сообщений об ошибках, которые могут повлиять на переход устройства NFC или сервисов в состояние приостановки [7]. В NFC отсутствуют какие-либо реальные контрмеры против повреждения данных, что связано с трудностью предотвращения радиопомех в целом. Однако, NFC является протоколом малого радиуса действия, поэтому потенциальный злоумышленник должен быть избирателен в своих действиях. Поскольку NFC обычно действует на расстоянии всего нескольких сантиметров во время нормальной работы, его сложнее заглушить, чем более дальние коммуникации с более слабыми сигналами [11].

Проблема несанкционированного доступа к устройству NFC, лежащему в кармане пользователя, должна решаться дополнительными мерами, например, запросом пароля для инициирования обмена, или использованием для этих целей дополнительных методов биометрической аутентификации.

Заключение

Технология NFC в настоящее время достаточно стандартизирована и совместима, что позволяет использовать её для бесконтактных платежей не только на физических картах, но и на мобильных устройствах.

С помощью NFC-платежей транзакции осуществляются одним касанием.

Связь в сантиметровом диапазоне может показаться более безопасной, но подобные технологии очень уязвимы для атак, поскольку практически не требуют дополнительной аутентификации, что, в свою очередь, может повлечь риски утечки данных.

Данные риски связаны с незашифрованной передачей данных, слабой защитой аппаратных средств или программного обеспечения. По этой причине необходимо своевременно выявлять потенциальные риски и вырабатывать контрмеры для обеспечения минимизации их вреда.

Литература

1. ISO/IEC 14443-1:2018 Cards and security devices for personal identification – Contactless proximity objects – Part 1: Physical characteristics // URL: <https://www.iso.org/standard/73596.html> (дата обращения: 25.02.2022).
2. Статистика национальной платежной системы // URL: <https://www.cbr.ru/statistics/nps/psrf/> (дата обращения: 03.03.2022).
3. EMV® Payment Tokenisation // URL: <https://www.emvco.com/emv-technologies/payment-tokenisation/> (дата обращения: 03.03.2022).
4. *Manmeet Mahinderjit Singh, Ku Aina Afiqah Ku Adzman, Rohail Hassan.* Near Field Communication (NFC) Technology Security Vulnerabilities and Countermeasures // *International Journal of Engineering & Technology*, 7 (4.31) (2018) 298-305.
5. *Reveilhac M. & Pasquet M.* Promising secure element alternatives in NFC architecture, Communication IEEE International Workshop on Near Field, 2009.
6. *Shobha, N. S. S., Aruna, K. S. P., Bhagyashree, M. D. P., & Sarita, K. S. J.* NFC and NFC payments: A review. 2016 International Conference on ICT in Business Industry & Government (ICTBIG). doi:10.1109/ictbig.2016.7892683.
7. *Albattah, A., Alghofaili, Y., & Elkhediri, S.* NFC Technology: Assessment Effective of Security towards Protecting NFC Devices & Services. 2020 International Conference on Computing and Information Technology (ICCIT-1441).
8. *MA Masyuk.* Information security of RFID and NFC technologies. 2019 J. Phys.: Conf. Ser. 1399 033093.
9. *Nelson, Dakota & Qiao, Mengyu & Carpenter, Andrew. (2013).* Security of the near field communication protocol: an overview // *Journal of Computing Sciences in Colleges*. 29. P. 94-104.
10. *Kim, Heon-june.* A Study on the Cryptographic Algorithm for NFC. *Indian Journal of Science and Technology*. 9. 10.17485/ijst/2016/v9i37/102543.
11. *Albattah, A., Alghofaili, Y., & Elkhediri, S.* NFC Technology: Assessment Effective of Security towards Protecting NFC Devices & Services. 2020 International Conference on Computing and Information Technology (ICCIT-1441). doi:10.1109/iccit-144147971.2020.9213758.
12. *Leigh-Anne Galloway, Tim Yunusov.* First Contact – Vulnerabilities in Contactless Payments // *Black Hat Europe Conference 2019*, 04.12.2019.
13. *Тимошук Ю.С., Маклачкова В.В.* Риски применения RFID-технологии в медицинских учреждениях. // *Телекоммуникации и информационные технологии*. 2021. Т. 8. № 2. С. 80-84.
14. *ISO/IEC 18092:2013 Information technology – Telecommunications and information exchange between systems – Near Field Communication – Interface and Protocol (NFCIP-1)* // URL: <https://www.iso.org/standard/56692.html> (дата обращения: 17.03.2022).
15. *Maklachkova V.V., Dokuchaev V.A., Statev V.Y.* Risks identification in the exploitation of a geographically distributed cloud infrastructure for storing personal data // 2020 International Conference on Engineering Management of Communication and Technology, EMCTECH 2020 - Proceedings, Vienna, 20-22 октября 2020 года. Vienna, 2020. P. 9261541. DOI 10.1109/EMCTECH49634.2020.9261541
16. *Dokuchaev V.A., Maklachkova V.V., Statev V.Yu.* Classification of personal data security threats in information systems // *T-Comm*. 2020. Vol. 14. No 1. P. 56-60. DOI 10.36724/2072-8735-2020-14-1-56-60.
17. *Докучаев В.А., Маклачкова В.В., Статьев В.Ю.* Требования к информационным системам при работе с «цифровым образом» субъекта // III Научный форум телекоммуникации: теория и технологии ТТТ-2019: Материалы XXI Международной научно-технической конференции, Казань, 18-22 ноября 2019 года. Казань: Казанский государственный технический университет им. А.Н. Туполева, 2019. С. 296-297.
18. *Д. В. Гадасин, А. В. Шведов, О. Г. Клыгина, Д. Д. Гадасин.* Реализация платформы туманных вычислений для предоставления сервисов IoT // *REDS: Телекоммуникационные устройства и системы*. 2021. Т. 11. № 2. С. 65-75.
19. *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.

20. *Гадасин Д.В., Шведов А.В.* Проблемы интеграции концепции "интернет вещей" и облачных вычислений // Технологии информационного общества: Материалы XIII Международной отраслевой научно-технической конференции, Москва, 20-21 марта 2019 года. М.: Издательский дом Медиа Паблишер, 2019. С. 22-23.
21. *Докучаев В.А., Маклачкова В.В., Статъев В.Ю.* Идентификация субъекта – ключевой момент в процессе обработки персональных данных // Технологии информационного общества: Сборник трудов XIV Международной отраслевой научно-технической конференции, Москва, 18-19 марта 2020 года. М.: Издательский дом Медиа Паблишер", 2020. С. 273-274.
22. *Мазуркевич Д.О., Орлов В.Г.* Эволюция систем безопасности сетей сотовой связи разных поколений // Т-Comm: Телекоммуникации и транспорт. 2011. Т. 5. № 1. С. 38-40.
23. *K. Lopatina, V. A. Dokuchaev and V. V. Maklachkova.* Data Risks Identification in Healthcare Sensor Networks, 2021 International Conference on Engineering Management of Communication and Technology (EMCTECH), 2021, pp. 1-7, doi: 10.1109/EMCTECH53459.2021.9619178.
24. *Докучаев В.А., Маклачкова В.В., Статъев В.Ю.* Цифровизация субъекта персональных данных // Т-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 6. С. 27-32.
25. *Pavlov S.V., Dokuchaev V.A., Maklachkova V.V., Mytenkov S.S.* Features of supporting decision making in modern enterprise infocommunication systems // Т-Comm. 2019. Т. 13. № 3. С. 71-74.
26. *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.
27. *Докучаев В.А., Кальфа А.А., Мытенков С.С., Шведов А.В.* Анализ технических решений по организации современных центров обработки данных // Т-Comm: Телекоммуникации и транспорт. 2017. Т. 11. № 6. С. 16-24.
28. *Pavlov S.V., Dokuchaev V.A., Mytenkov S.S.* Model of a fuzzy dynamic decision support system // Т-Comm. 2020. Т. 14. № 9. С. 43-47.
29. *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.

RISK ANALYSIS OF WIRELESS PAYMENT USING NFC

Julia S. Timoschuk

Graduate MTUCI, Moscow, Russia,
iul.tim2012@yandex.ru

Victoria V. Maklachkova

Director for scientific NO "ACEM", Senior Lecture, MTUCI, Moscow, Russia,
v.v.maklachkova@mtuci.ru

Abstract

NFC payments are becoming an increasingly common way of paying for everyday needs. NFC is a short-range RFID technology for data transmission security. However, the technology has a number of vulnerabilities that can put not only users' money but also their personal data at risk. This article discusses the current problems of NFC payments and possible solutions.

Keywords: risk, vulnerability, radio-frequency identification, NFC, RFID, data security, contactless payments.

АНАЛИЗ СЦЕНАРИЕВ РАСПРЕДЕЛЕНИЯ РЕСУРСА В СЛАЙСАХ ПРИ ГРУППОВОМ ПОСТУПЛЕНИИ ЗАЯВОК

Ассума Туль Лудениату,
магистрант МТУСИ, Москва, Россия,
loudeniathassouma@gmail.com

Степанов Сергей Николаевич,
Заведующий кафедрой ССисК, д.т.н., МТУСИ, Москва, Россия,
stpvnvsrc@gmail.com

Аннотация

Интернет Вещей (Internet of Things, IoT) на сегодняшний день является одной из динамично развивающихся технологических отраслей. В работе излагаются принципы распределения ресурса при обслуживании потоков заявок устройств IoT, основанных на технологии NB IoT. С использованием результатов теории марковских случайных процессов построена математическая модель формирования и обслуживания потоков заявок сессий связи устройств IoT. Поступление заявок подчиняется групповой модели пуассоновского потока. Время обслуживания заявки имеет экспоненциальное распределение. Построен рекурсивный алгоритм оценки вероятностей состояний модели. Разработана процедура оценки объема ресурса, требуемого для обслуживания заданных потоков трафика с необходимым качеством. Приведены численные примеры, иллюстрирующие особенности реализации построенных расчетных процедур.

Ключевые слова: Интернет Вещей, network slicing, NB IoT, групповое поступление заявок, марковский процесс

1. Оператор систем наблюдения

Рассмотрим процесс планирования ресурса передачи информации, необходимого для качественной работы оператора систем наблюдения. Анализируемая система включает в себя некоторое количество видеочкамер, реализующих функции видеонаблюдения, достаточно большое число датчиков, осуществляющих функции контроля за температурой, давлением, вторжением и т.п., и аналитические центры, где на основе результатов наблюдения принимаются решения, предназначенные для создания необходимого уровня безопасности работы технических систем. Схема работы оператора систем наблюдений представлена на рисунке 1.

Распределение телекоммуникационного ресурса между поступающими информационными потоками без учета особенностей их формирования по принципу единой мерки для всех «one-size-fits-all» уже не устраивает рынок связи [1]. Для создания условий по обслуживанию многочисленных коммуникационных приложений типа M2M и качественной доставки мультимедийного контента в беспроводных сетях требовалось изменить подход к использованию имеющегося ресурса передачи. Соответствующая концепция была разработана и носит название «network slicing». Она включает ряд процедур, позволяющих разделить имеющуюся физическую сетевую архитектуру на несколько независимых логических сетей, каждая из которых конфигурируется исходя из требований к качеству обслуживания коммуникационных приложений, использующих эти сети [1, 2, 10-19]. Составными частями концепции являются: набор сетевых функций (*network functions*), к которым относится, например, балансировщик нагрузки, использующийся как отдельный модуль при компоновке слайса; метод виртуализации «*virtualization*», рассматриваемый как возможность создания логически независимых сетей на одном физическом ресурсе, и система оркестровки «*orchestration*», которая представляет из себя процедуру координации компонентов, участвующих в жизненном цикле каждого слайса.

Для обоснованного применения сценария распределения ресурса, основанного на «network slicing» необходимо: построение новой математической модели, учитывающей особенности совместного обслуживания гетерогенного трафика; проведение ее исследования с целью построения алгоритма оценки характеристик; разработка примеров реализации модели и средств ее анализа для решения задач планирования объема ресурса слайса и обслуживания заданного потока трафика с требуемым качеством. Решение перечисленных задач будет рассмотрено в последующих разделах работы.



Рис. 1. Функциональная модель формирования входного потока данных IoT в сети оператора систем наблюдений

2. Параметры модели потока сессий связи

Свойства сеанса связи от видеокamer зависят от целого ряда параметров. Прежде всего, необходимо учесть характеристики работы кодека (например, с использованием стандарта H.264), а также способ буферизации данных перед отправкой (например, при помощи методов traffic shaping). В зависимости от требований целевого приложения (в частности, задержки воспроизведения) весь сеанс связи может быть «нарезан» при помощи traffic shaping на отдельные сессии, требующие некоторой минимальной скорости передачи данных. Например, в случае H.264 при передаче потока разрешением 1080P HD со средним уровнем качества и без изменения других настроек, средняя скорость потока от одной камеры составит приблизительно 1,4 Мбит/с. Конечно, это значение зависит от параметров видеопотока, для учета которых может быть использован упомянутое приложение «калькулятор».

Таким образом, сессия представляет собой фрагмент сеанса с постоянной скоростью, порожденный целевым приложением после проведения процедуры gate adaptation. Битрейт камер меняется в широких пределах от сотен килобит до нескольких мегабит в секунду. Время передачи видеоинформации определяется объемом используемого ресурса и может составлять величину от нескольких секунд до нескольких десятков секунд. Число видеокamer на одну соту может меняться от нескольких десятков до нескольких сотен. Интенсивность поступления заявок на передачу сессий можно связать с числом подключенных видеокamer.

Обозначим их количество через N_v , а через γ_v обозначим интенсивность поступления сессий от одной камеры. Тогда в силу высказанных предположений поступление заявок на передачу сессий от видеокamer подчиняется пуассоновскому закону с интенсивностью $\lambda_v = N_v \gamma_v$. Отметим, что учет числа камер носит весьма условный характер и требует от камер образования однородного семейства с близкими характеристиками. Если этого не удастся сделать в рамках одного потока, то можно разделить все видеокamer на несколько отдельных групп, каждая из которых порождает свой пуассоновский поток с отличными от других потоков параметрами использования ресурса передачи.

Рассмотрим свойства сессий связи от устройств NB-IoT. По рекомендациям 3GPP базовая станция LTE обслуживает около 50 тысяч устройств NB-IoT. Приведенные цифры соответствуют случаю, когда каждое устройство в среднем передает в час примерно 100 байт информации. В качестве ориентировочных значений в спецификациях стандарта NB-IoT указаны максимальные значения объема передаваемого сообщения порядка 1280 байт и скорости в направлении к базовой станции порядка 20 кбит/с для режима singletone, и 250 кбит/с для режима multitone. С развитием технологий скорости передачи данных будут увеличиваться. Предположим, что объем передаваемой информации имеет экспоненциальное распределение со средним значением F , выраженным в секундах. Аналогично тому, как это было сделано для потока сессий от видеокamer, частоту поступления заявок на передачу информационных файлов от устройств NB-IoT можно связать с количеством таких устройств. Обозначим их число через N_a , а через γ_a обозначим интенсивность поступления файлов от одной

камеры. Предполагается, что поступление файлов от устройств NB-IoT подчиняется пуассоновскому закону с интенсивностью $\lambda_a = N_a \gamma_a$.

3. Сценарии распределения ресурса

Рассмотрим отдельную соту сети стандарта LTE, представляющую из себя круг с базовой станцией, размещенной в центре. Для сетей стандарта LTE ресурс обслуживания трафика представлен множеством ресурсных блоков (РБ). Для **1,4 МГц – 6 РБ**, для **5 МГц – 25 РБ** и т.д. Скорость передачи информации для **20 МГц** может быть **50 Мбит/с** для uplink соединения и **100 Мбит/с** для downlink соединения. С дальнейшим развитием технологии эти величины достигнут значений **500 Мбит/с** и **1 Гбит/с**, соответственно. Гибкость в распределении ресурса, использование протокола IP на уровне доступа и ядра сети позволяют организовать обслуживание трафика сервисов реального времени и эластичных данных с предоставлением заданных показателей качества приема и передачи информационных потоков.

Воспользуемся возможностями сети стандарта LTE для организации работы оператора систем наблюдений. Предположим, что в соте выполняется обслуживание информационных потоков от обычных потребителей услуг беспроводной связи, и арендуется ресурс для передачи трафика оператора систем наблюдений. Разделение ресурса выполнено с использованием возможностей network slicing. Обозначим через T_0 объем ресурса в РБ, который выделен оператору. Часть из этого ресурса используется для обслуживания трафика видеокамер, а другая часть – для обслуживания трафика устройств IoT по технологии NB IoT. В модели наименьшее требование к ресурсу имеют соединения для обслуживания сессий NB IoT. Обозначим через u скорость, обеспечиваемую одним каналом, а через v обозначим общее число таких каналов.

Процедура network slicing предполагает использование ресурса для обслуживания однородного трафика, порожденного одним типом коммуникационных приложений с близкими по величине требованиями к скорости передачи. В анализируемом случае имеется два типа запросов на информационное обслуживание: от видеокамер и от устройств NB IoT [3]. Оба типа различаются требованиями к ресурсу, поэтому имеет смысл опять воспользоваться процедурой network slicing и разделить ресурс между этими двумя типами трафика. При этом будут исследованы более общие модели, в отличие от рассмотренных в данном разделе. Так, предполагается, что файлы от устройств NB IoT будут поступать группами случайного объема, а поступление сессий видеокамер будет задаваться произвольным числом потоков с индивидуальными требованиями к ресурсу передачи и общим характером функции распределения времени обслуживания сессии.

Жесткое разделение ресурса позволяет выровнять показатели обслуживания сессий, однако при этом ухудшается эффективность его использования. Отрицательные последствия этого явления можно отчасти уменьшить, если имеющийся ресурс использовать совместно, но при этом ограничить доступ для части сессий, чтобы достичь требуемых значений потерь поступающих запросов.

Достаточность ресурса в слайсе для обслуживания поступающего потока запросов на информационное обслуживание оценим долей потерянных запросов, а эффективность использования ресурса – коэффициентом занятия одного виртуального канала [4-6]. Для оценки введенных показателей необходимо построить математическую модель распределения ресурса в слайсах и провести ее исследование с использованием теории марковских процессов и численных методов линейной алгебры. Это исследование выполнено в следующем разделе.

4. Анализ моделей распределения ресурса в слайсах

Построим модель входного потока, учитывающую отмеченные особенности формирования поступающих запросов на информационное обслуживание. В модели Эрланга промежутки времени между появлением заявок распределены экспоненциально. Случайный характер моментов появления заявок означает, что в какие-то интервалы времени они могут поступать чаще, а в какие-то – реже. Однако этой особенности пуассоновского потока ещё недостаточно, чтобы отразить отмеченный выше импульсный характер поступления заявок на передачу файлов, который свойственен работе устройств NB IoT. Построим модель, которая, с одной стороны, отражает групповое поступление файлов, а с другой – сохраняет возможность оценки характеристик качества передачи файлов с вычислительной сложностью, сопоставимой с расчетами по модели Эрланга.

Перейдем к обсуждению основных этапов построения математической модели распределения канального ресурса, в которой учитывается импульсный характер появления файлов. В отличие от модели Эрланга здесь рассматривается двухуровневая схема реконструкции поступающего потока трафика. В модели имеется v каналов, которые обслуживают пуассоновский поток отдельных групп заявок на передачу файлов интенсивности λ . Интенсивность поступления групп заявок можно представить как функцию числа устройств NB IoT, если применить параметризацию входного потока [4-6].

Обозначим через $f(s)$ вероятность наличия в группе s файлов. Пусть $g = v$. Понятно, что $f(1) + \dots + f(v) = 1$. Объем файла распределен экспоненциально со средним F . Время передачи файла одним каналом распределено экспоненциально с параметром μ . Пусть b среднее число файлов в группе:

$$b = \sum_{s=1}^g f_s s.$$

Качество обслуживания поступающих заявок и процесс занятия каналов передачи информации определяют следующие показатели: π_t доля времени пребывания слайса в состоянии занятости всех каналов; π_c доля потерянных заявок на передачу файлов; π_f доля потерянных потенциальных соединений на передачу файлов и m – среднее число занятых каналов.

Перечисленные характеристики могут быть найдены, если известны значения доли времени $P(i)$ пребывания модели в состоянии с i занятыми каналами. Алгоритм вычисления $P(j)$ состоит из следующих шагов.

1. Положим значение $P(0) = 1$

2. Для значений j , меняющихся последовательно от 1 до v , находим ненормированные вероятности $P(j)$, используя рекурсию [4-6].

$$P(j) = \frac{\lambda}{j\mu} \left(P(0) \sum_{i=j}^v f_i + P(1) \sum_{i=j}^v f_i + \dots + P(j-1) \sum_{i=j}^v f_i \right). \quad (1)$$

3. Определяем нормировочную константу

$$N = \sum_{j=0}^v P(j).$$

4. Вычисляем нормированные вероятности $p(j)$, используя соотношение

$$p(j) = \frac{P(j)}{N}, j = 0, 1, \dots, v.$$

Таким образом, для оценки стационарных вероятностей $p(j)$ достаточно воспользоваться рекурсией (1). С вычислительной точки зрения выполнение этой процедуры не вызывает больших затруднений. Перейдём к определению и расчёту показателей качества обслуживания поступающих заявок.

Доля времени занятости всех каналов слайса π_t находится из соотношения $\pi_t = p(v)$.

Доля π_c потерянных заявок на передачу файла находится из выражения:

$$\pi_c = \frac{\text{Интенсивность потерянных заявок}}{\text{Интенсивность поступивших заявок}}$$

Окончательное выражение для оценки π_c записывается в виде соотношения:

$$\pi_c = \frac{1}{h} \sum_{i=j}^v p(i) \sum_{i=j}^v f_{v-k} (i - k) \quad (2)$$

Для модели с групповым поступлением заявок на передачу файлов рассмотрим решение двух задач:

1. Оценка объема ресурса слайса, достаточного для обслуживания заданной величины трафика NB IoT с требуемым качеством;

2. Оценка максимально возможного числа устройств NB IoT, которые могут быть обслужены на заданном слайсе с требуемым качеством.

Начнем с решения первой из перечисленных задач. Качество обслуживания файлов зададим значением $\pi_c = 0,03$. Пусть $F = 100$ байтов, $u = 100/c$, а $f_i = 1,5$ для всех i .

Предположим, что один датчик посылает файлы с интенсивностью γ , равной 1 файл в минуту. Число устройств NB IoT, примем равным 37500. Из приведенных данных следует, что среднее время передачи файла одним каналом NB IoT определяется из выражения:

$$\frac{1}{\mu} = \frac{F}{u} = 0,008 c = 8 \text{ мс}. \quad (3)$$

Интенсивность предложенного трафика a для всех сценариев одинакова и определяется из равенства, $a = 5$ Эрл. Отсюда находим значение $\lambda = \frac{a\mu}{b} = 208,33$ групп файлов в секунду. Определение требуемой величины ресурса передачи информации выполняется последовательным увеличением числа каналов v и проверкой условия $\pi_c \leq 0,03$. При первом выполнении неравенств поиск величины ресурса оканчивается. Результаты решения задачи показаны на рисунке 2. Искомое значение числа каналов $v = 14$, следовательно, ресурс слайса должен обеспечивать скорость 1,4 Мбит/с.

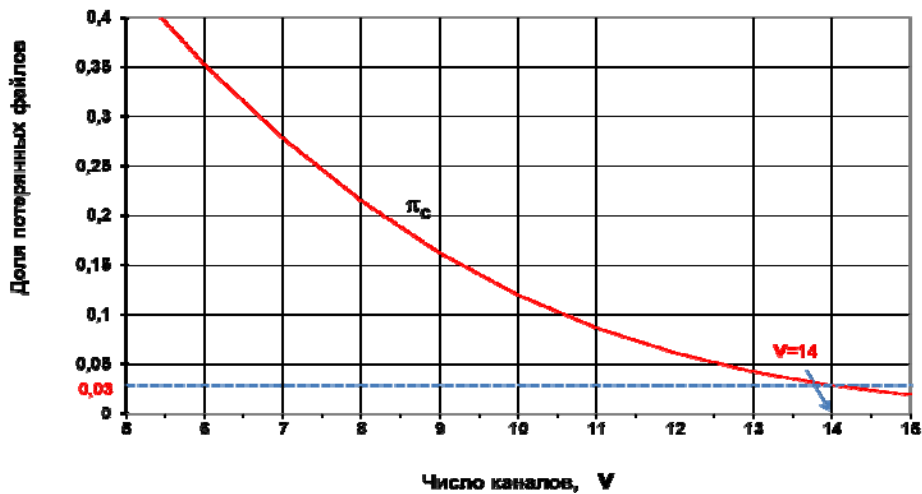


Рис. 18. Оценка объема ресурса слайса, достаточного для обслуживания заданной величины трафика NB IoT с требуемым качеством

Теперь обратимся к решению второй из перечисленных задач. Качество обслуживания файлов зададим значением $\pi_c = 0,03$.

Величины параметров поступления файлов возьмем те же, что были использованы при решении первой задачи: $F = 100$ байтов; $u = 100$ кбит/с; $f_i = \frac{1}{5}$, $i = 1,2,3,4,5$; $\frac{1}{\mu} = 8$ мс;

$\gamma = 1$ файл/мин, за исключением числа виртуальных каналов $v = 15$. Число N_d устройств NB IoT будет переменной величиной. Интенсивность поступления групп файлов $\lambda = \frac{N_d \gamma}{b_m}$.

Оценка максимально возможного числа устройств NB IoT, которые могут быть обслужены на заданном слайсе с требуемым качеством, выполняется последовательным увеличением числа устройств NB IoT N_d и проверкой $\pi_c \leq 0,03$. При первом выполнении противоположного неравенства поиск N_d оканчивается. Результаты решения задачи показаны на рисунке 3. Искомое значение числа устройств NB IoT не должно превосходить $N_d < 43000$.

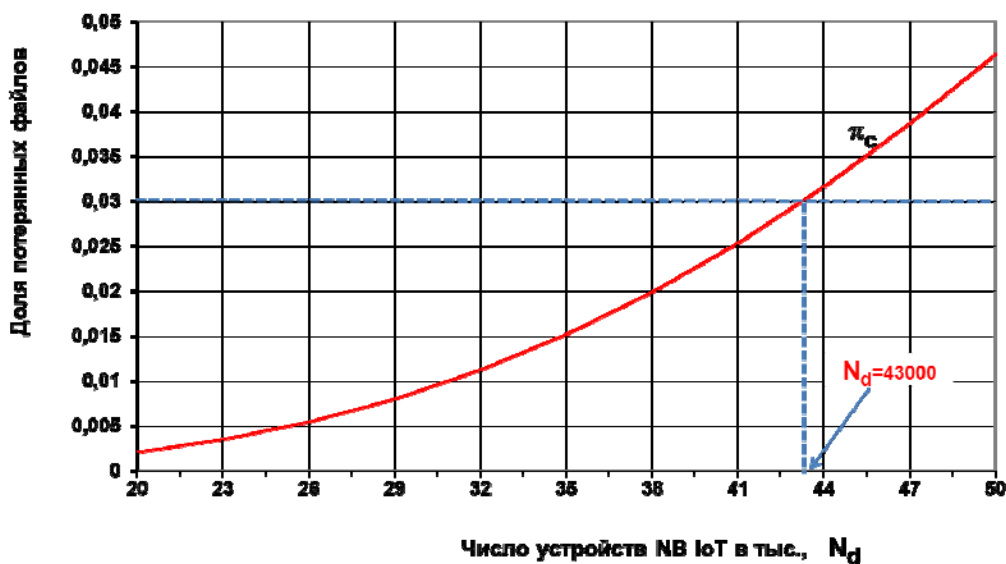


Рис. 19. Оценка максимально возможного числа устройств NB IoT, которые могут быть обслужены на заданном слайсе с требуемым качеством

Приведенные примеры свидетельствуют о возможности использования построенной модели и алгоритмов оценки ее характеристик для решения задачи планирования объема ресурса слайса для обслуживания заданного потока трафика с требуемым качеством, а также оценки максимально возможного числа устройств NB IoT, которые могут быть обслужены на заданном слайсе с требуемым качеством [7].

Расчет характеристик не вызывает затруднений и может быть произведен для любых значений структурных параметров модели, представляющих интерес с точки зрения практических приложений.

Заключение

В работе рассмотрены вопросы формирования потоков заявок на открытие сессий связи от устройств оператора систем наблюдений, представляющих из себя видеокамеры и другие подобные им средства фиксации информации. Предполагается, что заявки поступают группами в соответствии с пуассоновской моделью с фиксированной интенсивностью. Число заявок в группе задается с вероятностью, сумма которых равна единице. Время обслуживания каждой сессии имеет экспоненциальное распределение. С использованием построенной модели сформулированы определения характеристик обслуживания сессий. Среди них: доля потерянных заявок и средний объем используемого ресурса передачи. Построенный рекурсивный алгоритм оценки вероятностей стационарных состояний модели может быть использован для оценки введенных характеристик для любых значений структурных параметров: объема ресурса и размеров групп, поступающих заявок. Приведены примеры использования построенной модели и алгоритмов оценки ее характеристик для решения задач планирования требуемого объема ресурса слайса при обслуживании заданного потока трафика с требуемым качеством, а также оценки максимально возможного числа устройств NB IoT, которые могут быть обслужены на заданном слайсе с требуемым качеством. Результаты исследования могут быть обобщены на случай моделей с мультисервисными заявками [7-9].

Литература

1. Росляков А.В., Ваняшин С.В., Гребешков А.Ю., Самсонов М.Ю. Интернет Вещей, под ред. А.В. Рослякова. Самара: ПГУТИ, ООО «Издательство Ас Гард», 2014. 342 с.
2. Официальный сайт 3rd Generation Partnership Project (3GPP). www.3gpp.org.
3. Архитектура интернета вещей / пер. с англ. М. А. Райтмана. М.: ДМК Пресс, 2019. 454 с.
4. Степанов С.Н. Основы телетрафика мультисервисных сетей. М.: Эко – Трендз, 2010. 392 с.
5. Степанов С.Н. Теория телетрафика: концепции, модели, приложения. М.: Горячая линия – Телеком, 2015. 868 с.
6. Пиеничников А.П. Теория телетрафика. М.: Горячая линия – Телеком, 2017. 212 с.
7. Степанов С.Н., Степанов М.С. Планирование ресурса передачи при совместном обслуживании мультисервисного трафика реального времени и эластичного трафика данных // Автоматика и телемеханика. 2017. № 11. С. 79-93.
8. Степанов С.Н., Степанов М.С. Планирование ресурса передачи информации соединительных линий мультисервисных иерархических сетей доступа // Автоматика и телемеханика. 2018. № 8. С. 66-80.
9. Степанов С.Н., Степанов М.С. Эффективный алгоритм оценки требуемого объема ресурса беспроводных систем связи при совместном обслуживании гетерогенного трафика устройств интернета вещей // Автоматика и телемеханика. 2019. № 11. С. 108-126.
10. Stepanov S.N., Ndayikunda Ju., Kanishcheva M.G. Resource allocation model for ITE technology with functionality of NB-IOT and reservation // T-Comm. 2021. Т. 15. № 11. С. 69-76.
11. Umer M.A., Stepanov S.N., Ndayikunda J., Kanishcheva M.G. Cellular network resource distribution methods for the joint servicing of real-time multiservice traffic and grouped iot traffic // T-Comm. 2020. Т. 14. № 10. С. 61-69.
12. Stepanov S.N., Shishkin M.O., Stepanov M.S., Zhurko H.M. The construction and analysis of call-center model in overload traffic condition // T-Comm. 2020. Т. 14. № 7. С. 42-50.
13. Степанов С.Н., Васильев А.П. Построение и анализ модели с динамическим распределением канального ресурса при обслуживании мультисервисного трафика в сетях IMS/LTE // T-Comm: Телекоммуникации и транспорт. 2019. Т. 13. № 1. С. 29-36.
14. Stepanov S.N., Shishkin M.O., Sosnovikov G.K., Stepanov M.S., Vorobeychikov L.A., Zhurko H.M. The analysis of call center model in case of overload // T-Comm. 2019. Т. 13. № 10. С. 68-76.
15. 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.
16. Степанов С.Н., Васильев А.П. Построение и анализ модели совместного обслуживания линией доступа трафика реального времени и эластичного трафика данных // T-Comm: Телекоммуникации и транспорт. 2018. Т. 12. № 2. С. 55-61.
17. Степанов С.Н., Романов А.М. Математическая модель линии доступа при обслуживании трафика реального времени и эластичного трафика данных // T-Comm: Телекоммуникации и транспорт. 2018. Т. 12. № 4. С. 62-67.
18. Степанов С.Н., Романов А.М. Математическая модель линии доступа при обслуживании трафика реального времени и эластичного трафика данных // T-Comm: Телекоммуникации и транспорт. 2017. Т. 11. № 9. С. 74-79.
19. Васильев А.П., Степанов С.Н. Построение и анализ математической модели с ожиданием и динамическим распределением канального ресурса при групповом поступлении запросов на передачу данных // Научные технологии в космических исследованиях Земли. 2017. Т. 9. № 4. С. 6-12.

**ANALYSIS OF SCENARIOS OF RESOURCE DISTRIBUTION IN SLICES
WITH BATCH APPLICATIONS**

Assouma T. Loudeniatou,
Graduate MTUCI, Moscow, Russia,
loudeniatassouma@gmail.com

Stepanov S. Nikolaevich,
Head of the Department of CS&SC, Doctor of Technical Sciences, MTUCI, Moscow, Russia,
stpvnvsrg@gmail.com

Abstract

The Internet of Things (Internet of Things, IoT) is currently one of the fastest growing technology industries. The paper outlines the principles of resource allocation when servicing application flows of IoT devices built using NB IoT technology. Using the results of the theory of Markov random processes, a mathematical model for the formation and maintenance of application flows for communication sessions of IoT devices is constructed. The receipt of requests obeys the group model of the Poisson flow. The service time of a request has an exponential distribution. A recursive algorithm for estimating the probabilities of model states is constructed. A procedure has been developed for estimating the amount of resource required to service given traffic flows with the required quality. Numerical examples illustrating the features of the implementation of the constructed calculation procedures are given.

Keywords: *Internet of Things, network slicing, NB IoT, batch receipt of requests, Markov process*

РАЗРАБОТКА ЛАБОРАТОРНОГО ПРАКТИКУМА ПО РАЗДЕЛУ «СПЕКТРАЛЬНЫЙ АНАЛИЗ СИГНАЛОВ: ПЕРИОДИЧЕСКИХ, КОНЕЧНОЙ ДЛИТЕЛЬНОСТИ И ДИСКРЕТНЫХ»

Титов Александр Владимирович

студент МТУСИ, Москва, Россия,

cryogenmash97@gmail.com

Григорьева Елена Дмитриевна

доцент кафедры ТЭЦ, к.т.н., МТУСИ, Москва, Россия,

e.d.grigoreva@mtuci.ru

Аннотация

В статье представлены результаты разработки узкоспециализированного пользовательского приложения. В качестве объекта исследования была выбрана задача разработки комплекса виртуальных лабораторных работ для изучения методов спектрального анализа сигналов, применяемых в современном телекоммуникационном оборудовании. В лабораторных работах проводится исследование спектральных характеристик аналоговых и дискретных сигналов конечной длительности, периодических аналоговых и дискретных сигналов, а также амплитудно-модулированных сигналов. В качестве среды программирования выбрано ПО Python в силу того, что данное программное обеспечение обладает требуемой гибкостью. Все расчёты и графики, полученные в созданной программе, проверялись в среде программирования Mathcad.

Ключевые слова: *виртуальная лабораторная работа, программное обеспечение Python, Scilab, свободное программное обеспечение, проприетарное программное обеспечение, лабораторные стенды, спектральный анализ.*

Для обеспечения качественного образования, отвечающего требованиям ФГОС 3++, современный процесс обучения использует передовые инфокоммуникационные технологии, предоставляющие студентам возможность в кратчайшие сроки овладеть новыми знаниями. Таковыми технологиями являются виртуальные лабораторные работы, позволяющие упростить и ускорить процесс изучения нового материала. Основными преимуществами подобного подхода являются наглядность представления учебного материала и гибкость в его применении [7-9].

При изучении раздела «Спектральный анализ» учебный процесс включает в себя лекции, практические занятия и лабораторные работы на стендах, оснащённых измерительными приборами и радиотехническими элементами (рисунок 1). Кроме того, на кафедре имеются персональные компьютеры с программным обеспечением, позволяющим моделировать процессы, происходящие в электрических цепях.

Используемые на кафедре стенды были разработаны преподавателями кафедры Драго Г.Н. и Добровольским И.Н. [1]. Эти стенды в течение многих лет с успехом использовались как в российских университетах, так и в зарубежных. Неизбежно наступает время, когда встроенные в стенд приборы морально и физически устаревают, и возникает задача разработки нового лабораторного практикума.

Организовать проведение лабораторного практикума по теме «Спектральный анализ» возможно на основе следующих методик:

- использование современного лабораторного стенда,
- применение проприетарного бесплатного программного обеспечения *Micro-Cap*,
- использование собственного программного обеспечения на базе ПО *Python*.

Проанализируем преимущества и недостатки этих трёх методик.

1. Лабораторный практикум на базе лабораторных стендов (рисунок 1).

При использовании оборудования лабораторных стендов можно исследовать спектральные характеристики аналоговых сигналов, представляющих собой периодические последовательности импульсов различной формы: прямоугольной, треугольной, пилообразной, срезанной синусоиды, экспоненциальной. Для каждого вида сигналов можно варьировать длительность импульса и его размах. Период следования импульсов в этих экспериментах регулировать не представляется возможным.



Рис. 1. Блоки лабораторного стенда ЛКТЦ для проведения экспериментов по теме «Спектральный анализ»

При этом на экране двухлучевого осциллографа на одном канале можно наблюдать мгновенное значение периодического сигнала $u(t)$, а на втором канале – импульсы, соответствующие его амплитудному спектру (рисунок 2). При этом размах импульса соответствует значению амплитуды гармоники Um , а расположение импульса на горизонтальной оси соответствует частоте гармоники $k \cdot \omega_1$.

Преимущество данного эксперимента перед другими видами исследований заключается в наглядности: перестраивая форму импульса, можно одновременно наблюдать изменение амплитудного спектра.

Недостатком эксперимента является недостаточная информативность: количество определяемых гармонических составляющих ограничено; отсутствует информация о численных значениях амплитуд гармоник и соответствующих им значений частот.

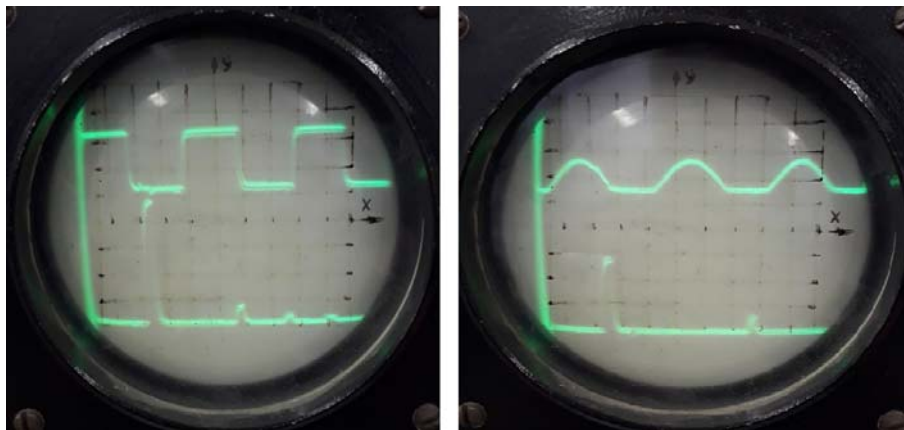


Рис. 2. Экспериментальные исследования спектральных характеристик периодических аналоговых сигналов на базе лабораторного стенда

2. Спектральный анализ на базе ПО *Micro-Cap*.

Для моделирования процессов в электрических цепях уже давно используется проприетарное ПО *Micro-Cap*. Следует подчеркнуть, что ПО *Micro-Cap* доступно на официальном сайте разработчика, распространяется бесплатно и давно и успешно используется студентами. Доступны учебные пособия для выполнения лабораторного практикума по электротехнике, например [2].

При использовании ПО *Micro-Cap* проведение спектрального анализа требует предварительного расчёта гармонических составляющих ряда Фурье периодического сигнала, а затем – формирования длинной записи суммы гармоник.

На рисунке 3 представлено изображение экрана монитора при исследовании периодического сигнала формы срезанной синусоиды. На верхнем графике представлен сигнал, содержащий постоянную составляющую и четыре гармоники ряда Фурье. На нижнем графике средствами ПО *Micro-Cap* представлен спектр амплитуд.

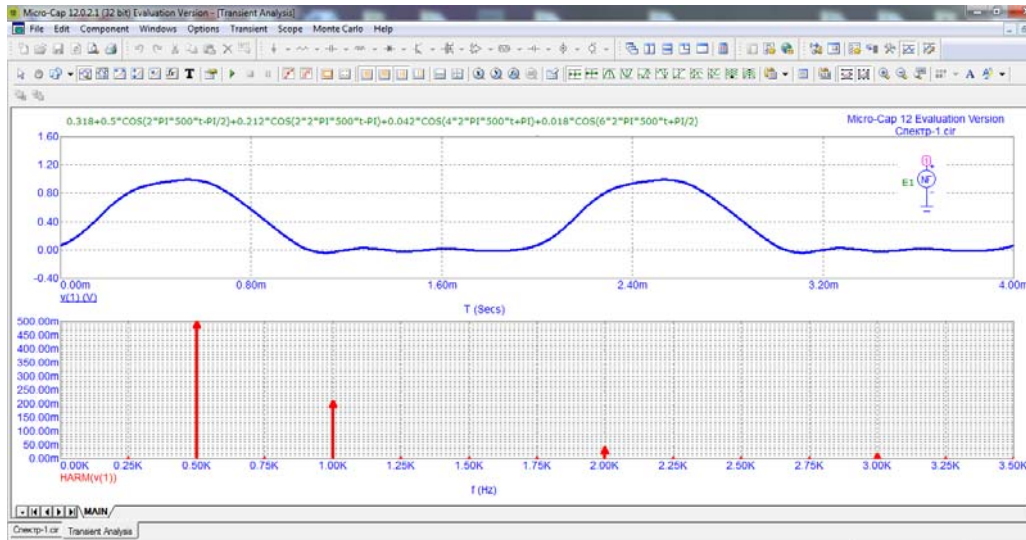


Рис. 3. Спектральный анализ периодической последовательности импульсов формы срезанной синусоиды в среде *Micro-Cap*

Недостатком проведения исследований в среде *Micro-Cap* является продолжительный процесс ввода данных спектра амплитуд периодического сигнала и невозможность сопоставления на одном графике результатов различных исследований при варьировании формы сигнала.

3. Спектральный анализ на базе разработанного пользовательского приложения

В настоящее время на кафедре ТЭЦ МТУСИ выполняется разработка нового современного лабораторного стенда [3], позволяющего проводить как натурные эксперименты, так и эксперименты с применением компьютерного моделирования сигналов, в том числе и дискретных.

Проведение натурального эксперимента по исследованию спектральных характеристик сигналов потребовало бы применения дорогостоящего оборудования. Поэтому по теме «Спектральный анализ» планируется разработать виртуальный лабораторный практикум.

3.1. Применение ПО *Scilab*

В рамках студенческой научно-исследовательской работы был разработан комплекс виртуальных лабораторных работ на базе пакета прикладных программ *Scilab*, который не только содержит сотни математических функций, но и предоставляет пользователям возможность добавлять свои математические функции, написанные на различных языках, таких как C, C++, Fortran и других. *Scilab* является проприетарным бесплатным программным обеспечением.

Разработанный комплекс виртуальных лабораторных работ [4] на базе использования ПО *Scilab* позволяет исследовать спектральные характеристики сигналов различных типов:

- периодических аналоговых и дискретных сигналов;
- аналоговых и дискретных сигналов конечной длительности.

На рисунке 4 представлено окно программы, демонстрирующее результаты исследование спектра амплитуд периодической последовательности импульсов прямоугольной формы и представлена формула, по которой выполняется расчёт комплексных коэффициентов ряда Фурье. Количество рассчитываемых коэффициентов q задаётся. Длительность импульса τ задаётся в секундах. Период следования импульсов T задаётся кратным длительности импульсов. Подобное исследование можно провести с применением лабораторного стенда ЛКТС (рисунок 2).

Недостатком натурального эксперимента является, прежде всего, старение оборудования, отсутствие возможности наблюдать большее количество спектральных составляющих (энергия высших гармоник убывает) и отсутствие возможности менять период следования импульсов из-за конструктивных особенностей стенда.

3.2. Разработка собственного пользовательского приложения

Создание пользовательских приложений позволяет избежать использования проприетарного ПО. Помимо того, последующая разработка аналогичных программ или усовершенствование уже существующих, является простой задачей, так как существует возможность использования ранее созданных *DLL*-библиотек, содержащих основные универсальные функции, в качестве основы для решения новых задач.

В рамках студенческой научно-исследовательской работы было разработано узкоспециализированное пользовательское приложение. В качестве среды программирования избран Python 3.9 в силу того, что данное программное обеспечение обладает требуемой гибкостью. Все расчёты и графики, полученные в созданной программе, проверялись в среде Mathcad.

Для создания программы, позволяющей решить поставленную задачу, был создан новый проект на языке Python 3.9.

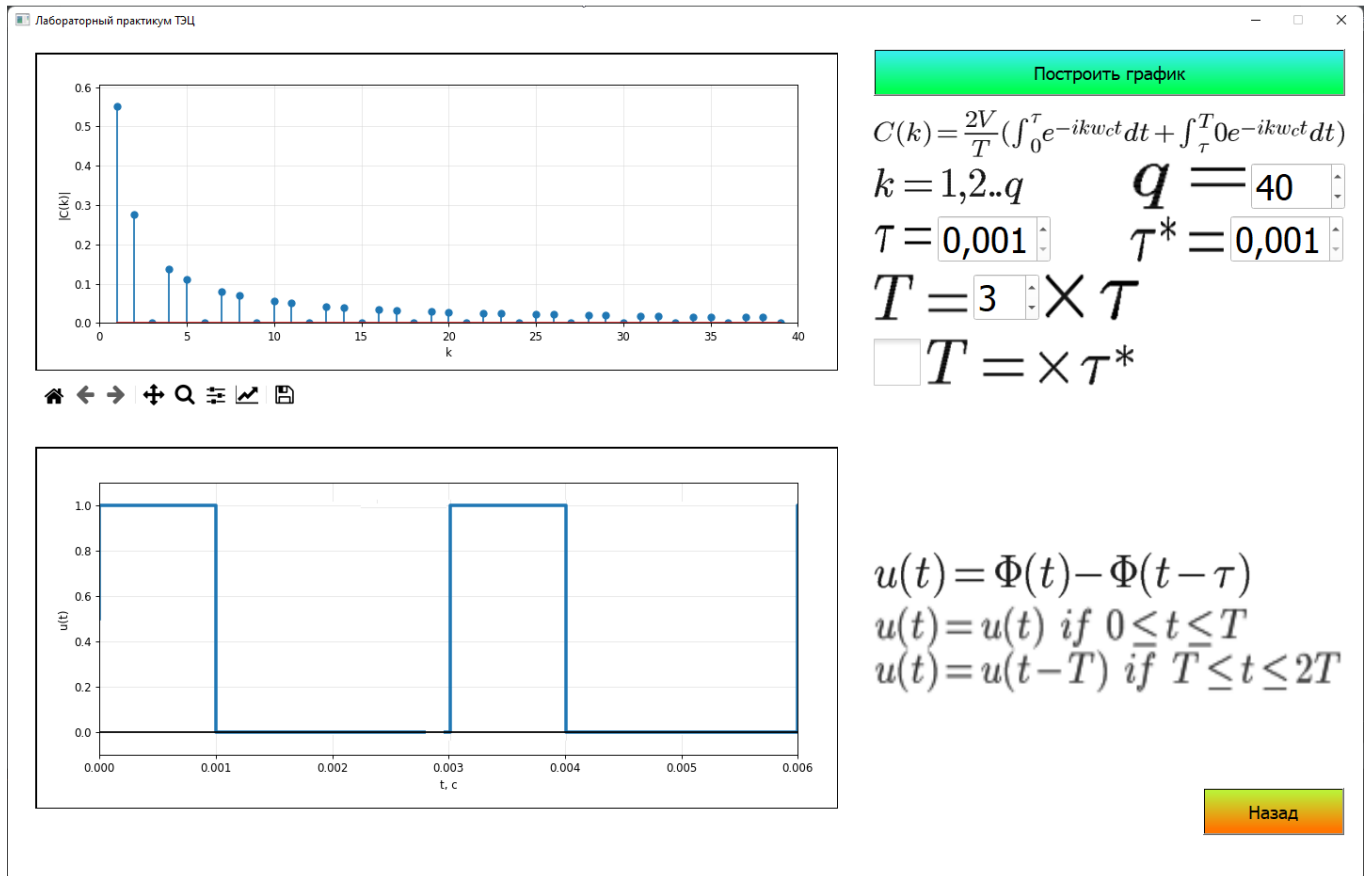


Рис. 4. Исследование спектральных характеристик периодического сигнала

В целях оптимизации процесса разработки была создана библиотека классов, содержащая базовые функции, используемые во всём проекте, а также была применена такая парадигма как объектно-ориентированное программирование. Помимо этого, при разработке программы использовались такие модули как:

- "cmath" – модуль, обеспечивающий доступ к математическим функциям для обработки комплексных чисел.
- "numpy" – библиотека высокоуровневых математических функций (синус, косинус, экспонента функции и др.) или значений констант (число пи и др.).
- "scipy" – библиотека, предназначенная для выполнения научных и инженерных расчётов; в нашем случае мы используем данный модуль для интегрирования.
- "matplotlib" – библиотека для создания статических, анимированных и интерактивных визуализаций; в нашем случае для отображения линейных и дискретных графиков.
- "PyQt5" – это набор Python библиотек для создания графического интерфейса на базе платформы Qt5.
- "sys" – модуль обеспечивает доступ к некоторым переменным, используемым или поддерживаемым интерпретатором, а также к функциям, тесно взаимодействующим с интерпретатором.

Каждое лабораторное исследование сопровождается функциональными модулями (рисунок. 5):

- модуль «Обучение», представляющий собой инструкцию по использованию программы;
- модуль «Практика», содержащий краткую теорию по теме исследования;
- модуль «Зачёт», представляющий собой тест для закрепления изученного материала.

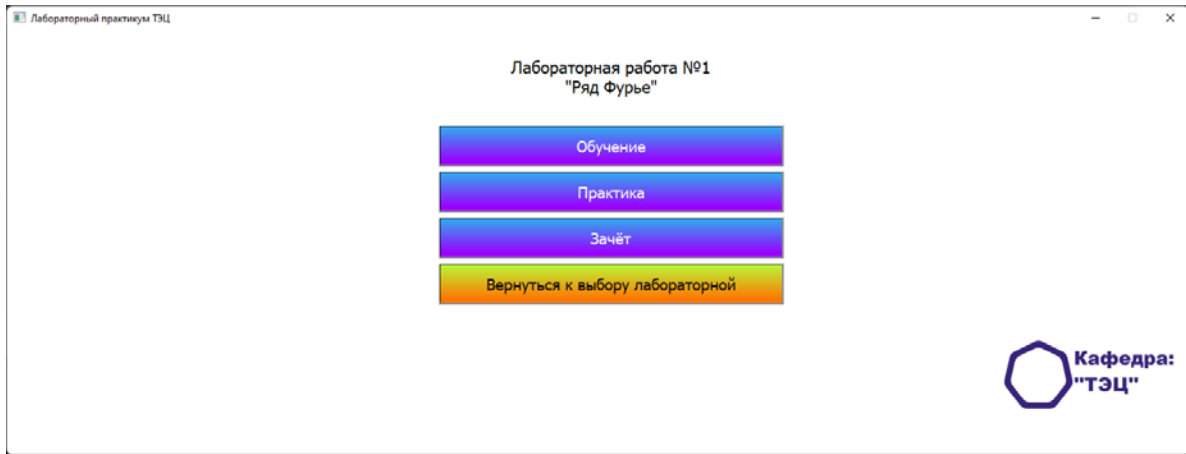


Рис. 5. Интерфейс меню модулей

Под каждым графиком имеется панель инструментов, для того чтобы приблизить/отдалить область, поменять тип линии графика, включить логарифмический масштаб, сохранить график в виде изображения в формате PNG.

Прежде чем график отображается, происходит расчёт всех значения в указанных диапазонах. Эти значения могут быть выведены в виде таблицы формата XLSX.

Важно отметить, что погрешность вычислений определяется при разработке программы, поэтому существует возможность определения необходимой точности расчётов в зависимости от поставленной задачи. Модульная структура языка *Python* позволяет использовать многие функции, написанные в ходе разработки данного проекта и при создании новых программ, что позволяет сэкономить время и упростить решение новых задач.

Безусловно, занятия в реальной, физической лаборатории очень важны, так как позволяют студентам получить практические навыки. Но проведение некоторых исследований требует слишком дорогостоящего оборудования, а проведение отдельных исследований может угрожать физической безопасности студентов.

Использование виртуальной лаборатории в процессе изучения электротехники имеет свои преимущества [5,6]:

- возможность дистанционного обучения;
- ускорение процесса обучения, так как можно быстро организовать эксперимент и быстро обнаружить возможные ошибки эксперимента и также быстро их устранить;
- возможность досконального изучения испытуемого устройства без риска его разрушить;
- возможность проведения экспериментов, которые невозможно выполнить с использованием имеющегося лабораторного оборудования.

В отличие от лабораторного стенда ЛКТЦ, который не позволяет проводить наглядные эксперименты по исследованию характеристик спектральной плотности сигналов конечной длительности, виртуальное лабораторное исследование предоставляет такую возможность для дискретных и аналоговых сигналов, как продемонстрировано на рисунках рисунки 6, 7 и 8.

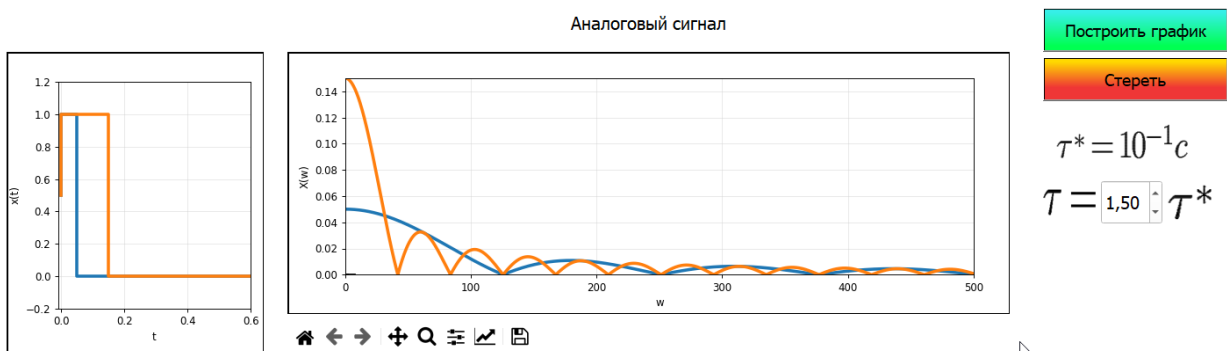


Рис. 6. Исследование спектральной характеристики аналогового сигнала конечной длительности

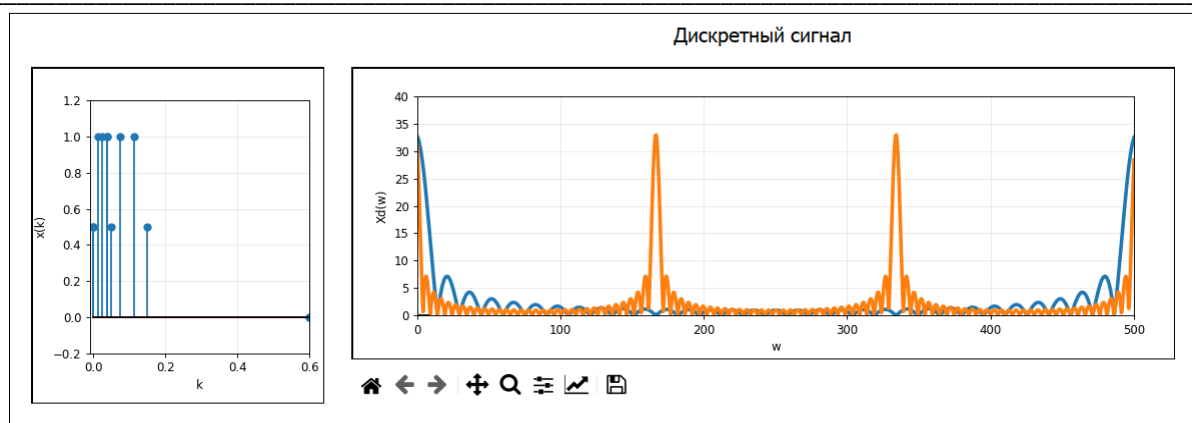


Рис. 7. Исследование спектральной характеристики дискретного сигнала конечной длительности

На рисунке 7 представлено окно программы, демонстрирующее результаты исследования спектральных характеристик дискретных сигналов прямоугольной формы конечной длительности. Данный эксперимент демонстрирует влияние выбора периода дискретизации на представление спектральной характеристики.

Упомянутый выше Лабораторный стенд ЛКТЦ не позволяет проводить наглядные эксперименты по исследованию спектральных характеристик амплитудно-модулированных сигналов. Виртуальное лабораторное исследование предоставляет такую возможность (рис. 8).

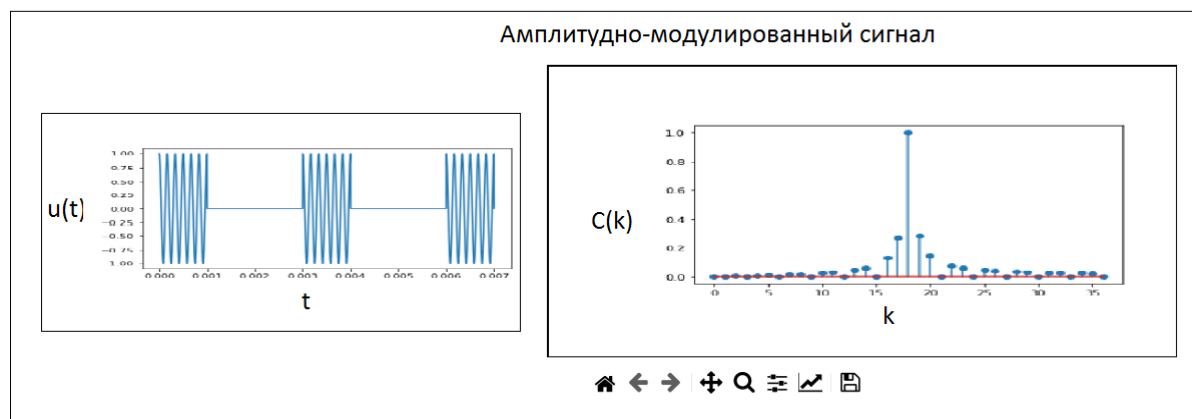


Рис. 8. Исследование спектральной характеристики амплитудно-модулированного сигнала

Заключение

Разработка и использование в учебном процессе виртуальных лабораторных работ является эффективным методом оптимизации учебного процесса, позволяющим наглядно представить решение тех или иных задач в рамках лабораторного практикума и, кроме того, обеспечивает возможность избежать применения проприетарного программного обеспечения. Разработанный виртуальный практикум по разделу «Спектральный анализ сигналов» планируется использовать в учебном процессе на кафедре ТЭЦ МТУСИ.

Литература

1. Добротворский И.Н. ТЭЦ. Лабораторный практикум. М: Радио и связь, 1990. 216 с.
2. Фриск В.В., Логвинов В.В. ТЭЦ, СТУ, РПрУСМС, РПрУСРС и РД Лабораторный практикум III на персональном компьютере. М.: СОЛОН-Пресс, 2016. 480 с.
3. Grigorieva E.D., Kreindelin V.B., Frisk V.V. Application of New Technologies and Software in the Development of a New Laboratory Stand for Studying the Discipline of Electrical Engineering // <https://ieeexplore.ieee.org/xpl/conhome/9744081/proceeding?isnumber=9744255&pageNumber=2> Published in: 2022 Systems of Signals Generating and Processing in the Field of on Board Communications
4. Тутов А.В., Григорьева Е.Д. Разработка лабораторного практикума на базе пакета прикладных программ Scilab по разделу «Спектральный анализ сигналов: периодических, конечной длительности и дискретных» // Труды НТК «Цифровые технологии радиосвязи и телерадиовещания». МТУСИ XII Молодёжный научный форум «Телекоммуникации и Информационные технологии», 12-21 апреля 2021. М., 2021.

5. *Sutchenkov A.A., Tikhonov A.I.* Electrical Engineering Materials Virtual Laboratory // В сборнике: 2018 4th International Conference on Information Technologies in Engineering Education, Inforino 2018 – Proceedings. 4. 2018. С. 8581843.

6. *Башарина С.О., Дубов В.М., Кубряков Е.А.* Переход на отечественное программное обеспечение в образовании // В сборнике: Информационные технологии в образовательном процессе вуза и школы. Материалы XIV Всероссийской научно-практической конференции. Редколлегия: В.В. Малев (науч. ред.), А.А. Малева (отв. ред.), М.В. Дюжакова, С.О. Башарина. 2020. С. 19-24.

7. *Крейнделин В.Б., Григорьева Е.Д.* Применение системы MATLAB в самостоятельной работе студентов при изучении дисциплины "теория электрических цепей" // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2018. Т. 7. № 2. С. 25-27.

8. *Григорьева Е.Д., Семёнова Т.Н., Степанова А.Г.* Организация самостоятельной работы студентов с применением электронного учебника при изучении дисциплины "теория электрических цепей" // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2018. Т. 7. № 2. С. 7-9.

9. *Григорьева Е.Д., Микиртичан А.Г., Степанова А.Г.* Особенности дистанционного обучения при преподавании электротехнических учебных дисциплин // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2021. Т. 10. № 1. С. 11-14.

**DEVELOPMENT OF A LABORATORY WORKSHOP
ACCORDING TO THE SECTION "SPECTRAL ANALYSIS OF SIGNALS:
PERIODIC, FINITE DURATION AND DISCRETE"**

Alexandr V. Titov,
Student MTUCI, MTUCI, Moscow, Russia,
[*cryogenmash97@gmail.com*](mailto:cryogenmash97@gmail.com)

Yelena D. Grigoreva,
Associate Professor of the Department of TEC, Ph.D., MTUCI, Moscow, Russia,
[*e.d.grigoreva@mtuci.ru*](mailto:e.d.grigoreva@mtuci.ru)

Abstract

This paper presents the results of the development of a highly specialized user application. As an object of research, the task of developing a set of virtual laboratory work to study the methods of spectral analysis of signals used in modern telecommunications equipment was chosen. The spectral characteristics of analog and discrete signals of finite duration, periodic analog and discrete signals, as well as amplitude-modulated signals are studied in laboratory work. Python software was chosen as the programming environment due to the fact that this software has the required flexibility. All calculations and graphs obtained in the created program were checked in the Mathcad programming environment.

Keywords: *virtual laboratory work, Python software, Scilab, free software, proprietary software, laboratory stands, spectral analysis.*

СИНТЕЗ НОВЫХ СИГНАЛОВ ДЛЯ ФИЛЬТРА БАТТЕРВОРТА ПО КРИТЕРИЮ МАКСИМУМА В ЦЕНТРЕ ЕГО ОТКЛИКА

*Мясищева Наталья Романовна,
студент МТУСИ, Москва, Россия,
mnatashka022@gmail.com*

*Санников Владимир Григорьевич,
доцент кафедры ОТС, к.т.н., МТУСИ, Москва, Россия,
v.g.sannikov@mtuci.ru*

Аннотация

В статье рассматривается проблема передачи данных по линейному, ограниченному по частоте, каналу связи без межсимвольной интерференции. Для решения данной проблемы проводится согласование формы передаваемых данных с временными и частотными характеристиками линейного канала связи. Для данного канала, с заданными импульсной реакцией и комплексным коэффициентом передачи, синтезируются новые сигналы по критерию максимума его отклика в центре символьного интервала. Получены новые общие выражения для оптимальных сигналов и новых множителей Лагранжа применительно к модели канала в виде фильтра Баттерворта. Приведены результаты моделирования сигналов для фильтра Баттерворта второго порядка.

Ключевые слова: линейный канал связи, фильтр Баттерворта, оптимальные финитные сигналы, синтез по максимуму в центре посылки, нивелирование межсимвольной интерференции.

В телекоммуникационных системах (ТКС) с повышенной скоростью передачи данных основной причиной снижения их помехоустойчивости является межсимвольная интерференция (МСИ) [1-3]. Для компенсации МСИ, вызываемой, в основном, нелинейностью АЧХ канала связи, в современных ТКС часто применяются эквалайзеры или выравниватели каналов. Последние, как известно [1-6], вносят значительную задержку в принимаемые в демодуляторе ТКС решения, и используют сложный приём в целом, основанный на том или ином варианте алгоритма Витерби [4-6]. Следует отметить, что помехоустойчивость такого приема оценивается приближенно.

В работе предлагается новый подход при организации ТКС. Основу его составляет синтез таких сигналов на передаче, согласованных с АЧХ канала связи, которые принципиально не вызывают МСИ на входе демодулятора ТКС. Синтез таких сигналов проводится путем решения новой вариационной задачи на условный экстремум. В решающем устройстве демодулятора ТКС при этом обеспечивается простой поэлементный прием на согласованный фильтр с импульсной реакцией, представляющей зеркальное отображение сигнала на выходе модулятора ТКС. Помехоустойчивость такого приема оценивается не приближенно, но точно. Этому направлению посвящены работы [7-9].

Вариационная задача синтеза новых сигналов, согласованных с АЧХ линейного канала связи

Дан линейный канал связи (КС) с известными импульсной реакцией $q(t)$, $t \geq 0$, и комплексным коэффициентом передачи $K(j\omega)$.

На вход канала воздействует финитный во времени сигнал $x(t)$, $0 \leq t \leq T$, с фиксированной энергией

$$E_x = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x^2(\omega) d\omega = \int_0^T x^2(t) dt = const, \quad (1)$$

где $S_x(j\omega)$ – комплексный спектр сигнала.

Требуется получить форму входного, ограниченного во времени сигнала $x(t)$, $0 \leq t \leq T$, такую, чтобы отклик канала $y(t)$ на этот сигнал остался также ограниченным, т.е. был задан на том же интервале $0 \leq t \leq T$, и в момент времени $t = T/2$ (середина посылки) было бы максимальным по всему множеству спектров сигнала $x(t)$, $0 \leq t \leq T$. Синтезированные по такому критерию сигналы обозначим как - ОФС1.1.

В отличие от аналогичной задачи, рассмотренной в [9], когда требовалось полное отсутствие МСИ на двух интервалах, а именно при $t \leq 0$ и $t \geq T$, в данной работе синтезируются ОФС1.1, обеспечивающие подавление МСИ на выходе линейного КС после окончания воздействия этого сигнала на входе КС, т.е. при $t \geq T$.

Выражения для отклика КС и его k -й производной при $k = 0, 1, \dots, (n-1) = \overline{0, (n-1)}$, равны

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega) K(j\omega) e^{j\omega t} d\omega, \quad \frac{d^k y(t)}{dt^k} = \frac{1}{2\pi} \int_{-\infty}^{\infty} (j\omega)^k S_x(j\omega) K(j\omega) e^{j\omega t} d\omega. \quad (2)$$

Критерий оптимальности представляется так

$$\max_{x(t)} y(T/2) = \max_{S_x(j\omega)} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega) K(j\omega) e^{j\omega \frac{T}{2}} d\omega \right\} \quad (3)$$

при дополнительных условиях

$$\frac{d^k y(t)|_{t=T}}{dt^k} = \frac{1}{2\pi} \int_{-\infty}^{\infty} (j\omega)^k S_x(j\omega) K(j\omega) e^{j\omega T} d\omega = 0, \quad k = \overline{0, (n-1)}. \quad (4)$$

Для решения поставленной вариационной задачи введем новый функционал

$$J = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(j\omega) K(j\omega) e^{j\omega \frac{T}{2}} d\omega - \sum_{k=0}^{n-1} \lambda_k \frac{1}{2\pi} \int_{-\infty}^{\infty} (j\omega)^k S_x(j\omega) K(j\omega) e^{j\omega T} d\omega. \quad (5)$$

где λ_k – новые, по сравнению с [9], множители Лагранжа, $k = 0, 1, \dots, (n-1) = \overline{0, (n-1)}$.

Максимальное значение функционала (5) по всем спектрам $S_x(j\omega)$ ищется из неравенства Гёльдера [8]

$$S_x(j\omega) = \varepsilon K(-j\omega) \left\{ e^{-j\omega \frac{T}{2}} - \sum_{k=0}^{n-1} \lambda_k (-j\omega)^k e^{-j\omega T} \right\} = \varepsilon K(-j\omega) \Phi(j\omega, T), \quad (6)$$

где $\varepsilon = E_x / y(T/2)$ – постоянная, определяемая из ограничения 1, и сигнальная функция вида

$$\Phi(-j\omega, T) = e^{-j\omega T/2} - \sum_{k=0}^{n-1} \lambda_k (-j\omega)^k e^{-j\omega T}. \quad (7)$$

Тогда спектр синтезированного сигнала $y(t)$, $0 \leq t \leq T$, на выходе КС находится по соотношению

$$S_y(j\omega) = S_x(j\omega) K(j\omega) = \varepsilon K^2(j\omega) \Phi(-j\omega, T), \quad (8)$$

В соответствие со спектрами (6) и (8) находятся искомые сигналы

$$x(t) = \varepsilon \left\{ q\left(\frac{T}{2} - t\right) - \sum_{k=0}^{n-1} \lambda_k \frac{d^k q(T-t)}{dt^k} \right\}, \quad -\infty \leq t \leq T, \quad (9)$$

$$y(t) = \varepsilon \left\{ G\left(\frac{T}{2} - t\right) - \sum_{k=0}^{n-1} \lambda_k \frac{d^k G(T-t)}{dt^k} \right\}, \quad -\infty \leq t \leq \infty, \quad (10)$$

где $q(t)$, $t \geq 0$ – импульсная реакция КС и $G(t)$ – канальная функция КС вида

$$G(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} K^2(j\omega) e^{j\omega t} d\omega = \int_0^{\infty} q(\tau) q(\tau + |t|) d\tau, \quad -\infty \leq t \leq \infty. \quad (11)$$

Сигналы в (9) и (10) обеспечивают максимум функционалу (5) при известных характеристиках линейного КС и найденных значений множителей Лагранжа.

В общем виде решение этой задачи затруднительно. Поэтому остановимся на модели линейного канала связи с комплексно сопряженными полюсами, к которой, в частном случае, относится фильтр Баттерворта четного порядка [9].

Новые финитные сигналы ОФС 1.1, согласованные с характеристиками фильтра Баттерворта

Полагаем, что в модуляторе ТКС используется линейный фильтр Баттерворта с АЧХ вида:

$$K(\omega, p) = 1 / \sqrt{1 + \nu[\omega / \omega_\gamma]^{2p}}, \quad (12)$$

где p - порядок фильтра, $F_\gamma = \omega_\gamma / 2\pi$ - граничная частота по уровню ослабления его АЧХ на L_γ дБ, ν - величина, зависящая от L_γ ; так при $L_\gamma = -10 \lg 2 \approx -3$ дБ, $\nu = 1$.

В работе [9] для фильтра Баттерворта с АЧХ (12) получены импульсная реакция, её производные, а также канальная функция и её производные $Q^{(k)}(t), k = \overline{1, p}$

$$q^{(k)}(t) = \frac{\omega_\gamma^{k+p/2}}{p} \sum_{l=1}^{p/2} q_l e^{-s_{F,l} t} \sin[c_{F,l} t + (k+1-p/2)(\pi/2 + \varphi_l)], \quad t \geq 0, \quad (13)$$

$$Q^{(k)}(t) = \frac{\omega_\gamma^{k+1}}{2p} \sum_{l=0}^7 e^{-s_{F,l} t} \sin[c_{F,l} t + \varphi_l + k(\pi/2 + \varphi_l)], \quad \omega_\gamma = 2\pi F_\gamma,$$

где $s_l = \omega_\gamma \sin \varphi_l$, $c_l = \omega_\gamma \cos \varphi_l$, $q_l = 1 / \left[c_l \prod_{i=1}^4 (s_i - s_l)_{i \neq l} \right]$, $\varphi_l = (2l-1)\pi/2p$.

Для фильтра Баттерворта с импульсной реакцией (13) сигналы в (9) и (10) будут финитными, определены на интервале $0 \leq t \leq T$, если параметры $\lambda_m, m = 0, 1, \dots, (p-1)$, удовлетворяют новой системе уравнений:

$$\begin{cases} 1 - \sum_{m=0}^{p-1} \lambda_m \omega_\gamma^m \left[e^{-\alpha_k \frac{T}{2}} \cos[\beta_k T / 2 - m(\pi/2 - \varphi_k)] \right] = 0, \\ \sum_{m=0}^{p-1} \lambda_m \omega_\gamma^m \left[e^{-\alpha_k \frac{T}{2}} \sin[\beta_k T / 2 - m(\pi/2 - \varphi_k)] \right] = 0, \quad m = \overline{1, p/2}, \end{cases} \quad (14)$$

Пример расчета форм ОФС1.1 и их спектров

Численные значения для множителей Лагранжа при решении поставленной выше вариационной задачи можно получить, если в качестве модели линейного канала связи выбрать фильтр Баттерворта второго порядка. Решая систему линейных уравнений второго порядка на основе (14), после достаточно сложных вычислений приходим к следующим новым формулам для множителей Лагранжа

$$\lambda_{1,0} = [\text{ctg}(b_T) - 1] \cdot \exp(b_T), \quad \lambda_{1,1} = \frac{\sqrt{2}}{\omega_\gamma} \exp(b_T) \cdot \sin(b_T), \quad b_T = \omega_\gamma T / 2\sqrt{2}. \quad (15)$$

На основе полученных соотношений (6) – (11), (13) и (15), проведен машинный эксперимент в среде MatLab (рис. 1). В расчетах принято $F_\gamma = 1,55$ кГц.

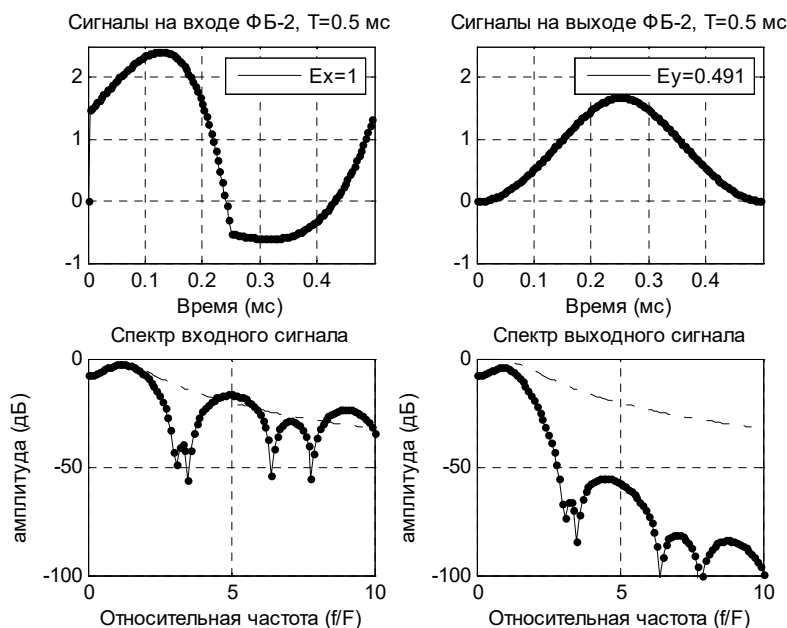


Рис. 1. ОФС1 и их спектры на входе и выходе ФБ-2 (метод 1)

Заключение

Таким образом, поставленная задача синтеза финитного сигнала $x(t)$, сохраняющего финитным отклик фильтра Баттерворта $y(t)$ и максимизирующего мгновенное значение этого отклика при $t = T/2$, полностью решена. Сигналы $x(t)$ и $y(t)$ будут финитными и определены на интервале времени $0 \leq t \leq T$. Откуда следует, что при последовательной передаче ОФС1.1 по линейному ограниченному по полосе частот каналу связи, типа фильтра Баттерворта, отсутствует явление межсимвольной интерференции, а помехоустойчивость ТКС рассчитывается точно, так как определяется только наличием аддитивного белого шума наблюдения.

Литература

1. Макаров С.Б., Цикин И.А. Передача дискретных сообщений по радиоканалам с ограниченной полосой частот. М.: Радио и связь, 1988. 304 с.
2. Николаев Б.И. Последовательная передача дискретных сообщений по непрерывным каналам с памятью. М.: Радио и связь, 1988. 204 с.
3. Зяблов В.В., Коробков Д.Л., Портной С.Л. Высокоскоростная передача сообщений в реальных каналах. М.: Радио и связь, 1991. 288 с.
4. Скляр Бернгард. Цифровая связь. Теоретические основы и практическое применение, 2-е изд.: Пер с англ. М.: Издательский дом «Вильямс», 2003. 1104 с.
5. Феер К. Беспроводная цифровая связь. Методы модуляции и расширения спектра: Пер. с англ. / Под ред. В.И. Журавлева. М.: Радио и связь, 2000. 520 с.
6. Прокис Дж. Цифровая связь: Пер. с англ./ Под ред. Д.Д. Кловского. М.: Радио и связь, 2000. 800 с.
7. Ключев В.И. Частотно-временные преобразования и прием дискретных сигналов в системах связи. М.: Радио и связь, 1990. 208 с.
8. Ямпольский Э.М. Вариационные принципы согласования сигналов с каналом связи. М.: Радио и связь, 1987. 136 с.
9. Санников В.Г. Синтез финитных сигналов Найквиста, согласованных с телефонным каналом связи // Электросвязь, № 5, 2012. С. 9-12.

SYNTHESIS OF NEW SIGNALS FOR THE BUTTERWORTH FILTER ACCORDING TO THE MAXIMUM CRITERION IN THE CENTER OF ITS RESPONSE

Natalia R. Myasisheva,
Student MTUCI, Moscow, Russia,
mnatashka022@gmail.com

Vladimir G. Sannikov,
Associate Professor of the Department of GCT, Ph.D., MTUCI, Moscow, Russia,
v.g.sannikov@mtuci.ru

Abstract

The problem of data transmission through a linear, frequency-limited, communication channel without intersymbol interference is considered. To solve this problem, the form of the transmitted data is coordinated with the time and frequency characteristics of the linear communication channel. For this channel, with a given pulse reaction and a complex transmission coefficient, new signals are synthesized according to the criterion of the maximum of its response in the center of the symbol interval. New general expressions for optimal signals and new Lagrange multipliers applied to the channel model in the form of a Butterworth filter are obtained. The results of signal modeling for the second-order Butterworth filter are given.

Keywords: *linear communication channel, Butterworth filter, optimal finite signals, synthesis to the maximum in the center of the parcel, leveling of intersymbolic interference.*

ОПТИМИЗАЦИЯ ПРОПУСКНОЙ СПОСОБНОСТИ СПУТНИКОВОГО КАНАЛА СВЯЗИ ЗА СЧЕТ ПРИМЕНЕНИЯ АДАПТИВНЫХ АЛГОРИТМОВ ПОМЕХОУСТОЙЧИВОГО КОДИРОВАНИЯ

*Лобов Сергей Сергеевич,
магистрант МТУСИ, Москва, Россия,
sl14032014@yandex.ru*

*Шаврин Сергей Сергеевич,
профессор кафедры МТС, д.т.н., МТУСИ, Москва, Россия,
sss@mtuci.ru*

Аннотация

Рассматривается проблема передачи информации по каналу спутниковой системы «Марафон IoT» с постоянной скоростью кода. Приводится сравнение и анализ передачи информации с постоянной и переменной скоростью кода. Основная цель – вычисление выигрыша от адаптивного помехоустойчивого кодирования, при определенных временных интервалах кодовой перестройки.

***Ключевые слова:** сверточные коды, алгоритм Витерби, частота битовых ошибок, BER, образующий полином, регистр сдвига, скорость кода, корректирующая способность, минимальное кодовое расстояние.*

Введение

В настоящее время в Российской Федерации существует проект разработки спутниковой системы «Марафон IoT», предназначенной для передачи коротких сообщений и обмена данными между устройствами «Интернет Вещей». Система строится на базе низкоорбитальной спутниковой группировки, расположенной на орбите высотой 750 км и углом места равном 10° [1].

Одним из факторов, пагубно влияющих на помехозащищенность передаваемого сигнала, является доплеровский сдвиг частоты между передаваемым и принимаемым сигналом, значение которого достигает $3 \cdot 10^{-5}$. Вдобавок, помехозащищенность сигнала уменьшается по мере отдаления космического аппарата (КА) от наземной станции в зоне радиообзора, и в точке наихудшего приема меньше на 9.6 дБ, чем в точке наилучшего приема. Поскольку во время сеанса связи наибольшее отношение сигнал/шум на входе приемника составляет порядка 14 дБ, обеспечить надежную передачу информации возможно при применении сверточного кода со скоростью $1/2$ с последующей корректировкой кодом Рида-Соломона. С учетом того, что максимальное изменение защищенности сигнала в процессе передачи данных составляет ≈ 10 дБ, очевидно, что применение кодирования сверточным кодом со скоростью $1/2$ на протяжении всего сеанса связи является избыточным.

Данная проблема может быть решена применением алгоритмов адаптивного сверточного кодирования, оптимизирующих пропускную способность спутникового канала связи. Использование адаптивных алгоритмов помехоустойчивого кодирования открывает новые возможности в области повышения скорости передачи информации и позволяет по-новому взглянуть на разработку спутниковых модемов.

Использование сверточных алгоритмов помехоустойчивого кодирования в спутниковых модемах российского производства

Сверточные алгоритмы помехоустойчивого кодирования были разработаны в 1955 году советским ученым Львом Матвеевичем Финком, и в настоящее время используются в ряде спутниковых модемах российского производства, например, в цифровых спутниковых модемах серии СЦМ-5: СЦМ-5, СЦМ-5А (СЦМ-5АК), СЦМ-5В (СЦМ-5ВК). В вышеперечисленных модемах и в большинстве спутниковых модемах других производителей широкое распространение получил сверточный код с образующими полиномами 171 в восьмеричной системе счисления, и 133 в восьмеричной системе счисления при входной длине кодового ограничения равной 7. Сверточный кодер с данными полиномами представлен на рисунке 1.

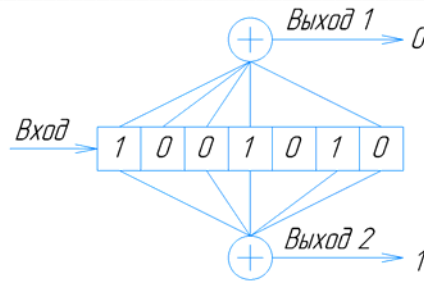


Рис. 1. Сверточный кодер

В качестве алгоритма декодирования сверточных кодов большую популярность получил разработанный в 1967 году алгоритм Витерби. При жестких решениях демодулятора данный алгоритм представляет собой поиск пути на решетчатой диаграмме кода с минимальной метрикой, а при мягких решениях – пути с наименьшим расстоянием Евклида. Согласно [2], использование мягких решений позволяет увеличить энергетический выигрыш от кодирования на 2 дБ. Широкое распространение сверточных кодов вместе с декодированием по алгоритму Витерби в спутниковых системах связи обосновано тем, что спутниковые каналы связи близки по своим свойствам к каналам с аддитивным белым гауссовским шумом, являющимися симметричными каналами без памяти [3,4].

С учетом вышесказанного, сигнально-кодую конструкцию спутникового модема «Марафон IoT» предлагается формировать двухпозиционной фазовой манипуляцией BPSK в сочетании со сверточным кодированием со скоростями 1/2, 3/4, 7/8 и декодированием по алгоритму Витерби с мягкими решениями. Данный вид модуляции характеризуется максимальной устойчивостью к воздействию Доплеровской деструкции спектра, относительное значение которой (ожидаемое) составляет величину $3 \cdot 10^{-5}$.

Модем предназначен для передачи информации со скоростью не менее 1 Мбит/с. В процессе пролета над зоной радиобзора при постоянных символьной скорости передачи и вероятности ошибки битовая скорость будет меняться за счет изменения кодовой скорости при адаптивной настройке средств помехоустойчивого кодирования. Снижение энергетического выигрыша от кодирования по мере приближения КА к наземной станции обеспечит возможность повышения скорости передачи до 1.75 Мбит/с. Резерв энергетического выигрыша от кодирования составляет 9.6 дБ.

Моделирование канала связи

Имитация канала связи проводилась при помощи модели, спроектированной в среде Simulink (рис. 2).

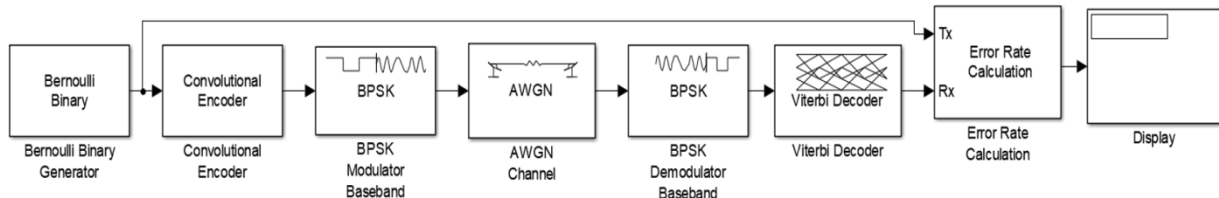


Рис. 2. Модель Simulink

Битовые псевдослучайные последовательности размером 1 Мбит, 1.5 Мбит, 1.75 Мбит подвергались сверточному кодированию $C(171_8, 133_8)$ со скоростями 1/2, 3/4 и 7/8 соответственно.

Таким образом, размер последовательности на выходе канального кодера составлял 2 Мбит. Вектор перфорирования для скорости 3/4 – 110110, для скорости 7/8 – 11010101100110. Закодированная битовая последовательность поступала на вход модулятора BPSK.

Модулированный сигнал передавался по каналу АБГШ. Значение отношения сигнал/шум на бит информации (E_b/N_0) в канале связи выбиралось из диапазона [4 дБ, 14 дБ] с шагом 0.5 дБ. Декодирование осуществлялось по алгоритму Витерби с мягким решением при ширине окна декодирования равной пяти длинам входного кодового ограничения – 5k (35).

С помощью блока «Error Rate Calculation» осуществлялось вычисление вероятности битовой ошибки в цифровом потоке на выходе сверточного декодера. По результатам моделирования построен график зависимость BER в битовом потоке на выходе декодера Витерби с мягкими решениями от E_b/N_0 канала связи (рис. 3).

Согласно результатам моделирования, при кодировании со скоростью 7/8 в диапазоне E_b/N_0 канала от 14 до 9.5 дБ ошибок в декодированном битовом потоке не возникает. При E_b/N_0 канала равной 9 дБ частота битовых ошибок на выходе декодера Витерби равна примерно $1 \cdot 10^{-6}$.

При кодировании со скоростью $3/4$ в диапазоне E_b/N_0 канала от 14 до 7 дБ ошибок в декодированном битовом потоке не возникает. При E_b/N_0 канала равной 6.5 дБ частота битовых ошибок на выходе сверточного декодера равна примерно $1 \cdot 10^{-6}$. При кодировании со скоростью $1/2$ в диапазоне E_b/N_0 канала от 14 до 5 дБ ошибок в декодированном битовом потоке не возникает. При E_b/N_0 канала равной 4.5 дБ частота битовых ошибок на выходе сверточного декодера равна примерно $6 \cdot 10^{-6}$, а при 4 дБ – $2.4 \cdot 10^{-5}$. Таким образом, в точке наилучшего приема BER на выходе декодера Витерби с мягкими решениями не превышает $3 \cdot 10^{-5}$.

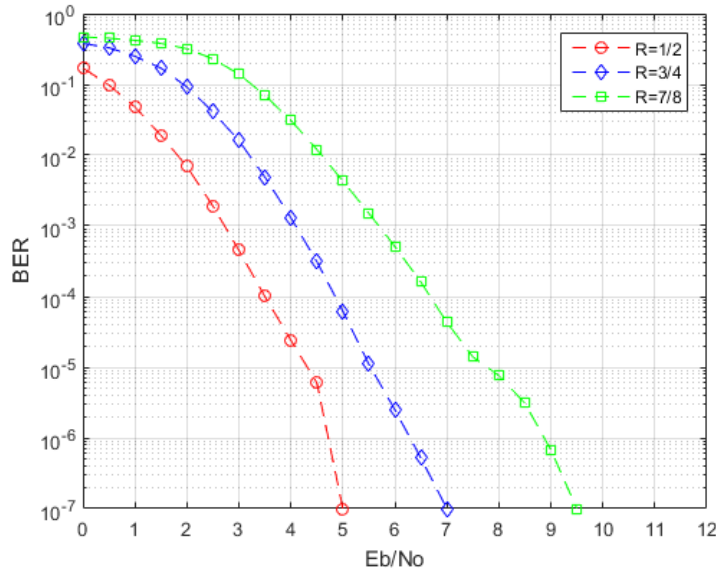


Рис. 3. Зависимость BER на выходе декодера Витерби с мягкими решениями от E_b/N_0 канала

Адаптивная перестройка параметров сверточного кода

Используя результаты моделирования, возможны различные варианты адаптивной перестройки сверточного кода. В данной статье предлагается следующая схема: при передаче данных с кодовой скоростью равной $7/8$, $3/4$ и $1/2$ частота битовых ошибок на выходе сверточного декодера должна быть не более $3 \cdot 10^{-5}$. При минимальном расстоянии между С и КА отношение сигнал/шум на входе приемника составляет не менее 14 дБ.

Согласно данным моделирования, данного значения вполне достаточно для передачи информации с кодовой скоростью равной $7/8$. По мере отдаления спутника от наземной станции отношение сигнал/шум на входе приемника уменьшается и при значении равном 8 дБ система изменяет скорость кодирования на $3/4$, изменение защищенности сигнала в этом случае будет составлять 6 дБ. При дальнейшем отдалении КА и уменьшении отношения сигнал шум до 6 дБ изменение защищенности будет составлять 8 дБ, а скорость кодирования уменьшится до $1/2$ (рис. 4).

В качестве синхропоследовательности, предназначенной для информирования приемного устройства о начале передачи информации с определенной кодовой скоростью рекомендуется применять линейную рекуррентную последовательность максимального периода (далее М-последовательность), длиной 31 бит вида: 1001011001111100011011101010000. Генерация М-последовательности такой длины производится с помощью регистра сдвига с линейной обратной связью и образующего полиномом 5 степени вида: $g(x)=x^5+x^2+1$. Преимущество использования синхропоследовательности такого типа заключается в том, что она обладает хорошей автокорреляционной функцией.

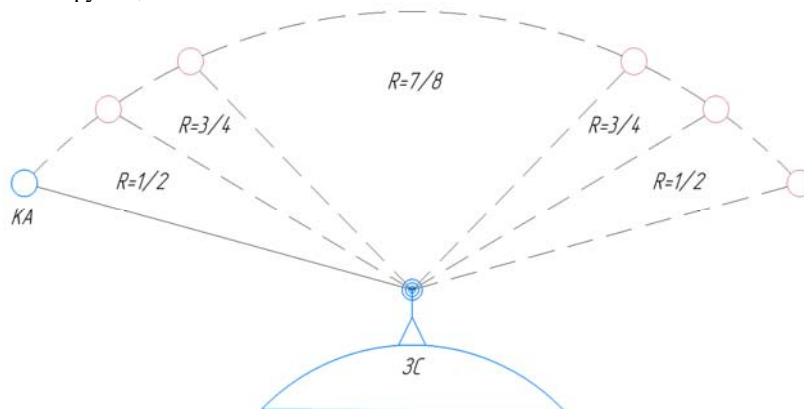


Рис. 4. Изменение скорости сверточного кода во время сеанса связи

При циклическом сдвиге количество совпадений у последовательности и её сдвинутой копии равно 15. Таким образом, для обнаружения синхрокомбинации достаточно совпадения $31-16 \div 2 = 23$ бит. Иными словами, синхронизация обеспечивается в случаях, если в синхропоследовательности искажено менее 8 бит. Полагая, что распределение ошибок в канале связи подчиняется биномиальному закону, при вероятности ошибки в канале связи $BER = 1 \cdot 10^{-2}$ вероятность того, что в 31-битной последовательности будет искажено 8 и более элементов определится по формуле Бернулли и равна $8.3 \cdot 10^{-10}$. Следовательно, вероятность пропуска посылки в точке наихудшего приема составляет $8.3 \cdot 10^{-10}$.

Соответственно, в упрощенном виде процесс кодовой перестройки можно описать следующим образом: при прохождении КА точек орбиты, в которой требуется перестройка сверточного кода, передающая сторона приостанавливает процесс сверточного кодирования (при этом не обнуляя регистр сдвига), генерирует соответствующую синхропоследовательность и отправляет в канал связи. Получив из канала последовательность определенного вида, приемная сторона изменяет параметры канального декодера и отправляет приемнику квитанцию о перестройке. Получив соответствующую квитанцию, приемник осуществляет перестройку канального кодера и возобновляет процесс передачи информации.

Для того, чтобы определить выигрыш при применении адаптивного алгоритма сверточного кодирования необходимо рассчитать время прохождения КА по дуге на определенной кодовой скорости. Для этого необходимо рассчитать расстояние от наземной станции до КА и длину дуги. Расстояние от наземной станции до КА «Марафон IoT» определяется по следующей формуле:

$$d = H \cdot 10^{\frac{\Delta A}{20}}, \tag{1}$$

где H – высота орбиты, ΔA – изменение защищенности сигнала.

Длина дуги определяется по следующей формуле:

$$L = 2 \cdot (R + H) \cdot \arcsin\left(\frac{d \cdot \cos \alpha}{R + H}\right) \tag{2}$$

где R – радиус Земли, H – высота орбиты, d – расстояние от наземной станции до КА «Марафон IoT», α – угол места.

Время прохождения КА по дуге на определенной кодовой скорости определяется по формуле:

$$t = \frac{L}{v}, \tag{3}$$

где $v = 7.8$ км/с – скорость движения КА «Марафон IoT» по орбите.

Количество информации, переданное при определенной скорости кодирования, определяется по формуле:

$$B = B_0 \cdot t, \tag{4}$$

где B_0 – скорость битового потока на входе сверточного кодера.

Данные результатов расчетов сведены в таблицу и представлены в таблице 1.

Таблица 1

Результаты расчетов

R	ΔA , дБ	d_{max} , км	L, км	t, с	B, Мбит
7/8	[0, 6)	1496.4	2968.8	380	666.1
3/4	[6, 8)	1883.9	785.1	101	151
1/2	[8, 9.6]	2264.9	783.7	100	100.5

Исходя из данных таблицы 1, можно сделать вывод, что подавляющую часть времени сеанса связи ($\approx 65\%$) осуществляется передача информации с кодовой скоростью равной 7/8 и приблизительно по 17% времени сеанса затрачивается на передачу данных со скоростью 3/4 и 1/2. При постоянном (неадаптивном) кодировании со скоростью равной 1/2 количество информации, переданное за время сеанса связи, будет составлять 581 Мбит, а при адаптивной скорости – 917.6 Мбит.

Таким образом, при данной схеме кодовой перестройки, выигрыш при применении адаптивного помехоустойчивого кодирования может составлять 336.6 Мбит, что подтверждает актуальность её использования.

Заключение

Предложена модель сигнально-кодовой конструкции спутникового модема «Марафон IoT», и приведены результаты исследования сверточных кодов. С учетом проделанных исследований, в течении сеанса связи рекомендуется осуществлять адаптивную перестройку сверточного кода не более 4 раз. Количество информации,

переданной за время сеанса связи при адаптивном помехоустойчивом кодировании, превосходит примерно на 58% количество информации, переданной при неадаптивном помехоустойчивом кодировании со скоростью $\frac{1}{2}$.

Литература

1. Анпилогов В.Р. Проект «Марафон» – система спутниковой связи для IoT: Доклад на Ежегодной конференции операторов и пользователей сети спутниковой связи и вещания в России, Москва. 03.10.2019.
2. Никитин Г. И. Сверточные коды: Учебное пособие. С-П.: Сов. радио, 2001. 78 с.
3. Морелос-Сарагоса, Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение. М.: Техносфера, 2005. 320 с.
4. Орлов В.Г., Шаврин С.С. Беспроводные мобильные приложения в системах мониторинга и диспетчеризации технологических служб // Фундаментальные проблемы радиоэлектронного приборостроения. 2008. Т. 8. № 1. С. 247-251.

THE OPTIMIZATION OF A SATELLITE COMMUNICATION CHANNEL CAPACITY THROUGH THE APPLICATION OF ADAPTIVE ALGORITHMS OF NOISE-RESISTANT CODING

Sergey S. Lobov

Graduate MTUCI, Moscow, Russia,
sl14032014@yandex.ru

Sergey S. Shavrin

Professor of the Department of MTS, Doctor of Technical Sciences,
MTUCI, Moscow, Russia,
sss@mtuci.ru

Abstract

The problem of information transmission over the channel of the Marathon IoT satellite system with a constant code rate is considered. Comparison and analysis of information transmission with constant and variable code rates is given. The main goal is to calculate the gain from adaptive error-correcting coding, at certain time intervals of code reconstruction.

Keywords: convolutional codes, Viterbi algorithm, bit error rate, polynomial BER, shift register, code rate, corrective power, minimum code distance.