

НАУЧНЫЙ ЖУРНАЛ

**ТЕЛЕКОММУНИКАЦИИ
И ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ**

№2-2022

(Дата издания: декабрь 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

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

Сидорин Виктор Викторович

д.т.н., профессор, Российский технологический университет МИРЭА, Москва, 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 №2-2022

«Информационные технологии, кибернетика и информационная безопасность»

<i>Андреева Я. А., Шармаев В. И., Василевский К. А.</i> АТАКИ «НУЛЕВОГО ДНЯ» И МЕТОДЫ ИХ ОБНАРУЖЕНИЯ	6
<i>Чубаров А. О., Докучаев В. А.</i> АНАЛИЗ ИСПОЛЬЗОВАНИЯ «ПУЗЫРЬКОВОЙ СОРТИРОВКИ» ДЛЯ ОЦЕНКИ ОПТИМИЗАЦИИ ПРОГРАММНОГО КОДА НА C#	14
<i>Мясников Д. Ю., Смирнов А. Э.</i> РАЗРАБОТКА АЛГОРИТМА ЗАЩИТЫ СЕРВЕРА БАЗЫ ДАННЫХ ОТ СЕТЕВЫХ АТАК	19
<i>Зайченко И. А., Большаков А. С.</i> ОБ ИСПОЛЬЗОВАНИИ СИСТЕМНЫХ ВЫЗОВОВ WIN-API ДЛЯ ОБНАРУЖЕНИЯ МОДИФИЦИРОВАННОГО ВРЕДНОСНОГО ПО	28
<i>Леухин А. И., Панков К. Н.</i> ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ БЛОКЧЕЙН ДЛЯ ЭКСПЕРТНЫХ СИСТЕМ	37
<i>Браславский Д В., Скородумова Е. А.</i> РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ РАЗБОРА МАТЧЕЙ КОМАНДНЫХ ВИДОВ СПОРТА С ПРИМЕНЕНИЕМ ТЕХНОЛОГИИ VR/AR	42
<i>Чекина Д. А., Пронина Е. Д.</i> ОБЗОР СУЩЕСТВУЮЩИХ СИСТЕМ ОТСЛЕЖИВАНИЯ СПУТНИКОВ	52
<i>Панов А. И.</i> ПРИМЕНЕНИЕ ЦИФРОВЫХ ДВОЙНИКОВ В ПРОМЫШЛЕННОМ ПРОИЗВОДСТВЕ	59

«Цифровые технологии радиосвязи и телерадиовещания»

<i>Власюк И. В., Киселева А. С.</i> АНАЛИЗ ЭФФЕКТИВНОСТИ БЕЗРЕФЕРЕНСНЫХ МЕТРИК ПРИМЕНИТЕЛЬНО К ОЦЕНКЕ КАЧЕСТВА ВИДЕО ПРИ ПОТОКОВОЙ ПЕРЕДАЧЕ	65
<i>Попов О.Б., Василевский К. А., Гаранин Т. Д., Орлов В. Г., Цыплаков Д. И.</i> ВЛИЯНИЕ ВОССТАНОВИТЕЛЬНОЙ КОРРЕКЦИИ НА ПАРАМЕТРЫ РЕЧЕВОГО СИГНАЛА	75
<i>Давыдова А. А., Егоров Д. А.</i> ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ МОДЕЛЕЙ УДАЛЕНИЯ ФОНА	84
<i>Максимов А. А., Комаров И. В., Горгадзе С. Ф.</i> МОДЕЛИРОВАНИЕ КОМПЕНСАЦИИ НЕЛИНЕЙНЫХ ИСКАЖЕНИЙ В УСИЛИТЕЛЕ МОЩНОСТИ НА ОСНОВЕ ЕГО ЦИФРОВОЙ ПОВЕДЕНЧЕСКОЙ МОДЕЛИ	92
<i>Пронина Е. Д., Свирская А. В.</i> ОБЗОР ПОСТРОЕНИЯ СОЛНЕЧНЫХ ПАНЕЛЕЙ ДЛЯ СПУТНИКОВ ТИПА CUBESAT	100

<i>Иванчев В.В., Власюк И.В.</i> МЕТОД РАСПРЕДЕЛЕНИЯ ВЫБОРКИ ЦВЕТОВЫХ ОТТЕНКОВ В ЦВЕТОВОМ ПРОСТРАНСТВЕ CIELCH	108
---	-----

<i>Ву Ши Дао., Горгадзе С. Ф.</i> ВЕРОЯТНОСТНЫЕ ХРАКТЕРИСТИКИ УСКОРЕННОГО ПОИСКА ШУМОПОДОБНЫХ СИГНАЛОВ	120
--	-----

<i>Калинина Л. В., Чернышева Т. В., Орлов В.Г.</i> ИССЛЕДОВАНИЕ ДОПОЛНИТЕЛЬНЫХ КРИТЕРИЕВ КАЧЕСТВА ЗВУКА В ПОМЕЩЕНИЯХ С КУПОЛЬНОЙ КОНСТРУКЦИЕЙ ПОТОЛКА	128
---	-----

«Сетевые технологии и системы телекоммуникаций»

<i>Рабенандрасана Жослен, Зайцев А. И.</i> ПРИМЕНЕНИЕ НАУКОМЕТРИЧЕСКОГО АНАЛИЗА ДЛЯ ВЫЯВЛЕНИЯ ПЕРСПЕКТИВНЫХ РАЗРАБОТОК В ОБЛАСТИ ВОЛС	135
---	-----

<i>Петров С. П., Бен Режеб Тауфик Бен Камель</i> МНОЖЕСТВЕННЫЙ ДОСТУП С РАЗДЕЛЕНИЕМ АБОНЕНТОВ ПО УРОВНЮ МОЩНОСТИ PD-NOMA	144
--	-----

<i>Деарт В. Ю., Лысых П. О.</i> ТЕНДЕНЦИИ РАЗВИТИЯ ТЕХНОЛОГИИ T-SDN	150
--	-----

<i>Носов И. Ю., Мирошниченко А. В.</i> СРАВНЕНИЕ МЕТОДОВ ОПРЕДЕЛЕНИЯ УГЛОВЫХ КООРДИНАТ В СИСТЕМАХ РАДИОМОНИТОРИНГА	158
--	-----

«Схемотехника и технологические решения»

<i>Фан Туан Ань., Логвинов В. В.</i> УПРАВЛЯЕМЫЙ СВЧ АВТОГЕНЕРАТОР НА КОМПЛЕМЕНТАРНОЙ ПАРЕ МОП – ТРАНЗИСТОРОВ	166
---	-----

<i>Борисов М. Д., Волков Д. В., Яновский А. С.</i> РАЗРАБОТКА ЛАБОРАТОРНОГО СТЕНДА ДЛЯ НАГЛЯДНОГО ОТОБРАЖЕНИЯ И ИЗУЧЕНИЯ СТОЯЧИХ ВОЛН	174
---	-----

<i>Иванчев В. В.</i> ЦВЕТОВОЕ КОНТРАСТИРОВАНИЕ В РЕНТГЕНОВСКОМ КОНТРОЛЕ	181
--	-----

<i>Чернышева Т. В., Лукашова Е. А., Попов О. Б., Орлов В. Г.</i> МОДЕЛИРОВАНИЕ И РАСЧЁТ АКУСТИЧЕСКИХ ПАРАМЕТРОВ ТЕАТРАЛЬНОГО ЗАЛА СО СЛОЖНОЙ АРХИТЕКТУРНОЙ ФОРМОЙ	193
---	-----

«Информационные технологии, кибернетика и информационная безопасность»

АТАКИ «НУЛЕВОГО ДНЯ» И МЕТОДЫ ИХ ОБНАРУЖЕНИЯ

Андреева Яна Андреевна,
магистрант МТУСИ, Москва, Россия,
andreevaya.00@mail.ru

Шармаев Вадим Игоревич,
аспирант МАИ, Москва, Россия
vadidq@ya.ru

Василевский Кирилл Антонович,
магистрант МТУСИ, Москва, Россия,
alaxtver@yandex.ru

Аннотация

В статье рассматриваются угрозы информационной безопасности, связанные с эксплуатацией уязвимостей «нулевого дня». Актуальность данной проблемы обусловлена стремительным развитием информационных технологий, влекущим за собой рост уязвимостей в компьютерных системах и, соответственно, рост количества атак. Злоумышленники очень быстро адаптируются к современным системам защиты информации и находят новые способы для проведения атак. В работе определены наиболее вероятные сценарии возникновения новых атак в компьютерной сети, проанализированы известные уязвимости «нулевого дня» последних лет, а также существующие методы и средства их обнаружения.

Ключевые слова: информационная безопасность, защита, атаки «нулевого дня», обнаружение атак, сетевые аномалии, машинное обучение, глубокое обучение, нейронные сети.

Введение

Программные продукты всех производителей имеют уязвимости, которые могут представлять угрозу информационной безопасности. Вредоносное программное обеспечение используется в качестве основного инструмента для эксплуатации этих уязвимостей. В наши дни скорость генерации вредоносных программ очень высока. Это связано с тем, что злоумышленники владеют различными механизмами, позволяющими генерировать новые варианты вредоносных программ, то есть создавать мутации. Например, полиморфные и метаморфические механизмы модифицируют некоторые части исходного кода вредоносного ПО, вследствие чего антивирусное программное обеспечение не в состоянии его обнаруживать.

Исторически люди использовали системы обнаружения вторжений (СОВ) для защиты своих сетей от злонамеренных воздействий. Традиционные системы обнаружения вторжений включают в себя мониторинг сети и обнаружение сетевых атак с помощью сигнатур. Такие СОВ успешно справляются с задачей обнаружения известных атак, но терпят неудачу в случае возникновения новых. Кроме того, появление новых шаблонов требует обновления базы данных сигнатур, содержащей определение атак, что увеличивает потенциальное окно злонамеренного воздействия. Трудно предсказать тип, место и сценарий атак «нулевого дня». Неизвестная природа таких атак делает их непредсказуемыми и сложными как для обнаружения, так и для защиты от них.

Альтернативным решением является применение подходов машинного обучения. Популярность машинного обучения стремительно растет последнее десятилетие, и теперь методы машинного обучения применяются во многих областях жизнедеятельности человека, таких как здравоохранение, рекомендательные системы и фильтрация спама по электронной почте и многих других.

Определение понятия атак «нулевого дня»

Программное обеспечение часто содержит уязвимости, которые могут использоваться злоумышленники для своих целей. При создании нового продукта и поддержания уже существующего, разработчики программного обеспечения проводят тестирование безопасности на наличие разного рода уязвимостей, которые необходимо устранить.

Уязвимость «нулевого дня» – общий термин, обозначающий обнаруженную недавно уязвимость в программном обеспечении, которую злоумышленники могут использовать для создания эксплойта. Атака «нулевого дня» происходит тогда, когда злоумышленник использует неизвестную разработчикам уязвимость.

Уязвимость возникает в момент внесения ошибки, которая не была выявлена разработчиками на этапе тестирования. После того, как уязвимость или эксплойт обнаружены и описаны, поставщик уязвимого программного обеспечения выпускает патч для закрытия уязвимости, а разработчики средств безопасности обновляют антивирусные сигнатуры для обнаружения эксплойта или конкретных атак.

Жизненный цикл атаки «нулевого дня» включает шесть основных этапов (рис. 1).



Рис. 1. Жизненный цикл атаки «нулевого дня»

Жизненный цикл атаки «нулевого дня» начинается в момент добавления ошибки в процессе разработки нового программного обеспечения, подготовки обновления или в момент обнаружения «дыр» безопасности в применяемых технологиях. На следующем этапе злоумышленник обнаруживает уязвимость, создает рабочий эксплойт и использует его для проведения скрытых атак против выбранных целей. Производитель узнаёт об уязвимости в ходе тестирования или из сторонних отчетов, оценивает ее серьезность, назначает приоритет исправления и начинает работу над созданием патча. Далее уязвимость раскрывается публично. Уязвимость добавляется в базу данных общеизвестных уязвимостей информационной безопасности, ей присваивается идентификационный номер вида CVE-год-номер, описание и ряд общедоступных ссылок.

После раскрытия уязвимости разработчики антивирусных программ выпускают сигнатуры для новой атаки и создают эвристические методы обнаружения эксплойта. Вследствие этого атаки могут быть обнаружены антивирусными программами с обновленными сигнатурами. Завершающий этап жизненного цикла атаки «нулевого дня» – выпуск исправления разработчиками уязвимого программного обеспечения. После этого, хосты, которые применили патч, будут в дальнейшем не подвержены эксплойту.

Проведенные исследования атак «нулевого дня» показывают, что уязвимости могут существовать в течение значительного периода времени, прежде чем будут обнаружены: в среднем десять месяцев, а после публичного раскрытия новых уязвимостей продолжительность их воздействия увеличивается примерно в пять раз [1].

Кроме того, в исследовании [2] авторы приводят график по годам среднего количества дней, необходимых разработчикам для выпуска патча для новой уязвимости. Несмотря на то, что за последние 10 лет процесс создания исправления сократился, выпуск патча все еще занимает значительное время – более трех месяцев.

Таким образом, можно сделать вывод, что атаки «нулевого дня», вопреки своему названию, остаются актуальными достаточно длительный период времени. А после публичного раскрытия, когда об уязвимости становится известно широкому кругу лиц, такие атаки оказываются еще более опасными.

Обзор сценариев атак «нулевого дня»

Атаки «нулевого дня» могут быть направлены на разные объекты компьютерной инфраструктуры: операционные системы, веб-браузеры, офисные приложения, аппаратное обеспечение, прошивки, интернет вещей и так далее. Следовательно, круг потенциальных жертв таких атак достаточно широкий:

- Обычные пользователи, использующие уязвимые системы, такие как веб-браузеры, операционные системы для собственных нужд;
- Устройства интернет вещей, сетевое оборудование и другие устройства;
- Лица, имеющие доступ к ценным данным;
- Предприятия и организации;
- Государственные органы.

Таким образом, атаки «нулевого дня» способны нанести ущерб большому количеству пользователей разного уровня. Атаки «нулевого дня» принято делить на целевые и нецелевые.

Целевые атаки чаще всего направлены против потенциально значимых объектов, таких как крупные организации, правительственные учреждения и т.д.

К нецелевым относятся атаки, эксплуатирующие уязвимости программного обеспечения общего назначения и направлены на пользователей этого уязвимого программного обеспечения.

Злоумышленники обычно используют следующие приёмы для распространения уязвимостей нулевого дня:

1. Фишинговые электронные письма с прикрепленными в них файлами с эксплойтами.
2. Спам-письма, социально ориентированные на то, чтобы побудить получателей перейти по URL-ссылке на мошеннические веб-сайты, на которых размещены эксплойты.
3. Наборы эксплойтов, чьи цепочки атак включают в себя вредоносную рекламу и вредоносные сайты, на которых размещены уязвимости нулевого дня.

Наиболее известные уязвимости «нулевого дня» последних лет

Согласно отчету компании Microsoft [3], в последние годы наблюдается резкое увеличение количества неизвестных кибератак. В течение последних нескольких лет было выявлено ряд серьезных уязвимостей в программном обеспечении общего назначения, среди них уязвимость «нулевого дня» Google Chrome, Zoom, Apple iOS, Microsoft Word.

В 2021 году была выявлена уязвимость веб-браузера Google Chrome. Причиной уязвимости стали ошибки, возникшие в JavaScript-движке. Прежде чем компания выпустила патч, Chrome подвергся серии атак, связанных с эксплуатацией этой уязвимости.

В 2020 у платформы Zoom была обнаружена уязвимость. В результате эксплуатации этой уязвимости злоумышленники могли получить удаленный доступ к компьютерам пользователей, при этом, если жертвой атаки становились администраторы, злоумышленники могли получить доступ ко всем файлам.

Еще одной компанией, столкнувшейся с атакой «нулевого дня» в 2020 году, стала Microsoft [4]. Эта атака была нацелена на уязвимость удаленного выполнения кода в банкомате. Эксплуатации уязвимости позволила злоумышленникам удаленно запускать вредоносные сценарии, которые рассылаются через спам.

Другая атака «нулевого дня» эксплуатировала уязвимость CVE-2020-12271 и была направлена на изменение настроек брандмауэра Sophos XG. Посредством эксплуатации данной уязвимости злоумышленникам удавалось получить несанкционированный доступ к системе.

Apple iOS, которую также считают самой безопасной операционной системой для мобильных устройств, в 2020 году была подвержена как минимум двум атакам «нулевого дня». Злоумышленниками было обнаружено несколько уязвимостей, одна из которых позволяла удаленно скомпрометировать устройство.

В 2019 году была раскрыта уязвимость Microsoft Windows, позволяющая повысить локальные привилегии. В данном случае злоумышленники воспользовались уязвимостью локальных привилегий Microsoft Windows для запуска произвольного кода и установки программ, а также для просмотра и изменения данных о скомпрометированных программах.

В 2017 году результатом эксплойта нулевого дня Microsoft Word стала компрометация личных банковских счетов пользователей. Жертвами этой атаки стали люди, открывшие на своем устройстве вредоносный документ Word. В самом документе отображалось предложение загрузить удаленное содержимое – всплывающее окно, запрашивающее внешний доступ из другой программы. После нажатия на кнопку «Да» на устройства пользователей устанавливалось вредоносное ПО, которое впоследствии перехватывало учетные данные для входа в интернет-банк.

Рассмотренные примеры, позволяют убедиться в том, что от такого вида злонамеренного воздействия, как атаки «нулевого дня», не могут защититься даже крупные разработчики программных продуктов.

В исследовании [5] авторы отмечают, что количество атак «нулевого дня» в 2019 году превышает предыдущие три года, а авторы статьи [6] ссылаются на статистическое исследование, которое показывает, что 62% атак удается обнаружить лишь тогда, когда система уже скомпрометирована. Все эти соображения подчеркивают очевидную и насущную потребность в более эффективных моделях обнаружения атак «нулевого дня» и методов их предотвращения.

Методы обнаружения атак «нулевого дня»

Методы обнаружения атак «нулевого дня» принято разделять две группы по объекту рассмотрения: методы обнаружения аномалий и методы обнаружения злоупотреблений (рис. 2).

Методы обнаружения злоупотреблений основаны на описании известных атак: если поведение наблюдаемого объекта совпадает с описанием известной атаки, поведение объекта считается атакой.

Основной сложностью при создании системы обнаружения злоупотреблений является вопрос об эффективном механизме формирования правил. Очевидно, что создание полноценной базы правил для выявления всевозможных атак является невозможным в силу нескольких факторов.

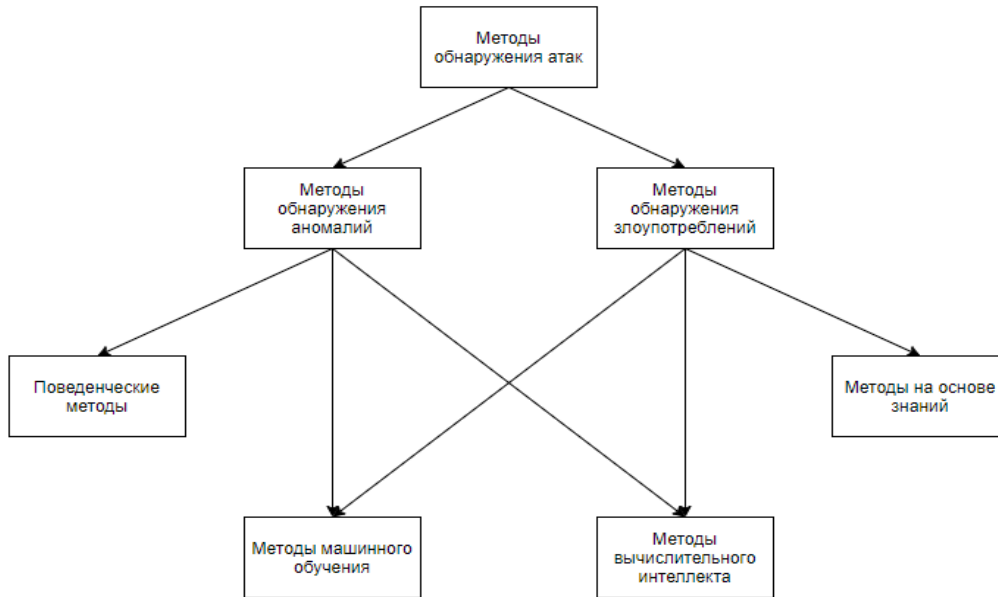


Рис. 2. Классификация методов обнаружения атак

Один из этих факторов заключается в том, что описание различных вариаций атакующих действий негативно сказывается на производительности системы. А поскольку даже несущественные изменения в атаке приводят к невозможности ее обнаружения методами на основе злоупотреблений, то задаваемые правила должны быть универсальными и покрывать как можно большее число известных модификаций сетевых атак [7]. Особенности методов обнаружения злоупотреблений делают их применимыми в задачах обнаружения уже известных атак, но в выявлении новых атак их использование не эффективно.

Системы обнаружения вторжений на основе аномалий используют эвристические методы для создания модели сети, определяющей допустимые шаблоны и поведение пользователей в качестве основы. Затем модель используется для выявления отклонений от базовой модели, чтобы найти потенциальные вторжения.

Системы обнаружения аномалий исходит из предположения, что поведение зловредного программного обеспечения отличается от типичного поведения легитимных программ. Основное преимущество этого метода заключается в том, что с помощью этой модели можно обнаружить даже неизвестные атаки. Однако, поскольку методы, основанные на аномалиях, классифицируют любые отклонения от нормы как зловредные, легитимное поведение системы может быть отождествлено как вторжение, что приводит к высокому уровню ложных срабатываний.

Сетевые аномалии имеют различные причины и могут быть связаны с деятельностью злоумышленников, некомпетентных пользователей, неисправностью аппаратуры и дефектами программного обеспечения [8].

Сетевые аномалии с точки зрения объекта воздействия можно разделить на две группы: программно-аппаратные отклонения и нарушения безопасности (рис. 3).

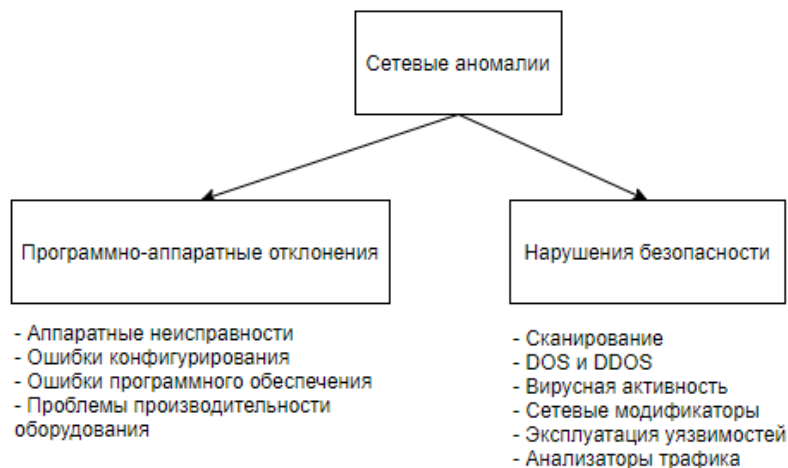


Рис. 3. Классификация сетевых аномалий

К программно-аппаратным отклонениям относятся: аппаратные неисправности, ошибки конфигурирования, ошибки программного обеспечения и проблемы производительности оборудования.

Программно-аппаратные отклонения могут повлечь за собой перевод системы в нештатный режим функционирования, что нарушает общую работоспособность системы или вызывает полный выход из строя отдельных компонентов информационной системы.

К нарушениям безопасности относятся такие сетевые атаки, как сканирование сети, анализ трафика, атаки на отказ обслуживания, вирусы, сетевые модификаторы и эксплуатация уязвимостей информационной системы.

Среди методов обнаружения аномалий выделяют:

- поведенческие методы;
- методы машинного обучения;
- методы вычислительного интеллекта.

Поведенческие методы

Поведенческие методы – это методы, основанные на использовании информации о нормальном поведении системы и ее сравнении с параметрами наблюдаемого поведения. Данная группа методов ориентирована на построение модели штатного, или нормального, функционирования системы или пользователя. В процессе своей работы системы, использующие данный подход, сравнивают текущие показатели активности с профилем нормальной деятельности, и в случае значительных отклонений может рассматриваться как свидетельство наличия атаки [9].

К поведенческим методам обнаружения атак относят вейвлет-анализ, статистический анализ и т. д. Для этих методов характерно большое количество ложноположительных срабатываний, то есть признаний легитимных действий за атаку, что значительно снижает эффективность системы обнаружения вторжений. Такие ошибки возникают в основном из-за сложности полного описания легитимных действий пользователя. Большинству подобных систем необходим достаточно длительный период предварительной настройки, во время которого у системы формируется модель нормального поведения.

Методы машинного обучения и вычислительного интеллекта

В настоящее время методы машинного обучения и методы вычислительного интеллекта довольно часто применяются в задачах обнаружения аномалий и злоупотреблений. Это связано с тем, что указанные подходы в качестве данных для обучения используют как нормальное, так и аномальное поведение, что значительно увеличивает точность обнаружения.

Системы обнаружения вторжений на основе аномалий обнаруживают сетевые аномалии, анализируя сетевые шаблоны и обнаруживая действия, которые не соответствуют нормальным шаблонам. Системы обнаружения аномалий должны тщательно воспринимать характеристики нормальных данных, чтобы иметь возможность различать аномальные закономерности.

Машинное обучение – это область искусственного интеллекта, целью которого является разработка алгоритмов и математических моделей, которые могут обучаться на образцах данных, без явных инструкций. Модели, разработанные с использованием машинного обучения, извлекают шаблоны и используют выводы для изучения характеристик данных, для прогнозов относительно новых данных.

В зависимости от метода, используемого при обучении, методы машинного обучения можно разделить на две категории: обучение с учителем (контролируемое обучение) и обучение без учителя или неконтролируемое обучение.

В контролируемом методе обучения математическая модель обучается с использованием полностью размеченного набора данных. Методы с учителем можно разделить на две категории: регрессия и классификация. Алгоритмы регрессии используются, когда выходные данные находятся в непрерывном диапазоне, а алгоритмы классификации используются, когда выходные данные представляют собой дискретные категории.

Деревья решений, k ближайших соседей, метод опорных векторов и искусственные нейронные сети — одни из самых популярных алгоритмов машинного обучения с учителем. Все эти алгоритмы использовались в области обнаружения вторжений с разной степенью успеха [10].

Неконтролируемые методы обучения используют неразмеченные наборы данных. Как правило, неконтролируемые алгоритмы группируют входные данные по категориям на основе общности точек данных. Неконтролируемые алгоритмы используются в случае, когда помеченный набор данных недоступен. Алгоритмы кластеризации, такие как кластеризация K-средних и иерархическая кластеризация, являются примерами алгоритмов машинного обучения без учителя.

Иногда гораздо проще собрать данные, принадлежащие одному классу, чем разным. Например, в случае обнаружения аномалий сбор не аномальных данных проще, чем сбор данных об аномалиях, поскольку аномалии встречаются реже. Это также относится и к области обнаружения вторжений, так как проще собирать сетевые данные без атак, чем собирать данные о сетевых вторжениях. В этом случае можно использовать

полууправляемые методы, чтобы начать с частично помеченных данных, изучить характеристики нормальных данных и найти отклонения от нормы. Метод опорных векторов и автокодировщики являются примерами алгоритмов, которые могут работать частично контролируемым образом.

Хотя алгоритмы машинного обучения доказали свою эффективность в обучении на основе данных без явных инструкций, производительность алгоритмов машинного обучения во многом зависит от представления данных.

Каждая часть информации в представлении называется параметром, и для того, чтобы алгоритмы машинного обучения работали хорошо, параметры должны быть тщательно подобраны специалистом в изучаемой предметной области.

Глубокое обучение

Глубокое обучение – это область машинного обучения, которая может решить проблему определения наиболее оптимальных параметров. Модели глубокого обучения имеют многоуровневую архитектуру, что позволяет им самостоятельно выбирать необходимые параметры. Первые уровни моделей изучают более простые представления из данных, которые затем объединяются для создания сложных представлений [11].

Большинство используемых в настоящее время современных моделей глубокого обучения основаны на искусственных нейронных сетях.

Искусственные нейронные сети – это универсальные аппроксиматоры функций. Цель обучения нейронной сети состоит в том, чтобы найти функцию, которая может описать отношение между входными и выходными данными выборки, которую можно обобщить на новые данные, не содержащиеся в обучающей выборке.

Искусственные нейронные сети состоят из вычислительных блоков, называемых искусственными нейронами, которые созданы по аналогии с биологическими нейронами. Искусственный нейрон принимает ряд входных данных для формирования выходного сигнала.

Искусственные нейроны в нейронной сети организованы слоями. Сеть принимает входные данные и обрабатывает их слой за слоем для получения выходных данных. Первый слой сети называется входным. Последний слой является выходным слоем, и выход этого слоя также является выходом нейронной сети. Остальные слои называются скрытыми слоями. Каждый нейрон принимает выходные данные нейронов предыдущего слоя и создает выходные данные для следующего слоя. Информационный поток начинается с входного слоя, проходит через скрытые слои и достигает выходного слоя.

Цель обучения нейронной сети – найти оптимальные значения весов для каждого нейрона, чтобы выходные данные, выдаваемые моделью, были ближе всего к фактическим выходным данным, заданным во входных данных выборки.

Чтобы измерить способность нейронной сети предсказывать правильный результат для данного входа, используется функция потерь. Функция потерь измеряет разницу между реальным выходным значением нейронной сети и выходным значением, выдаваемым сетью. Одной из наиболее популярных функций потерь является среднеквадратическая ошибка, которая вычисляет среднеквадратическое отклонение значений реальных и расчетных выходных значений.

Другой распространенной функцией потерь является бинарная Кросс-энтропия. Кросс-энтропия используется, когда выходные данные могут принимать значения 0 и 1.

Целью обучения нейронной сети является минимизация выбранной функции потерь путем корректировки значений весов и смещений нейронов. Обучение выполняется с использованием контролируемого алгоритма обучения, называемого обратным распространением, который использует выборочные данные в качестве входных и выходных пар.

Алгоритм обучения начинается со случайной инициализации весов сети. После этого две фазы – проход вперед и проход назад – повторяются до тех пор, пока потери не будут минимизированы. На этапе прямого прохода вычисляются выход сети и значение функции потерь. Затем, на этапе обратного прохода, с помощью алгоритма градиентного спуска определяется, как обновить веса, чтобы уменьшить потери. После этого, веса обновляются на небольшую величину.

Алгоритм градиентного спуска используется для определения того, как обновить веса, чтобы уменьшить потери. Градиент определяется как соответствующее изменение значения потерь при обновлении параметра веса. Алгоритм обучения постепенно уменьшает потери небольшими шагами до тех пор, пока они не перестанут уменьшаться.

Нейронные сети имеют два типа параметров: обучаемые параметры и необучаемые параметры. Обучаемые параметры сети – это параметры, значения которых изучаются в процессе обучения. Эти параметры включают веса и смещения нейронов. Необучаемые параметры сети называются гиперпараметрами. Эти параметры определяются до начала обучения и включают в себя такие параметры, как количество скрытых слоев и нейронов в нейронной сети, скорость обучения и алгоритм оптимизации. Значения этих параметров должны быть заданы эмпирическим путем.

Нейронные сети используются в области обнаружения вторжений с 2008 года. Классификаторы на основе глубоких нейронных сетей применяются во многих исследованиях в области обнаружения вторжений, а их

результаты доказывают эффективность использования систем машинного обучения для задач обнаружения атак «нулевого дня» [12-15].

Заключение

Анализ уязвимостей «нулевого дня» последних лет показал, что злоумышленникам удается находить и эксплуатировать уязвимости в программном обеспечении, установленном на большом количестве устройств. При этом уязвимыми оказываются программные продукты от крупных разработчиков, таких как Microsoft, Google, Zoom, Apple. В таком случае любое устройство, на котором установлено уязвимое ПО, может быть атаковано злоумышленниками.

Произведенный анализ жизненного цикла атак «нулевого дня» показал, что окно воздействия неизвестных атак начинается в момент создания эксплойта для обнаруженной злоумышленником новой уязвимости и завершается с выпуском патча разработчиками программного обеспечения. При этом, для обнаружения новой атаки в среднем требуется десять месяцев [1], а для устранения - еще около трех [2], то есть злоумышленники могут более года пользоваться эксплойтами, прежде чем уязвимость будет устранена. Это лишний раз подчеркивает важность борьбы с такого рода злонамеренными воздействиями.

На основании проведенного обзора методов и средств обнаружения неизвестных атак можно сделать вывод, что наиболее перспективными методами для решения задач детектирования неизвестных атак являются методы машинного и глубокого обучения.

Литература

1. *Bilge L., Dumitras T.* Before we knew it: An empirical study of zero-day attacks in the real world // ACM Conference on Computer and Communications Security – 2012, pp. 833-844.
2. *Roumani Y.* Patching zero-day vulnerabilities: an empirical analysis // Journal of Cybersecurity – 2021. Vol.7, pp. 1-13.
3. *Kumari A.* “Exploiting a crisis: How cybercriminals behaved during the outbreak.”, 2021, (Интернет ресурс, дата обращения 03.10.2022) - [<https://www.microsoft.com/security/blog/2020/06/16/exploiting-a-crisis-how-cybercriminals-behaved-during-the-outbreak/>].
4. *John E.* Windows has a zero-day that won't be patched for weeks, 2020 (Интернет ресурс, дата обращения 11.04.2022) - <https://nakedsecurity.sophos.com/2020/03/25/windows-has-a-zero-day-that-wont-be-patched-for-weeks>.
5. *Metrick K., Najafi P.* Zero-Day Exploitation Increasingly Demonstrates Access to Money, Rather than Skill // J Intelligence for Vulnerability Management – 2020.
6. *Nguyen T., Reddi V.* Deep Reinforcement Learning for Cyber Security, 2019.
7. *Borkar A., Donode A.* A survey on Intrusion Detection System (IDS) and Internal Intrusion Detection and protection system (IIDPS) // International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 949-953.
8. *Кожевникова. И. С.* Анализ методов обнаружения аномалий для обнаружения сканирования портов // Молодой ученый. 2017. № 14 (Интернет ресурс, дата обращения 26.04.2022) <https://moluch.ru/archive/148/41829/>.
9. *Li R., Zhai S., Zhi Y.* An intrusion detection method based on behavior characteristics for business logic // Journal of Physics: Conference Series – 2020.
10. *Özgür A., Erdem J.* Genetic algorithm in intrusion detection systems using feature selection and multiclassifier fusion // Pattern Recognition. Vol. 40. 2018, pp. 2373-2391.
11. *Lecun Y., Bengio G.* Deep learning. Hinton Nature, 2015, p. 10.
12. *Miller M.* Are we protected yet? developing a machine learning detection system to combat zero-day malware attacks // IEEE International Conference on Big Data (Big Data), 2018.
13. *Shun J., Malki A.* Network Intrusion Detection System Using Neural Networks // Fourth Int. Conf. Nat. Comput. 2009. Vol. 5, pp. 242-246.
14. *Roy S., Mallik A., Gulati R.* A deep learning based artificial neural network approach for intrusion detection // Communications in Computer and Information Science, 2017.
15. *Tang T., Mhamdi L., McLernon D.* Deep learning approach for Network Intrusion Detection in Software Defined Networking // International Conference on Wireless Networks and Mobile Communications, WINCOM, 2016.

ZERO-DAY ATTACKS AND METHODS OF THEIR DETECTION

Yana A. Andreeva,
Graduate MTUCI, Moscow, Russia
andreevaya.00@mail.ru

Vadim I. Sharmaev,
Postgraduate MAI, Moscow, Russia
vadidq@ya.ru

Kirill A. Vasilevskii,
Graduate MTUCI, Moscow, Russia
alaxtver@yandex.ru

Abstract

This article discusses the threats to information security associated with the exploitation of zero-day vulnerabilities. The urgency of the problem is due to the development of information technology, which entails an increase in vulnerabilities in systems and, accordingly, an increase in the number of attacks. Attackers adapt very quickly to modern information security systems and find new ways to carry out their attacks. The paper identified the most likely scenarios for the occurrence of new attacks in a computer network, analyzed the known zero-day vulnerabilities of recent years, as well as existing methods and means of detecting them.

Keywords: *information security, protection, zero-day attacks, attack detection, network anomalies, machine learning, deep learning, neural networks.*

АНАЛИЗ ИСПОЛЬЗОВАНИЯ «ПУЗЫРЬКОВОЙ СОРТИРОВКИ» ДЛЯ ОЦЕНКИ ОПТИМИЗАЦИИ ПРОГРАММНОГО КОДА НА C#

Чубаров Андрей Олегович,
магистрант МТУСИ, Москва, Россия,
andrew.chubarov@internet.ru

Докучаев Владимир Анатольевич,
заведующий кафедрой СИТuС МТУСИ, д.т.н.,
профессор, МТУСИ, Москва, Россия,
v.a.dokuchaev@mtuci.ru

Аннотация

Рассматриваются широко используемые в настоящее время классические методы повышения эффективности и быстродействия конечного кода ПО. Проведён анализ двух типов оптимизации для кода, написанного на C#. Приводятся результаты применения метода замены части кода на более низкоуровневый с целью повышения производительности ПО.

Ключевые слова: оптимизация кода, C#, C-подобные языки программирования, анализ быстродействия, рефакторинг кода.

С появлением недорогих жёстких дисков, память которых удовлетворяет запросам, как обычных пользователей, так и владельцев современных серверных и коммерческих решений, разработчики программного обеспечения (ПО) стали уделять всё меньше времени на оптимизацию кода ради экономии памяти на жёстких дисках. Связано это с тем, что любое крупное приложение, которое пишется компаниями со слабой инфраструктурой, часто получает недостаточную поддержку средствами или кадрами высококвалифицированных программистов. Такие компании, разрабатывая свои продукты с целью раннего выпуска конечного продукта на рынок, редко заботятся о качестве кода и создаваемое ПО часто оказываются набором из отдельных фрагментов перегруженного кода. Также многие считают, что дешевле купить лишний жёсткий диск для хранения избыточной информации, чем нанимать опытных специалистов или задержать релиз. С созданием твердотельных накопителей вновь встал вопрос об экономии места на диске.

С появлением кроссплатформенных интерпретаторов и облачных сервисов вопрос экономии памяти почти исчерпал себя. Для крупных приложений были разработаны специальные фреймворки, оптимизацией которых занимаются отдельные компании. Память стала гораздо дешевле в производстве, если сравнивать с ситуацией, которая была пять-десять лет назад. Но при этом перегруженные приложения так и продолжают представлять в релиз. Особенно часто этим занимаются «молодые» компании, желающие быстро выйти на коммерческий ИТ-рынок.

В настоящее время возникает новая проблема – оптимизация приложений для повышения их быстродействия при большой их загруженности [1,2]. В общем случае под оптимизацией программного кода понимают различные методы его преобразования с целью улучшения его характеристик и повышения эффективности работы. Среди целей оптимизации можно указать уменьшения объема кода, объема используемой программой оперативной памяти, ускорение работы программы и уменьшение количества операций ввода-вывода [3].

Оптимизация кода ПО должна проводиться в соответствие с системными принципами [4], а также со следующими основными принципами программирования: естественность (код должен быть аккуратным, модульным и легко «читабельным»), естественно встраиваться в программу, легко поддаваться редактированию, интегрированию или удалению отдельных функций), производительность (в результате оптимизации должен получиться прирост производительности ПО – оптимизированная программа увеличивает быстродействие минимум на 20-30% в сравнение с исходным вариантом кода), затраченное время (оптимизация и последующая отладка ПО должны занимать небольшой период времени – оптимальными считаются сроки, не превышающие 10-15% времени, затраченного на написание самого программного продукта) [5].

Для решения этих задач существуют разнообразные методы [6,7], большинство из которых можно разделить на «классические» (с применением современных инструментов быстродействия кода) и «современные» (когда самые громоздкие части кода переписываются на более низкоуровневых языках программирования). В связи с этим актуальным является анализ подходов к выбору лучшего метода оптимизации. При этом необходимо протестировать быстродействие каждого отдельного метода и определить целесообразность его использования в разработке ПО, т.е. доказать эффективность оптимизации.

С той целью используем классическую пузырьковую сортировку 100000 элементов, отсортированных изначально в обратном порядке (количество элементов обусловлено быстродействием современных компьютеров).

Для обеспечения максимальной точности исследований использовались средства Windows и были пропущены первые два запуска после компиляции исходного кода. На такой шаг необходимо было пойти, так как при измерении времени первых запусков не учитываются такие свойства оперативной памяти, как хранение информации о ранее запущенных приложениях.

Для проверки эффективности каждого метода сначала были произведены замеры среднего значения времени работы пузырьковой сортировки без проведённых оптимизаций (время, за которое происходила полная сортировка). Затем каждый метод рассматривался по отдельности, чтобы в конце исследования можно было наглядно увидеть, какой из них можно считать самым быстродействующим. Использование пузырьковой сортировки обусловлено тем, что она может достоверно показать степень оптимизации кода в зависимости от проведённых манипуляций в случае, если приходится сортировать большое количество чисел, полностью отсортированных в обратном порядке.

Ресурс «Professor Web» [8] провёл полное исследование вопроса оптимизации кода на C#. Представленный в нём материал рассчитан на профессионалов в данной области. В работе предоставлено исчерпывающее руководство по использованию эффективных инструментов тестирования. Также там приводится пример, показывающий почему тестирование оптимизации также должно быть проведено корректно: порой инструменты для тестирования могут нагружать систему даже сильнее, чем тестируемый код.

Для исследования был выбран язык программирования C# т.к. он является быстроразвивающимся кроссплатформенным, постоянно обновляется и получает с каждым крупным патчем обновлённую стандартизацию. Также быстро в своё время развивался и сейчас продолжает развиваться ещё один крупный язык программирования – C++. Это обусловлено тем, что его развитием занимается компания Microsoft, которая может позволить себе одновременно заниматься и его постоянной оптимизацией: следует обратить внимание на значимый пример с точкой входа, которую в других языках уже давно сделали независимой от главной функции Main.

Требования к производительности должны выражаться в показателях, которые можно измерить. Кроме того, требования к производительности должны содержать некоторую информацию о программном окружении: общую или конкретную для конкретной задачи [8].

В роли окружения в данном исследовании выступали операционная система Windows, язык программирования C#, среда разработки (IDE) Visual Studio, которые представляют собой основной набор инструментов для разработки на языке C#. Инструментом тестирования была та же среда, в которой и происходит написание кода, т.е. процесс будет реализовываться с максимальным упором на производительность и быстродействие, которое может дать программное окружение. Главной целью тестирования является проверка целесообразности применения тех или иных методов оптимизации, а не её важность в целом. При этом сравнивалась производительность кода каждого этапа, не считая промежуточных, так как они влияют на производительность на уровне погрешности.

Во многих случаях классическая оптимизация представляет собой рефакторинг уже имеющегося кода (переработка исходного кода программы, чтобы он стал более простым и понятным). То есть изначально пишется непосредственно сам код, и только потом он редактируется в соответствии с принятыми правилами. Это может быть как небольшое внесение изменений, так и переписывание с нуля большей части функций. Примером такого подхода является ядро Linux, где с каждым патчем разработчики переписывают модули ядра с нуля, даже если проще было бы внести небольшие коррективы (пример: замена сервера отображения Xorg на Wayland).

Определим, какие изменения требуется внести в проект с пузырьковой сортировкой, чтобы наглядно продемонстрировать процесс оптимизации кода. Сборка проекта осуществлялась и тестировалась с классической конфигурацией Visual Studio Debug. В соответствии с [9] проект начинался с использования в своём коде классового объектно-ориентированного программирования (ООП). Это не только повышает удобство чтения кода, но и позволяет вынести побочные методы за пределы основного кода. Когда в последних версиях языка программирования C# пришлось отказаться от привычной точки входа программы, то это стало уже обязательным условием, а не процессом оптимизации. Но так как требуется рассмотреть все возможные сценарии, то начинать следует с вынесения побочных методов путём их отделения от главного, где и происходит запуск консольного приложения.

Следующий этап заключается в явном указании типа переменных. ООП – это основа оптимизации, но тип данных – это её фундамент. Сейчас всё большую популярность обретают языки программирования с динамической типизацией. Часто такой подход вызывает всё больше вопросов, и программисты начинают реже задумываться о возможных неправильных вхождениях. Это приводит к частому написанию исключений, и код получает всё больше ненужных нагромождений. В идеале необходимо каждый раз указывать тип данных, когда инициализируется новая переменная.

Исследование начинается с замеров с помощью таймера времени (запускается со старта выполнения кода), которое уходит на сортировку уже отсортированного в обратном порядке кода. Чем больше таких замеров, тем точнее будут конечные результаты. Для тестирования кода было выбрано число измерений в количестве шести на каждый код. Первые измерения времени были произведены для оригинального неоптимизированного кода, а

последующие проводились для уже оптимизированного под быстроедействие кода. Результаты измерений представлены в таблице 1.

Таблица 1

Измерения скорости полной сортировки массива

Время на исполнение оригинального проекта, (с)	55.02	55.25	55.63	55.31	55.48	66.01
Время на исполнение проекта с классической оптимизацией, (с)	44.82	55.03	44.94	55.12	44.80	55.34

Как можно заметить, при повторном измерении уже оптимизированного кода наблюдается несущественный прирост в скорости работы. В реальных условиях это очень незаметная оптимизация для конечного пользователя, но в своё время философия Apple строилась как раз на таких мелочах лишь для того, чтобы просто сэкономить время конечному потребителю.

Данный тип оптимизации характерен для рефакторинга кода, но сложно применяем на начальных этапах разработки.

Рассмотрим метод оптимизации кода, основанный на подходе, изложенном в [9], когда сборки библиотек классов группируют типы в легко развертываемые модули (DLL-файлы) и помещают в сборки библиотек классов, чтобы код стал доступен для других проектов (подход корпорации Microsoft). Второй тип оптимизации нельзя отнести к классическому, в связи с этим он рассматривается довольно редко, так как для крупных проектов это будет достаточно дорогостоящая процедура. Для не больших проектов данный тип оптимизации будет гораздо более целесообразен. Отличается он от первого тем, что подходит как для рефакторинга уже имеющегося кода, так и для написания любого кода с нуля.

Идея такого подхода заключается в построении кода таким образом, чтобы отдельные его компоненты были написаны на разных языках и, желательно, чтобы это были низкоуровневые языки программирования. Низкоуровневые языки программирования могут напрямую работать с памятью, большинство из них являются кроссплатформенными и, кроме того, они обеспечивают достаточную производительность.

Для анализа производительности оптимизации кода на C# был выбран язык программирования C++. Обусловлено это тем, что C++ имеет многолетнее развитие, поддержку и постоянно обновляемую стандартизацию. Отметим, что он остаётся одним из самых производительных. Также имеется большое число компиляторов, позволяющих компилировать код как угодно и под любую операционную систему.

Рассмотрим основные способы подключения внешнего кода. Отметим, что при этом есть несколько проблем: совместимости двух и более языков программирования (в случае возникновения легко решается); производительности (какой из способов подключения будет надёжнее и при этом будет обеспечивать большее быстроедействие); схожести языков (желательно, чтобы эти языки были достаточно схожими, чтобы «не изобретать заново велосипед»).

Первый способ достаточно простой и заключается в том, чтобы откомпилировать код на родном компиляторе одного языка программирования и передавать через уже скомпилированный код управление в исходное приложение, написанное на другом языке. Такой способ является ненадёжным в плане безопасности, т.к. высок риск, что данные могут перехватить. В то же время он является очень быстрым и, как правило, недорогим в разработке.

Второй способ, основанный на использовании скриптовых языков (например, Lua) отличается большей надёжностью. В настоящее время данный подход начал терять свою актуальность, но продолжает использоваться в крупных продуктах от компании Valve. Его главное преимущество заключается в скорости разработки, однако при этом возникают проблемы с производительностью, которые решить достаточно проблематично.

Третий способ – полноценное серверное решение. К его преимуществам следует отнести скорость работы (зависит от серверной машины) и надёжность. В целом грамотная реализация третьего способа - эталон бэкэнд разработки. К числу недостатков данного способа следует отнести высокую стоимость его реализации, вызванную необходимостью установки локального сервера на виртуальной машине. Это скорее замедлит процесс, чем ускорит его.

Последний способ – использование динамических библиотек. Недостаток данного способа только один - достаточно трудная настройка совместимых друг с другом компонентов. К числу преимуществ способа можно отнести скорость работы конечного кода, к которому подключают такие библиотеки, и его достаточно недорогая реализация. При его применении производительность будет зависеть от правильности скомпилированного кода. Однако на момент проведения исследования при использовании Visual Studio эта проблема не наблюдалась.

Очевидно, что самые интересные и верные решениями достижения поставленной цели обеспечивают третий и четвёртый способы. Для исследования лучше всего подойдёт четвёртый способ, так как арендовать сервер для пузырьковой сортировки нецелесообразно. Для определения способа сборки исходного кода на C++ были использованы инструменты Microsoft Visual Studio Community. Отметим, что вопрос громоздкости такого кода не так актуален, т.к. исследуется проблема быстрогодействия оптимизированного кода. Пусть создание массива и его заполнение останется на C#, а вот сама сортировка будет переписана на C++ и собрана в динамическую

библиотеку. Метод сортировки подключаем через эту библиотеку и делаем замеры времени, за которые происходит полная сортировка.

В таблице 2 представлены результаты исследования, из которых следует, что производительность исполнения повысилась и при этом обеспечивается более высокое быстродействие.

Таблица 2

Измерение скорости работы с заменой кода и без

Время на исполнение оригинального проекта, (с)	55.02	55.25	55.63	55.31	55.48	66.01
Время на исполнение проекта с заменой части кода, (с)	44.15	44.33	55.11	44.32	33.98	44.33

Подведём итоги и определим целесообразность применения рассмотренных проектов оптимизации кода. Сводные результаты замеров времени их исполнения представлены в таблице 3, а результаты расчётов средних значений приведены на рисунке 1.

Таблица 3

Сводные результаты измерения времени исполнения для рассмотренных проектов

Время на исполнение оригинального проекта, (с)	55.02	55.25	55.63	55.31	55.48	66.01
Время на исполнение проекта с классической оптимизацией, (с)	44.82	55.03	44.94	55.12	44.80	55.34
Время на исполнение проекта с заменой части кода, (с)	44.15	44.33	55.11	44.32	33.98	44.33

Среднее время выполнения неоптимизированного кода составляет 5,45 с. После классической оптимизации время работы незначительно уменьшилось до 5,01 с. Замена кода привела к снижению времени работы до 4,37с., что является уже значимым показателем. Очевидно, что оптимизация кода позволяет ускорить его выполнение. При этом повышается точность и быстродействие ПО. Отметим, что даже экономия времени выполнения кода менее секунды, в зависимости от размера проекта и его кода, может привести к экономии, выражаемой уже в минутах.

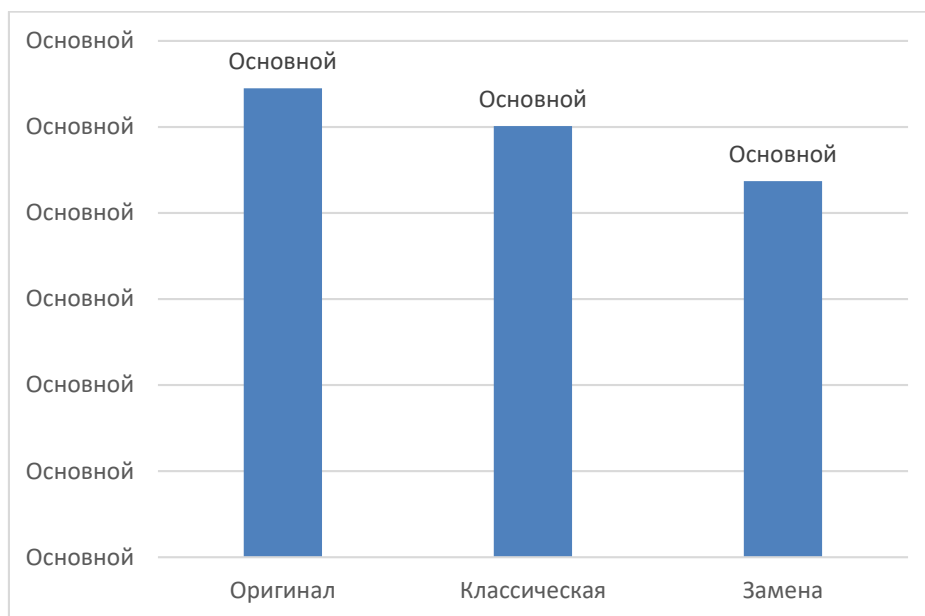


Рис. 1. Результаты измерения среднего времени исполнения кодов (в секундах) для разных проектов его оптимизации

Заключение

Результаты исследования и измерения показывают, что оптимизация кода позволяет ускорить его выполнение. При этом повышается точность и быстродействие ПО. Отметим, что даже экономия времени выполнения кода менее секунды, в зависимости от размеров проекта и его кода, может привести к значительной экономии временных затрат. Наиболее эффективным и актуальным представляется метод оптимизации с заменой части кода, учитывая быстрое развитие совместимости каждого из его отдельных компонентов.

Литература

1. Конов А. А., Докучаев В. А. Особенности проектирования системы мониторинга показателей веб-приложений и сайтов для гостиничного бизнеса // Перспективные технологии в средствах передачи информации: материалы 14-ой международной научно-технической конференции, Владимир, 06-07 октября 2021 года. Владимир: Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых, 2021. С. 301-306. EDN KRCJMK.
2. Докучаев В. А., Конов А. А. Актуальность разработки и применения системы мониторинга веб-показателей в гостиничном бизнесе // Актуальные проблемы и перспективы развития экономики : Труды Юбилейной XX Всероссийской с международным участием научно-практической конференции, Симферополь – Гурзуф, 11-13 ноября 2021 года. Симферополь: Издательский дом КФУ, 2021. С. 39-42. EDN FOWHDE.
3. Оптимизация программного кода. URL: <https://it-black.ru/optimizacija-programmnogo-koda> (дата обращения: 08.11.2022).
4. Dokuchaev V. A., Maklachkova V. V., Pavlov S. V., Mytenkov S. S. Comparison of the Spectral and Hierarchy Analysis Methods Applicability for Technical Systems Choosing // 2021 Systems of Signals Generating and Processing in the Field of on Board Communications, Conference Proceedings, Moscow, 16-18 марта 2021 года. Moscow, 2021. P. 9416017. DOI 10.1109/IEEE-CONF51389.2021.9416017. – EDN ZWJYTX.
5. Оптимизация программного кода. URL: <https://techrocks.ru/2019/01/25/code-optimization-tips> (дата обращения: 08.11.2022).
6. Конов А. А., Гадасин Д. Д., Докучаев В. А. Исследование проблемы мониторинга ошибок неконтролируемого программного окружения // Технологии информационного общества : Сборник трудов XV Международной отраслевой научно-технической конференции «Технологии информационного общества», Москва, 03–04 марта 2021 года. М.: Издательский дом Медиа публицер, 2021. С. 200-202. EDN MGCBBF.
7. Летунов А. С., Докучаев В. А. W-code платформы, как инструмент разработки программного обеспечения // Телекоммуникации и информационные технологии. 2019. Т. 6. № 2. С. 123-128. EDN DTBNNQ.
8. Характеристики производительности приложений .NET Framework: сайт. URL: https://professorweb.ru/my/csharp/optimization/level1/1_1.php (дата обращения: 03.10.2022).
9. Прайс Марк. С# 9 и .NET 5. Разработка и оптимизация. СПб.: Питер, 2022. 832 с. (Серия «Для профессионалов»). ISBN 978-5-4461-2921-8.

C# CODE ANALYSIS AND OPTIMIZATION

Andrey O. Chubarov,
Graduate MTUCI, Moscow, Russia,
andrew.chubarov@internet.ru

Vladimir A. Dokuchaev
Head of the Department NIT&S MTUCI, Doctor of Technical Sciences,
Professor, MTUCI, Moscow, Russia,
v_dok@mail.ru

Abstract

Considered are currently widely used classical methods for improving the efficiency and speed of the final software code. The analysis of two types of optimization for code written in C# was carried out. The results of applying the method of replacing part of the code with a lower-level one in order to improve software performance are given.

Keywords: code optimization, C #, C-like programming languages, performance analysis, code.

РАЗРАБОТКА АЛГОРИТМА ЗАЩИТЫ СЕРВЕРА БАЗЫ ДАННЫХ ОТ СЕТЕВЫХ АТАК

Мясников Денис Юрьевич,
магистрант МТУСИ, Москва, Россия,
myasnicov2011@yandex.ru

Смирнов Алексей Эдуардович,
доцент кафедры ИБ МТУСИ, к.т.н., Москва, Россия,
smirnov.al.ed@gmail.com

Аннотация

В статье приведены результаты исследования основных механизмов защиты сервера базы данных от сетевых атак. Представлены схемы работы механизмов защиты сервера базы данных от сетевых атак. Разработан алгоритм защиты сервера базы данных от сетевых атак и обосновано назначение каждый из выбранных механизмов защиты базы данных. Актуальность использования разработанного алгоритма обусловлена увеличением количества атак на базы данных с использованием сетевых атак, а также возрастающим количеством утечек конфиденциальных данных.

Ключевые слова: информационная безопасность, защита информации, сетевые атаки, защита баз данных, алгоритм защиты от атак.

Введение

В настоящее время в организациях и крупных компаниях совершается огромное множество утечек конфиденциальных данных из баз данных [1]. Даже при высоких требованиях к уровням защищённости баз данных, количество утечек ежегодно возрастает, а так как современный мир не может обойтись без информационных технологий, это является серьёзной проблемой.

Одной из ключевых опасностей и угроз конфиденциальности, целостности и доступности защищаемым серверам баз данных является угроза сетевых атак. Сетевые атаки не только могут вывести из работы сервера баз данных компании, но также могут и способствовать утечке конфиденциальных данных [2], что может повлечь очень серьёзные проблемы, штрафы и издержки компании. Поэтому разработка алгоритмов защиты сервера базы данных от сетевых атак весьма актуальна.

1. Механизмы защиты сервера базы данных от сетевых атак

Для разработки алгоритма защиты сервера базы данных от сетевых атак необходимо рассмотреть все основные механизмы защиты сервера.

Среди основных механизмов защиты сервера баз данных от сетевых атак выделим следующие механизмы: межсетевой экран; *SSL/TLS* (*Secure Sockets Layer/Transport Level Security*; Протокол защищённых сокетов/Протокол защиты транспортного уровня); *SSH* (*Secure Shell*, защищённая оболочка) ключи; *port knocking* (простукивание портов); мониторинг и аудит; системы обнаружения и предотвращения вторжений; шифрование; автоматизированные системы защиты.

1.1. Межсетевой экран

Межсетевые экраны являются устройствами или системами, которые управляют потоком сетевого трафика между сетями с различными требованиями к безопасности [3]. Межсетевые экраны могут быть программными или программно-аппаратными элементами в компьютерной сети. Обычно межсетевые экраны рассматриваются в контексте стека протоколов *TCP/IP*. Как правило, чем мощнее межсетевой экран, тем больше большее количество уровней стека протоколов *TCP/IP* он покрывает.

Механизмы фильтрации трафика в межсетевых экранах разделяются на простые и сложные. Существуют две базовые стратегии работы межсетевых экранов:

- 1) разрешено всё, что явно не запрещено;
- 2) запрещено всё, что явно не разрешено.

При применении первого варианта межсетевой экран разрешает все подключения, которые явно не запрещены правилами фильтрации. При применении второго варианта межсетевой экран запрещает все подключения, кроме тех, которые явно указаны в правилах фильтрации. Чаще всего применяется второй вариант работы межсетевых экранов, так как контролировать только разрешённые подключения намного легче, чем контролировать запрещённые.

На рисунке 1 приведён процесс работы межсетевого экрана.

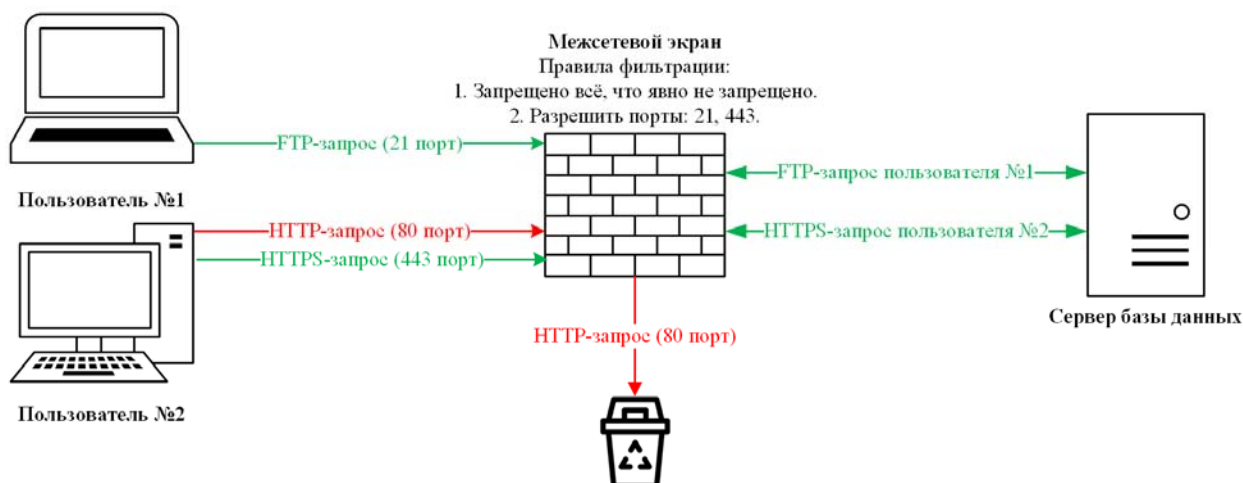


Рис. 1. Процесс работы межсетевого экрана

1.2. SSL/TLS-шифрование

SSL/TLS-протоколы – это криптографические протоколы, которые позволяют установить безопасное соединение для передачи информации в сети Интернет. Чаще всего, данные протоколы используются совместно с протоколом HTTP. Сетевые протоколы SSL/TLS могут использоваться не только для аутентификации, но и для шифрования передаваемых данных.

Протокол SSL спроектирован для обеспечения конфиденциальности обмена между двумя прикладными процессами клиента и сервера [4]. Опционально данный протокол может не только аутентифицировать сервер, но и может аутентифицировать клиента. При такой двусторонней аутентификации сервер посылает сообщение *CertificateRequest* для получения сертификата у пользователя.

Одним из преимуществ SSL является его независимость от прикладного уровня протоколов, таких, как: TELNET, HTTP, FTP. Данные протоколы могут работать с SSL, а протокол SSL будет согласовывать алгоритмы шифрования, процессы аутентификации сервера и клиента, ключи сессии до начала работы приложения с ранее приведёнными протоколами.

Протокол SSL предоставляет безопасный канал, у которого есть три основных свойства [4]:

1. Частный канал. После простой последовательности запросов для установления закрытого ключа все остальные сообщения шифруются.
2. Аутентифицированный канал. Сервер всегда аутентифицируется, а также есть возможность аутентифицировать и клиента.
3. Надёжный канал. Проверяется целостность сообщений с использованием MAC.

Протокол TLS применяется для защищённой передачи данных между двумя узлами сети Интернет. У данного протокола имеется два уровня: протокол записей TLS и протокол диалогов TLS. Данный протокол работает поверх транспортных протоколов TCP/UDP. Так же, как и SSL, TLS обладает тремя основными свойствами: частность, надёжность, аутентифицированность. Для шифрования используется симметричная криптография (RSA, DES, RC4), а ключ генерируется независимо для каждого подключения.

Стоит отметить, что TLS основан на действующей спецификации SSL, а так как SSL устарел, то разработчики программного обеспечения используют связку двух сертификатов SSL/TLS, что позволяет поддерживать и старые, и новые устройства. Именно поэтому SSL и TLS объединены.

На рисунке 2 детально показан механизм работы SSL/TLS-протоколов.

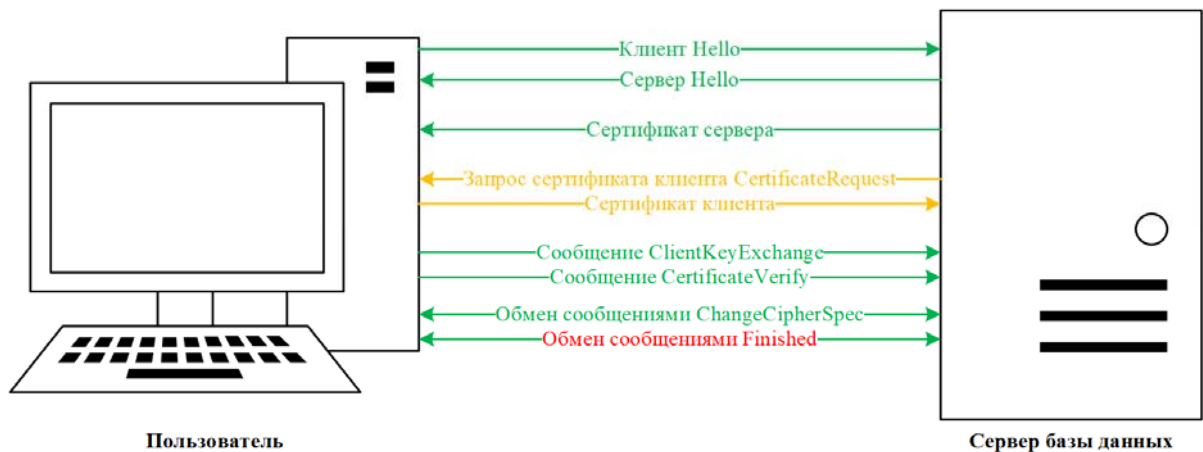


Рис. 2. Механизм работы SSL/TLS-протоколов

1.3. SSH-ключи

SSH – это сетевой протокол, который предназначен для безопасного защищённого удалённого доступа [5]. Основной целью SSH является обеспечение защищённого зашифрованного доступа между двумя узлами в защищённой сети. Для достижения защищённого зашифрованного соединения SSH использует криптографические методы с открытыми ключами. Стоит отметить, что SSH в основном работает по принципу клиент-сервер. Одним из основных преимуществ SSH является то, что SSH позволяет удалённо выполнять команды и отправлять файлы, но при этом требует аутентификацию и шифрует каналы связи. То есть, сервер позволяет клиенту подключиться по зашифрованному каналу связи с использованием закрытых и открытых ключей SSH.

Для конфигурирования аутентификации с помощью SSH-ключей необходимо, чтобы закрытый ключ клиента хранился в тайне, а открытый ключ был размещён на сервере. Далее, клиенту необходимо предоставить ответ серверу с помощью закрытого ключа, для получения доступа к серверу. Сервер зашифровывает сообщение с помощью открытого ключа и отправляет клиенту. При перехвате злоумышленником данного сообщения ничего не произойдёт, так как злоумышленник не сможет расшифровать информацию с помощью открытого ключа. В конце с помощью закрытого ключа клиент отвечает серверу.

На данный момент есть две версии SSH-1 и SSH-2. Так как SSH-2 – это обновленный SSH-1, то из SSH-2 были убраны устаревшие алгоритмы, которые подвергали опасности защищённое соединение. Также, во вторую версию добавили более стойкие алгоритмы шифрования и исправили некоторые ошибки предыдущей версии. На рисунке 3 показан механизм работы сетевого протокола SSH.

1.4. Port knocking

Port knocking – это защитный сетевой механизм, который позволяет открывать определённые порты с помощью попыток подключения к закрытым портам в определённой комбинации. После получения правильной комбинации попыток подключения к портам открывается нужный порт, что позволяет клиенту установить безопасное соединение с сервером.

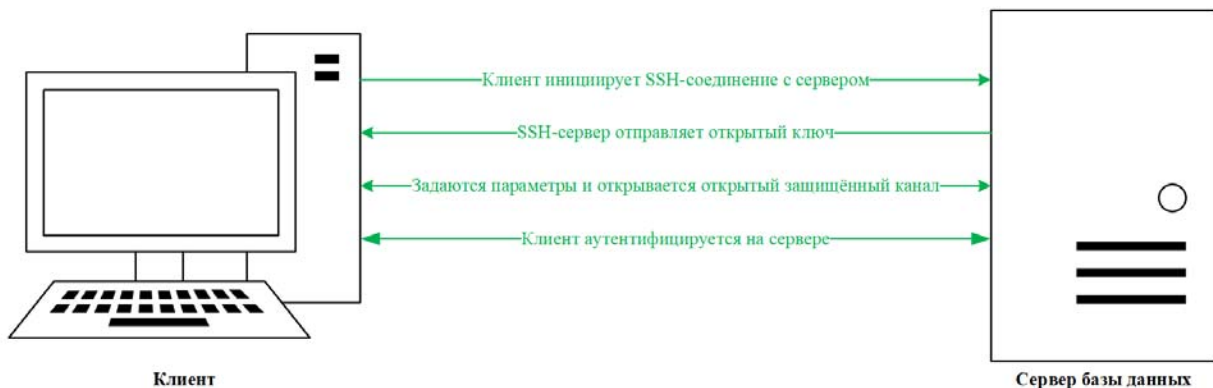


Рис. 3. Механизм работы сетевого протокола SSH

Одной из основных целей *port knocking* является защита от сканирования портов злоумышленником, так как если злоумышленник не отправит правильную комбинацию, то все защищённые порты будут казаться для него закрытыми.

Port knocking обычно реализуется с помощью межсетевой экрана, так как даже если порты закрыты и защищены, межсетевой экран всё равно способен обрабатывать попытки подключения.

С помощью *port knocking* можно не только открывать порты, но и выполнять другие полезные действия [6]. Например, можно настроить систему так, что после введения определённой комбинации система будет отключаться или перезагружаться, а при введении другой комбинации будет открываться необходимый порт.

Реализация *port knocking* возможна на нескольких уровнях, начиная от уровня ядра и заканчивая пользовательскими уровнями. Сами попытки подключения могут быть составлены из любого количества пакетов *TCP*, *UDP*, *ICMP* и пакетов других протоколов передачи данных. Также можно задать, чтобы проверялась не только комбинация, но и длина пакетов, что усилит безопасность, а также в какой-то степени защитит от угрозы сканирования трафика, передаваемого между клиентом и сервером. Как правило, прослушивание портов задаётся с помощью двух способов:

- 1) Перехват отдельных пакетов;
- 2) Журнал межсетевой экран.

На рисунке 4 показан процесс работы защитного сетевого механизма *port knocking*.

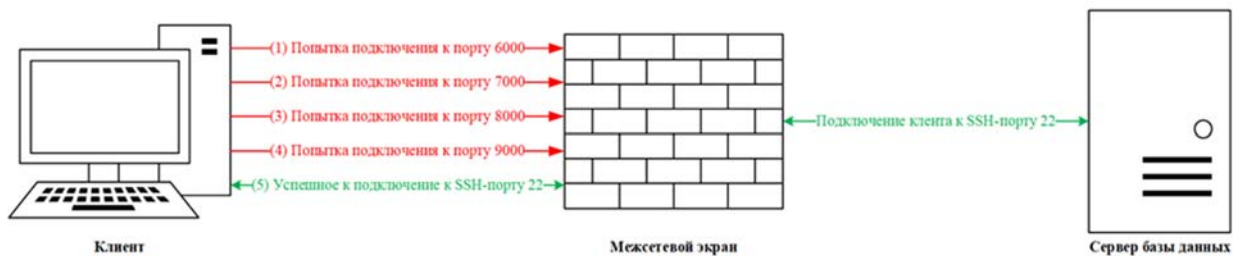


Рис. 4. Процесс работы механизма *port knocking*

1.5. Мониторинг и аудит

Аудит – это средства учёта и наблюдения, обеспечивающие возможности обнаружения и фиксирования важных событий, связанных с безопасностью, а также любые попытки получения доступа или удаления системных ресурсов [7]. Мониторинг очень близок по значению к аудиту. Он вводится для наблюдения за работой операционной системы, за базами данных и за пользователем. Аудит и мониторинг снижают риск внутренних угроз [7].

Аудит рекомендуется проводить регулярно, чтобы можно было оперативно оценить и найти слабые места в системе для дальнейшего исправления. Аудит и мониторинг можно проводить встроенными средствами систем управления базами данных. Такие системы анализируют базы данных, данные систем управления взаимоотношениями с клиентами и данные систем планирования ресурсов. После анализа аудиторю предоставляются предупреждения с возможностью создания встроенного или пользовательского отчёта. Также, в системах управления базами данных есть встроенные журналы запросов, которые ведутся независимо от всего. Они фиксируют кто, когда и что сделал.

На рисунке 5 показано функционирование аудита и мониторинга базы данных.

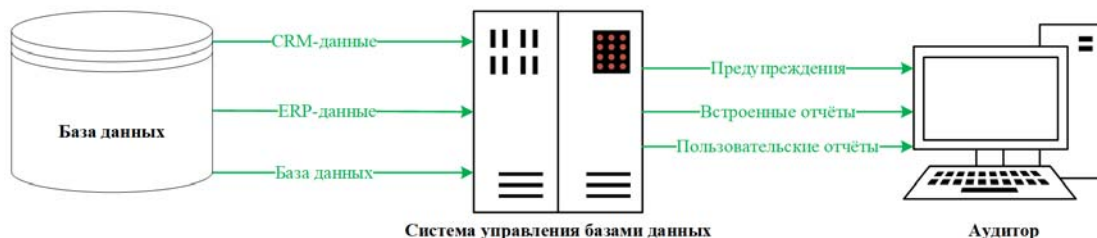


Рис. 5. Аудит и мониторинг базы данных

1.6. Системы обнаружения и предотвращения вторжений

Системы обнаружения и предотвращения вторжений представляют программные или программно-аппаратные решения, автоматизирующие процессы сбора, хранения и анализа событий, происходящих в компьютерной системе с возможностью их самостоятельного анализа [8]. Сетевые системы обнаружения и предотвращения вторжений производят мониторинг сетевого трафика в режиме реального времени на предмет выявления заранее определённых сигнатур аномального поведения. К основным функциям систем обнаружения и предотвращения вторжений можно отнести:

- Обнаружение сетевых атак.
- Контроль качества соединений.
- Сканирование на предмет уязвимостей сети.
- Распознавание источников атак.

Системы обнаружения и предотвращения вторжений делятся на два типа:

1. Системы обнаружения вторжений.
2. Системы предотвращения вторжений.

Системы обнаружения вторжений стремятся как можно быстрее обнаружить атаку, зарегистрировать её и уведомить об атаке пользователя.

На рисунке 6 показан процесс работы системы обнаружения вторжений.

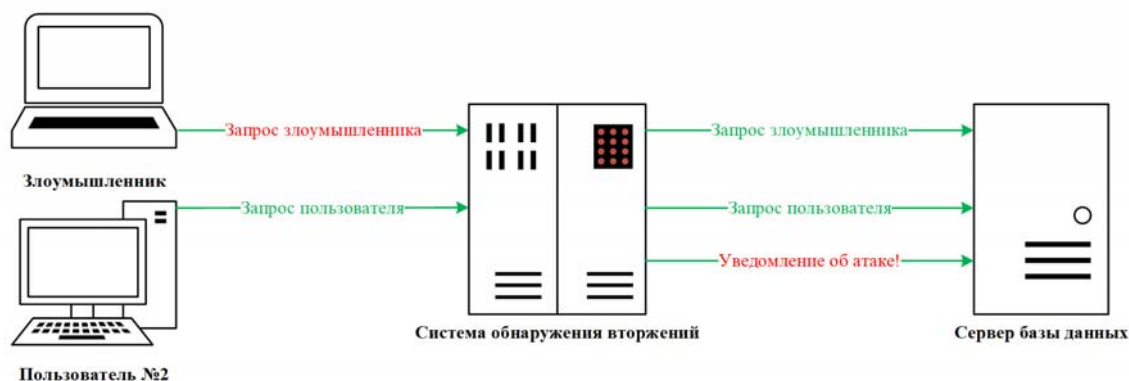


Рис. 6. Процесс работы системы обнаружения вторжений

В ходе совершенствования систем обнаружения вторжений появились улучшенные системы предотвращения атак. Системы предотвращения атак основаны на системах обнаружения атаки и, следовательно, они могут выполнять те же функции, что и системы обнаружения атак. Помимо этого, к их функционалу добавляются системы предотвращения вторжений для моментального реагирования на обнаружение сетевых атак.

На рисунке 7 показан процесс работы системы предотвращения вторжений.



Рис. 7. Процесс работы системы предотвращения вторжений

1.7. Шифрование

Шифрование – это кодирование данных при помощи специальных алгоритмов, в результате чего данные становятся недоступными для чтения [9]. Чтобы данные было возможно прочитать, требуется знание секретного ключа. Подразумевается, что для баз данных будет использоваться устойчивый криптоалгоритм. Как правило, шифруются не только база данных, но и все возможные настройки, и резервные копии. В случае получения доступа злоумышленником к данным, он увидит информацию в «нечитаемом» виде и, тем самым, не сможет ею

воспользоваться. При шифровании данных перестаёт работать индексация, что является огромным минусом шифрования. Также происходит очень большая потеря в производительности, так как для системы шифрования-дешифрования требуют значительного времени и ресурсов. Но, некоторые системы управления базами данных предлагают встроенные методы шифрования данных с поддержкой индексации, и с гораздо меньшими потерями производительности.

На рисунке 8 показан процесс работы механизма шифрования.

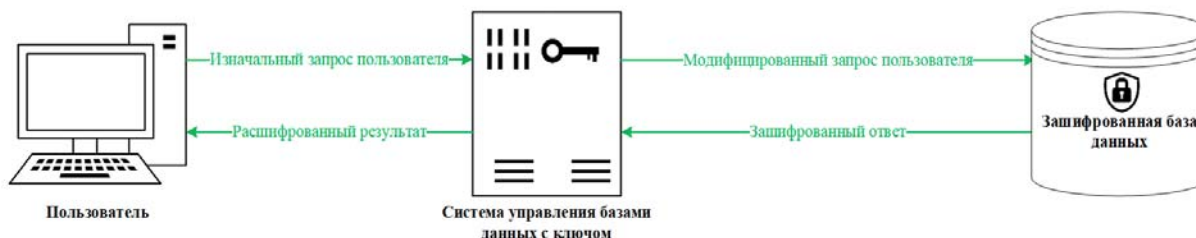


Рис. 8. Процесс работы механизм шифрования

Автоматизированные системы защиты

Автоматизированные системы защиты – это специальные системы защиты безопасности баз данных. Автоматизированные системы защиты представлены двумя типами систем [10].

1. Система мониторинга активности базы данных (*DAM*).
2. Система брандмауэра базы данных (*DBF*).

Главным отличием данных систем является их независимость от систем управления базами данных. Система мониторинга активности баз данных работает пассивно с копией трафика базы данных, тем самым не замедляя работу.

Данная система позволяет классифицировать трафик, вести полный аудит, использовать персонализированную систему фильтрации, а также выявлять аномалии в работе баз данных. К примеру, если в определённый момент пользователь произведёт в десять раз больше запросов к базе данных чем обычно, то данная система отправит уведомление с аномалией администратору. При этом система гибко анализирует весь трафик, сводя его к простым графикам и диаграммам. К основным функциям системы мониторинга активности баз данных можно отнести:

- анализ трафика пользователя;
- фильтр статистики;
- аудит запросов;
- классификация информации;
- проверка уязвимостей.

На рисунке 9 показан процесс работы системы мониторинга активности базы данных.

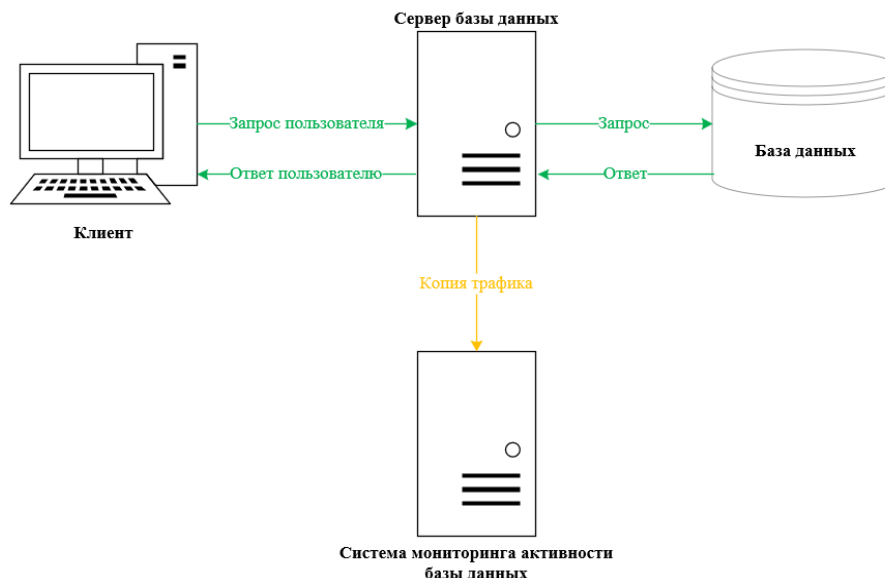


Рис. 9. Процесс работы системы мониторинга активности базы данных

Система брандмауэра может работать только активно с трафиком, в отличие от системы мониторинга активности базы данных. Данная система представляет собой брандмауэр с дополнительным анализом и мониторингом пользователей с помощью поведенческого анализа.

На рисунке 10 показан процесс работы системы брандмауэра базы данных.

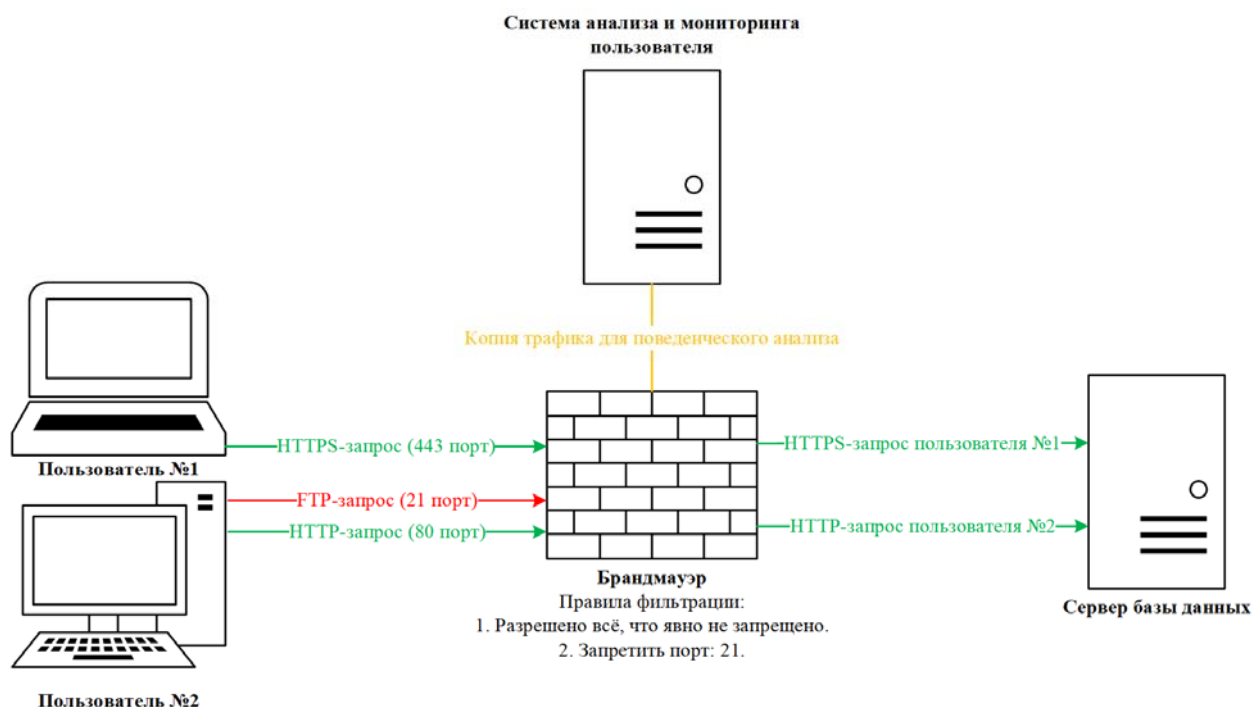


Рис. 10. Процесс работы системы брандмауэра базы данных

2. Разработка алгоритма защиты сервера базы данных от сетевых атак

Для разработки алгоритма были задействованы следующие механизмы защиты сервера базы данных от сетевых атак:

- SSL/TLS-шифрование.
- Port knocking.
- Мониторинг.
- Системы обнаружения и предотвращения вторжений.
- Криптографическое шифрование.

SSL/TLS-шифрование будет использоваться для установления безопасного соединения между клиентом и сервером, а также для безопасной передачи данных в сети Интернет посредством симметричной криптографии.

Port knocking будет использоваться для открытия необходимых портов на сервере, путём ввода неправильной комбинации. Port knocking поможет защититься от сканирования портов сервера базы данных злоумышленником.

Мониторинг позволит построить статистические профили пользователей и провести поведенческий анализ для поиска аномалий по отклонениям от статистического использования.

Системы обнаружения и предотвращения вторжений помогут обнаружить атаки вида: DoS (SYN Flood, UDP Flood, ICMP Flood), сканирования портов, SQL-инъекции.

Использование встроенного криптографического шифрования для хранения данных в зашифрованном виде на уровне приложения защитит данные от несанкционированного использования.

В итоге, на рисунке 11 приведена блок-схема разработанного алгоритма защиты сервера базы данных от сетевых атак.



Рис. 11. Блок-схема разработанного алгоритма защиты сервера базы данных от сетевых атак

Заключение

На основе анализа основных известных механизмы защиты серверов баз данных от сетевых атак продемонстрированы их ключевые особенности и эффективность функционирования

Разработан алгоритм для защиты сервера базы данных от сетевых атак, который включает в себя: *SSL/TLS*-шифрование, *port knocking*, мониторинг, систему обнаружения и предотвращения вторжений и криптографическое шифрование. Обосновано использование в разработанном алгоритме соответствующих механизмов защиты, а также приведены примеры сетевых атак, которые блокируются данными механизмами, что повышает общую защищённость серверов баз данных, а также снижает количество утечек конфиденциальных данных.

Литература

1. Утечка данных «Яндекс.Еды» и её последствия. URL: <https://versia.ru/utechka-dannyx-yandeks-edy-i-eyo-posledstviya>, Версия, 2021 г. (дата обращения: 05.10.2021).
2. Боршевников А. Е. Сетевые атаки. Виды. Способы борьбы. 2011. 8 с.
3. Лапонина О. Р. Межсетевое экранирование. Основы информационных технологий. 2014. 3 с.
4. Семёнов Ю. А. Телекоммуникационные и информационные технологии. 2019. 701 с.
5. Полное руководство по SSH в Linux и Windows. URL: <https://hackware.ru/?p=10059>, 2019 г. (дата обращения: 10.02.2022).
6. Защита сервера при помощи Port Knocking. URL: <https://putty.org.ru/articles/port-knocking.html>, 2022 г. (дата обращения: 17.02.2022).
7. Туриканов Т. С., Айтхожаева Е. Ж. Организация аудита и мониторинга в MS SQL SERVER. 2015. 621 с.
8. Грицкевич В. И., Петров С. Н. Особенности современных средств обнаружения вторжений. 2018. 67 с.
9. Скаун В. В. Защита информации в базах данных и экспертных системах. 2015. 13 с.
10. Обзор Гарда БД 4.19, российской DAM-системы для защиты СУБД. URL: <https://www.anti-malware.ru/reviews/Garda-DB-419>, 2022 г. (дата обращения: 03.03.2022).

**DEVELOPMENT OF AN ALGORITHM FOR PROTECTING A DATABASE SERVER
FROM NETWORK ATTACKS**

Myasnicov Denis Yurevich,
Graduate MTUCI, Moscow, Russia,
myasnicov2011@yandex.ru

Smirnov Aleksei Eduardovich,
Associate Professor of the Department of IS MTUCI,
Ph. D., Moscow, Russia,
smirnov.al.ed@gmail.com

Abstract

This article explored the basic mechanisms for protecting the database server from network attacks. The schemes of operation of mechanisms for protecting the database server from network attacks are presented. Based on the studied mechanisms, an algorithm was developed to protect the database server from network attacks. It was justified and told what each of the selected database protection mechanisms is used for. The relevance of the developed algorithm is due to the increase in the number of attacks on databases using network attacks, as well as the increased number of leaks of confidential data.

Keywords: *information security, information protection, network attacks, database protection, attack protection algorithm.*

ОБ ИСПОЛЬЗОВАНИИ СИСТЕМНЫХ ВЫЗОВОВ WIN-API ДЛЯ ОБНАРУЖЕНИЯ МОДИФИЦИРОВАННОГО ВРЕДНОСНОГО ПО

Зайченко Иван Андреевич,
студент МТУСИ, Москва, Россия,
ivan.zaychenko.2001@mail.ru

Большаков Александр Сергеевич,
доцент кафедры ИБ МТУСИ, к.т.н., Москва, Россия,
alexbol57@mail.ru

Аннотация

Рассмотрен альтернативный статическому способу, способ определения признаков вредоносного программного обеспечения (ВПО) в ОС Windows путем мониторинга системных вызовов Win-API, способу. В качестве метода выделения признаков вредоносности из последовательности системных вызовов исследуемого приложения предложен алгоритм LCS. Показано, что, используя алгоритм LCS, мониторинг системных вызовов является актуальным способом исследования ПО на предмет предотвращения нарушения ИБ. Обоснован вывод о целесообразности построения «поведенческой» БД из последовательностей системных вызовов известных «первозданных» вирусов в качестве компоненты системы антивирусной защиты (САЗ) для поиска модифицированного ВПО с логикой работы, адекватной «первозданному» ВПО.

Ключевые слова: системные вызовы, антивирусная защита, информационная безопасность, метаморфическое вредоносное ПО, LCS, Python, C++.

Введение

Одними из самых популярных способов распознавания вредоносного ПО являются: расчет хэш-суммы исполняемого файла программы и сверка ее с некоторой базой данных известных вирусов; анализ исполняемого кода программ для поиска известных «вирусных» блоков кода. Указанные способы анализа называются статическими [1], так как они анализируют код программы без ее запуска.

Проблема статических подходов заключается в том, что они сильно полагаются на (почти) неизменный вид анализируемых программ. Злоумышленник же может сильно изменить структуру вредоносного ПО, добавив туда, например, дополнительный код и данные, или например, зашифровав вирусный блок кода, что сильно затруднит обнаружение или сделает даже сделает его невозможным статическим способом.

Однако, если сама логика работы вируса не претерпела больших изменений, при запуске программы он будет делать те же операции, что и его версия до модификации. Таким образом, возникает идея анализа действий запущенных программ и выявления в них некоторых последовательностей операций, которые характерны для различных вирусов.

Поскольку все программы работают в некоторой ОС, то операции, выполняемые ими для взаимодействия с ОС, представляют собой системные вызовы. В различных ОС существуют свои наборы системных вызовов для использования программистами (так называемые API). В случае ОС Windows имеется набор Win-API [2]. Анализируя последовательности системных вызовов программ, можно выделять в них подпоследовательности, характерные для логики работы известных вирусов, и тем самым успешно определять вредоносное ПО даже если структура кода программы претерпела значительные изменения.

Целью данной работы является исследование возможности использования мониторинга Win-API в качестве метода обнаружения признаков ВПО в приложениях, работающих в ОС Windows. В данной статье предложен способ анализа «входящих» программ на предмет возможного нарушения ИБ путем сравнения последовательностей системных вызовов Win-API исследуемой программы с заведомо известными последовательностями системных вызовов Win-API для вирусов определенных классов. Данный подход использует алгоритм LCS (Longest Common Subsequence) [3] поиска наибольшей общей подпоследовательности среди двух последовательностей.

Выделение наибольшей общей подпоследовательности вызовов API программ как способ оценки принадлежности программ к некоторому классу вирусного ПО

Алгоритм нахождения общей наибольшей подпоследовательности среди двух последовательностей – LCS работает следующим образом: например, если имеется две последовательности целых чисел: $A = \{12, 25, 33, 55, 11, 9\}$ и $B = \{89, 12, 25, 100, 55, 9\}$, то их наибольшей общей подпоследовательностью будет последовательность: $L = \{12, 25, 55, 9\}$. Таким образом, наибольшая общая подпоследовательность должна содержаться в обеих

последовательностях, для которых она ищется (отсюда следует, что размер подпоследовательности не может быть больше минимальной по размеру из A и B). Значит, по размеру полученной подпоследовательности возможно судить о количестве общих элементов из A и B. Если A некий «эталон» последовательности, которую необходимо выделить в B (или хотя бы какую-либо ее часть), то имеет смысл просто найти отношение размера наибольшей общей подпоследовательности A и B к размеру A, чтобы показать, сколько процентов от числа элементов A удалось найти в B. Так, для примера, указанного выше положим $n = \text{кол} - \text{во элементов в } A = 6$, $m = \text{кол} - \text{во элементов в } L = 4$. Записав отношение m / n , получим $4/6 \approx 0,66$. Значит 66% элементов последовательности A содержатся в B.

Важно отметить, что сущность объектов для работы алгоритма не важна, главное, чтобы для объектов было определено понятие идентичности (т.е. для целых чисел из примера – это точное равенство). Если в качестве объектов взять слова, то понятие идентичности определяется следующим образом: одинаковое количество букв в каждом слове и совпадение каждой k-ой из букв. Очевидно, что если обе входные последовательности абсолютно идентичны, то на выходе алгоритм даст подпоследовательность также идентичную входным последовательностям.

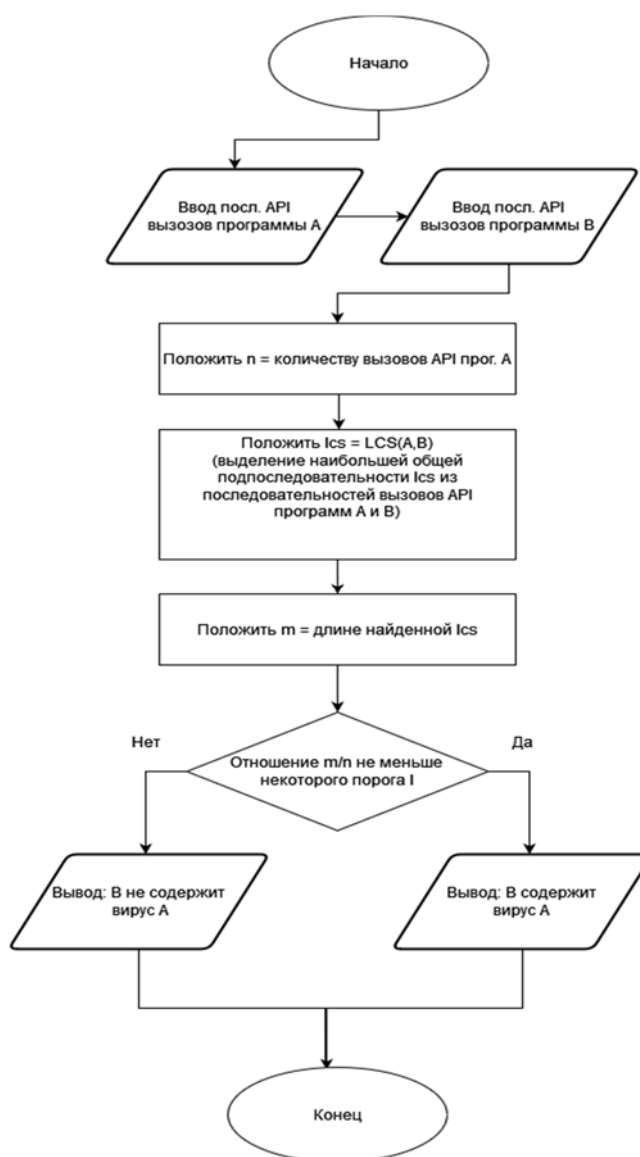


Рис 1. Применение алгоритма LCS для выявления вредоносного ПО

С целью определения возможности применения алгоритма LCS для построения САВЗ и целесообразности использования мониторинга Win-API были выполнены следующие процедуры:

1. Использованы три программы:

- программа «А», априори вредоносная, информация о структуре вредоносного кода которой заранее известна и хранится в открытой БД;

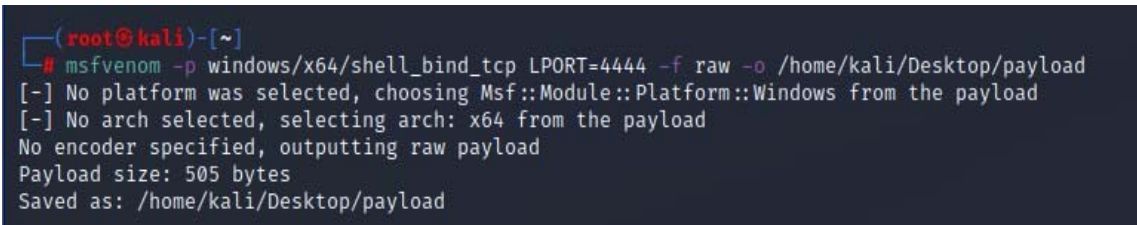
- программа «В», выполняющая такие же вредоносные операции, что и «А», но модифицированная путем внедрения в нее «полезной» не вредоносной нагрузки;
 - программа «С», априори не вредоносная;
2. Записаны системные вызовы *API* программ «А», «В» и «С», совершенные ими во время работы.
 3. Алгоритм LCS, изображенный на рисунке 1, реализован для последовательностей системных вызовов двух пар программ:
 - «А» и «В»
 - «А» и «С»
 4. Согласно алгоритму LCS, изображенному на рисунке 1, получены следующие результаты для пар программ п. 3:
 - в вызовах *API* программы «В» успешно распознаны вредоносные последовательности вызовов *API*, характерные для программы «А»;
 - в вызовах *API* программы «С» не распознаны вредоносные последовательности вызовов *API*, характерные для программы «В».

Смысл данных действий заключается в том, что если в последовательности системных вызовов *API* анализируемой программы содержится подпоследовательность (какой-либо части или всех) вызовов вредоносной программы «А», то алгоритм *LCS* найдет эту подпоследовательность. Причем, это означает, что последовательность вызовов анализируемой программы может содержать и другие вызовы (т.е. помимо вредоносных вызовов *API* «А», анализируемая программа может, параллельно делать и не вредоносные операции).

Тогда, записав отношение размера подпоследовательности найденной алгоритмом *LCS*, к размеру последовательности вызовов «А», можно судить насколько анализируемая программа «близка» к вредоносной «А». Поскольку для выявления в анализируемой программе *API* вызовов вируса «А» требуется знать лишь *API* вызовы, выполненные ею во время работы, то структура исполняемого файла программы для этого метода абсолютно не важна.

Пример использования алгоритма LCS в качестве метода оценки принадлежности программ к некоторому классу вирусного ПО

1. Выбор программ «А», «В» и «С»:
 - Программа «А» запускает блок кода вируса «*Shell Bind TCP*», сгенерированного в ОС *Kali-Linux* (рис. 2) в специально созданном для этого средствами ОС дочернем потоке выполнения исполняемого кода. Главный поток программы не выполняет никаких действий и лишь ожидает завершения потока, в котором исполняется вирус.



```
(root@kali)-[~]
└─# msfvenom -p windows/x64/shell_bind_tcp LPORT=4444 -f raw -o /home/kali/Desktop/payload
[-] No platform was selected, choosing Msf::Module::Platform::Windows from the payload
[-] No arch selected, selecting arch: x64 from the payload
No encoder specified, outputting raw payload
Payload size: 505 bytes
Saved as: /home/kali/Desktop/payload
```

Рис 2. Создание блока кода вируса «*Shell Bind TCP*» в *Kali-Linux*

На рисунке 3 показана логика работы программы «А», а на рисунке 4 показан пример реализации программы «А». Такая архитектура приложения позволяет определить все системные вызовы *API*, совершаемые вирусом.

Таким образом, программа «А» – это «эталон» для вируса класса «*Shell Bind TCP*».

- Программа «В» - модифицированная версия «А», где блок кода вируса «*Shell Bind TCP*» модифицирован (это нужно для того, чтобы вирус в файле программы не выглядел знакомым для САВЗ). При запуске программа «В» возвращает вирус в исходное состояние и запускает его в дочернем потоке, а в главном потоке – выполняет «полезную» работу, не имеющую вредоносный характер.

На рисунке 5 показана логика работы программы «В».

На рисунке 6 показан пример реализации программы «В».

- Таким образом, программа «В» маскирует блок кода вируса «*Shell Bind TCP*» в своем исполняемом файле, а также маскирует его операции, выполняя дополнительную «полезную» работу. В качестве приложения «С», не содержащего вируса «*Shell Bind TCP*», выбран стандартный в ОС *Windows* «Блокнот», в котором был набран случайный текст.



Рис 3. Логика работы программы «А»

```

//Программа А - эталон вируса
#include <windows.h>
#include <iostream>

//Блок исполняемого (ассемблерного) кода вируса,
//записанного как массив байт в шестнадцатеричном представлении.
char VIRUS_ASM[] = {
    0xAA, 0xBB, 0xCC, 0xDD,
    0xEE, 0xFF, 0xAB, 0xAC,
    0xAD, 0xAE, 0xAF, 0xAF,
    //В целях безопасности и для краткости,
    //приведенные байты не являются реальным вирусом,
    //а лишь служат примером того как можно встроить вирус в программу.
};

//Данная функция запускает ассемблерный код вируса,
//который передается ей в виде массива байт
HANDLE loadExecutableCode(const char* execCode, size_t size) {
    //Выделение памяти для кода (с разрешением на исполнение!)
    PVOID pExec = VirtualAlloc(0, size, MEM_COMMIT | MEM_RESERVE, PAGE_EXECUTE_READWRITE);
    //Копирование кода в эту память
    RtlCopyMemory(pExec, execCode, size);
    //Создание потока выполнения кода
    DWORD threadID;
    HANDLE hThread = CreateThread(NULL, 0, (PTHREAD_START_ROUTINE)pExec, NULL, 0, &threadID);
    //Дескриптор потока возвращается,
    //для возможности его ожидания и удаления в main()
    return hThread;
}

//Главная функция программы
int main() {
    //Запуск вируса в дочернем потоке
    auto hThread = loadExecutableCode(VIRUS_ASM, sizeof(VIRUS_ASM));
    //Ожидание завершения исполнения кода вируса в дочернем потоке
    WaitForSingleObject(hThread, INFINITE);
    //Удаление дескриптора потока
    CloseHandle(hThread);
    return 0;
}
    
```

Рис 4. Пример реализации программы «А» на C++



Рис 5. Логика работы программы «В»

```

//Программа В - модификация А
//Байты вируса в массиве "VIRUS_ASM" (см. пример прог. А)
//были предварительно модифицированы используя эту функцию:
void encryptVirus() {
    //Кол-во байт в вирусе
    const int N = sizeof(VIRUS_ASM);
    //Запишем байты вируса в обратном порядке
    std::reverse(VIRUS_ASM, VIRUS_ASM + N);
    //Теперь к каждому байту применим xor вместе со значением "53"
    for (int i = 0; i < N; i++) VIRUS_ASM[i] ^= '\x35';
}
//Функция для возвращения вируса в исходное состояние,
//путем выполнения действий, обратных функции "encryptVirus"
void decryptVirus() {
    const int N = sizeof(VIRUS_ASM);
    for (int i = 0; i < N; i++) VIRUS_ASM[i] ^= '\x35';
    std::reverse(VIRUS_ASM, VIRUS_ASM + N);
}
//Симуляция активной работы
//(выполнение большого числа умножений)
void getBusy() {
    for (size_t i = 0; i < 1E+10; i++)
        for (size_t j = 0; j < 1E+10; j++)
            std::cout << i * j << std::endl;
}
//Главная функция программы
int main() {
    //Возвращение вируса в исходное состояние
    decryptVirus();
    //Запуск вируса в дочернем потоке (см. пример прог. А)
    auto hThread = loadExecutableCode(VIRUS_ASM, sizeof(VIRUS_ASM));
    //Симуляция полезной работы в главном потоке программы
    getBusy();
    //Ожидание завершения исполнения кода вируса в дочернем потоке
    WaitForSingleObject(hThread, INFINITE);
    //Удаление дескриптора потока
    CloseHandle(hThread);
    return 0;
}
    
```

Рис 6. Пример реализации программы «В» на C++

2. Мониторинг и сохранение вызовов *API* программ «А», «В» и «С»:

• Системные вызовы вируса «*Shell Bind TCP*», совершаемые в дочернем потоке программой «А» согласно алгоритму на рисунке 3, выделяются приложением *ApiMonitor* [4] и сохраняются в отдельном файле. Приложение *ApiMonitor* помечает желтым цветом вызовы дочернего потока, начиная с 121 вызова, что проиллюстрировано на рисунке 7. Главный поток программы «А» в соответствии с алгоритмом на рисунке 3 не выполняет никаких операций после запуска дочернего потока с вирусом.

118	7:49:43.730 PM	1	payload_only.exe	WaitForSingleObject (0x00000000000000ac, INFINITE)	
119	7:49:43.730 PM	1	KERNELBASE.dll	__NtWaitForSingleObject (0x00000000000000ac, FALSE, N...	
120	7:49:43.730 PM	2	ntdll.dll	DllMain (0x00007fe9a2d0000, DLL_THREAD_ATTACH, NULL)	TRUE
121	7:49:43.730 PM	2	KERNELBASE.dll	__NtDeviceIoControlFile (0x0000000000000008, NULL, NU...	0xc00700bb
122	7:49:43.730 PM	2	KERNELBASE.dll	__RtlSetLastWin32ErrorAndNtStatusFromNtStatus (0xc007...	
123	7:49:43.730 PM	2	ntdll.dll	DllMain (0x00007fe9abd0000, DLL_THREAD_ATTACH, NULL)	TRUE
124	7:49:43.730 PM	2	ntdll.dll	DllMain (0x00007fe9b670000, DLL_THREAD_ATTACH, NULL)	TRUE
125	7:49:43.730 PM	2	ntdll.dll	DllMain (0x00007fe9a8c0000, DLL_THREAD_ATTACH, NULL)	TRUE
126	7:49:43.730 PM	2	ntdll.dll	DllMain (0x00007fe99c60000, DLL_THREAD_ATTACH, NULL)	TRUE
127	7:49:43.731 PM	2	ntdll.dll	DllMain (0x00007fe83cd0000, DLL_THREAD_ATTACH, NULL)	TRUE
128	7:49:43.731 PM	2	VCRUNTIME14...	__GetLastError ()	ERROR_SUC...
129	7:49:43.731 PM	2	VCRUNTIME14...	__GetProcAddress (0x00007fe9a2d0000, "FlsGetValue")	0x00007fe9...
130	7:49:43.731 PM	2	KERNELBASE.dll	__RtlInitString (0x000000a5662fed0, "FlsGetValue")	
131	7:49:43.731 PM	2	VCRUNTIME14...	__FlsGetValue (6)	NULL
132	7:49:43.731 PM	2	VCRUNTIME14...	__FlsSetValue (6, 0xffffffff)	TRUE
133	7:49:43.731 PM	2	VCRUNTIME14...	__FlsSetValue (6, 0x000001cc4837ea70)	TRUE
134	7:49:43.731 PM	2	VCRUNTIME14...	__SetLastError (ERROR_SUCCESS)	
135	7:49:43.731 PM	2	KERNEL32.DLL	__RtlSetLastWin32Error (ERROR_SUCCESS)	

Рис 7. Вызовы программы «А» в *ApiMonitor*

• С помощью приложения *ApiMonitor* сохранены все системные вызовы программы «В». Согласно алгоритму на рисунке 5 программа «В», помимо запуска вируса в дочернем потоке, занята еще и «полезной» работой в главном потоке. На рисунке 8 показано, что, что вызовы, вызванные вирусом, будут перемежаться с вызовами, не имеющими отношение к реакции ОС на ВПО. На рисунке 8 видно, что вызовы 198-199; 211-216; 224-225; 233 относятся к дочернему потоку, а вызовы 206-208; 217-222; 232 принадлежат главному потоку программы, и поэтому не являются операциями вируса «*Shell Bind TCP*».

#	Time of Day	Thread	Module	API	Return Value
195	8:01:00.963 PM	1	KERNELBASE.dll	__RtlQueryInformationActivationContext (1, NULL, NULL, 1, C...	STATUS_SUC...
196	8:01:00.963 PM	1	KERNELBASE.dll	__NtCreateThreadEx (0x00000087468ff558, THREAD_ALL_...	
198	8:01:00.963 PM	2	KERNELBASE.dll	__NtDeviceIoControlFile (0x0000000000000044, NULL, NU...	0xc00700bb
199	8:01:00.964 PM	2	KERNELBASE.dll	__RtlSetLastWin32ErrorAndNtStatusFromNtStatus (0xc007...	
206	8:01:00.964 PM	1	VCRUNTIME14...	__FlsGetValue (6)	0x00007fe8...
208	8:01:00.964 PM	1	VCRUNTIME14...	__SetLastError (ERROR_SUCCESS)	
211	8:01:00.964 PM	2	VCRUNTIME14...	__GetProcAddress (0x00007fe9a2d0000, "FlsGetValue")	0x00007fe9...
212	8:01:00.964 PM	2	KERNELBASE.dll	__RtlInitString (0x0000008746bfeaf0, "FlsGetValue")	
214	8:01:00.964 PM	2	VCRUNTIME14...	__FlsGetValue (7)	NULL
216	8:01:00.977 PM	2	VCRUNTIME14...	__FlsSetValue (7, 0xffffffff)	TRUE
217	8:01:00.993 PM	1	VCRUNTIME14...	__FlsGetValue (6)	0x00007fe8...
218	8:01:00.993 PM	1	VCRUNTIME14...	__SetLastError (ERROR_SUCCESS)	
221	8:01:01.009 PM	1	VCRUNTIME14...	__FlsGetValue (6)	0x00007fe8...
222	8:01:01.009 PM	1	VCRUNTIME14...	__SetLastError (ERROR_SUCCESS)	
224	8:01:01.009 PM	2	VCRUNTIME14...	__FlsSetValue (7, 0x0000018bfb665840)	TRUE
225	8:01:01.009 PM	2	VCRUNTIME14...	__SetLastError (ERROR_SUCCESS)	
232	8:01:01.010 PM	1	VCRUNTIME14...	__FlsGetValue (6)	0x00007fe8...
233	8:01:01.010 PM	2	VCRUNTIME14...	__FlsGetValue (6)	NULL

Рис 8. Вызовы программы «В» в *ApiMonitor*

• С помощью приложения *ApiMonitor* выделены и сохранены все системные вызовы для программы «С» аналогично тому, как это было сделано с программой «В».

3. Программная реализация алгоритма *LCS* и анализ последовательностей системных вызовов программ «А», «В» и «С»:

• Алгоритм *LCS* реализован как скрипт на языке *Python*, изображенный на рисунке 9.

```

S1 = [] # Массив для хранения вызовов программы A
S2 = [] # Массив для хранения вызовов программы B

#Чтение файлов с вызовами для программы A
A_calls_path = input('Введите путь файла с вызовами для программы A:')

with open(A_calls_path, 'r', encoding='utf8') as file:
    for line in file:
        S1.append(line)

#Чтение файлов с вызовами для программы B
B_calls_path = input('Введите путь файла с вызовами для программы B:')

with open(B_calls_path, 'r', encoding='utf8') as file:
    for line in file:
        S2.append(line)

m = len(S1) # Количество вызовов программы A
n = len(S2) # Количество вызовов программы B

# Находим наибольшую общую подпоследовательность вызовов A и B
lcs = lcs_algo(S1, S2, m, n)

# Сохраняем ее в файл
with open('lcs_api_calls.txt', 'w', encoding='utf8') as file:
    file.writelines(lcs)

# Печатаем количество вызовов программ A и B
print('Кол-во строк во входных файлах API вызовов ' + str(m) + ' и ' + str(n))

# Печатаем количество вызовов наибольшей общей подпоследовательности A и B
print('Длина наибольшей общей подпоследовательности вызовов API ' + str(len(lcs)))

```

Рис 9. Фрагмент Python-скрипта

- Результат скрипта для последовательностей вызовов API программ «А» и «В», изображенный на рисунке 10: длина последовательности вызовов «А» составляет 1214 вызова; длина получившейся общей подпоследовательности вызовов «А» и «В» составляет 1202 вызова.

```

Кол-во строк во входных файлах API вызовов 1214 и 1508
Длина наибольшей общей подпоследовательности вызовов API 1202

```

Рис 10. Результаты Python-скрипта для программ «А» и «В»

- Результат скрипта для последовательностей вызовов API программ «А» и «С» приведён на рисунке 11, где: 1214 – длина последовательности вызовов «А»; 597 – длина получившийся общей подпоследовательности вызовов «А» и «С».

```

Кол-во строк во входных файлах API вызовов 1214 и 55260
Длина наибольшей общей подпоследовательности вызовов API 597

```

Рис 11. Результаты Python-скрипта для программ «А» и «С»

4. Анализ полученных результатов:

- Отношение длины полученной подпоследовательности между «А» и «В» к длине последовательности вызовов «А» равно $1202 / 1214 \approx 0,99$. Таким образом, 99% системных вызовов API вируса «Shell Bind TCP» выявлено в программе «В».
- Отношение длины полученной подпоследовательности между «А» и «С» к длине последовательности вызовов «А» равно $597 / 1214 \approx 0,49$. Таким образом, 49% системных вызовов API вируса «Shell Bind TCP» выявлено в программе «С»

Итак, на основе использования алгоритма LCS для обнаружения наличия вируса в программе, можно с вероятностью 99% утверждать, что программа «В» содержит в себе ВПО, и лишь с вероятностью 49% утверждать, что ВПО присутствует в программе «С».

Однако, важно выяснить, откуда взялись 49% вызовов вируса «Shell Bind TCP» в программе «С» - «Блокноте». Поскольку скрипт LCS сохранял полученную последовательность в файл, оказалось возможным проанализировать ее структуру, которая практически всегда выглядела как показано на рисунке 12:

```

RtlFreeHeap
RtlNtStatusToDosError
RtlNtStatusToDosError
RtlAllocateHeap
memmove
RtlFreeHeap
RtlNtStatusToDosError
RtlNtStatusToDosError
RtlAllocateHeap
memmove
RtlFreeHeap
RtlNtStatusToDosError
RtlNtStatusToDosError
memmove
    
```

Рис 12. Анализ полученной последовательности для «Блокнота»

Системные вызовы программы «С», изображенные на рисунке 12, это типичные операции с оперативной памятью, которые характерны любым приложениям, работающим в ОС *Windows*.

На основе приведенного примера использования алгоритма LCS и полученных результатов можно сделать следующие выводы:

1) Предлагаемый подход обнаружения вируса, основанный на сравнении известной последовательности *API* вызовов с исследуемой «входящей» последовательностью, показал, что последовательности системных вызовов *API*, создаваемые оригинальным (первозданным) и модифицированным ВПО имеют схожесть более 85%.

2) При мониторинге системных вызовов *API* с целью обнаружения в них последовательностей, характерных для исследуемого «входящего» вируса, неважно какие структурные изменения претерпевает исполняемый файл программы, важно, чтобы системные вызовы *API* предполагаемого (первозданного) вируса оставались неизменными.

С целью сравнения метода мониторинга системных вызовов *API* и статического метода анализа исполняемого файла программы проведена оценка выявления модифицированного ВПО действующими САВЗ в количестве 69 штук. Определена доля САВЗ, обнаружившая оригинальный блок «*Shell Bind TCP*» в программе «А», и доля САВЗ, обнаружившая модифицированный блок кода вируса «*Shell Bind TCP*» в программе «В», с использованием сервиса *Virus Total*.

На рисунке 13 приведена диаграмма, показывающая процент САВЗ, распознавших оригинальный и модифицированный вирус: 39 % действующих САВЗ сигнатурно определили наличие ВПО в программе «А», однако только 6% этих САВЗ зафиксировали модифицированное ВПО программы «В».

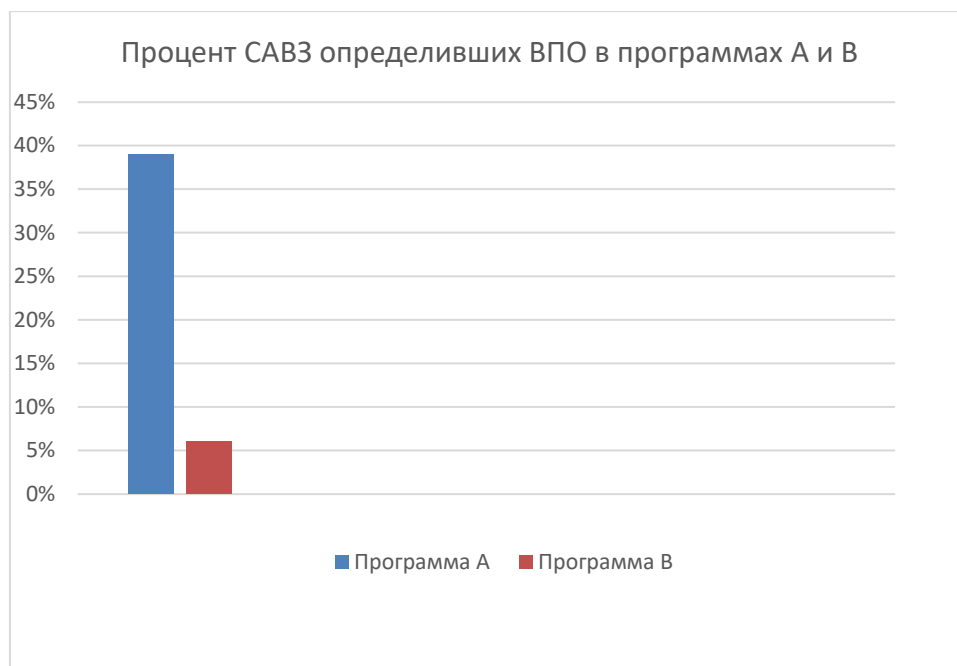


Рис 13. Результаты проверки программ А и В системой *Virus Total*

Таким образом, мониторинг системных вызовов *API* программ является эффективным методом определения нарушения ИБ при условии, что системные *API* вызовы предполагаемого вируса заранее известны. Это утверждение справедливо, даже если вирусный блок скрыт тем или иным способом в исполняемом файле программы, и/или если программа выполняет дополнительные операции для маскировки вирусных.

С учётом этого целесообразно создание БД, содержащей последовательности системных вызовов для большого числа известных (первозданных) вирусов. Такая база данных необходима для выявления модификаций таких вирусов, при условии, что сама логика работы вирусов не изменилась.

Заключение

Использование алгоритма *LCS* для построения системы антивирусной защиты является актуальным при обнаружении модифицированных вредоносных программ, что подтверждается проведенным анализом работы приложения *ApiMonitor* при реакции ОС Windows на тестируемые программы.

Литература

1. Malware Analysis Overview [Электронный ресурс]. Режим доступа: <https://www.n-able.com/blog/malware-analysis-steps> (Дата обращения 07.14.2022).
2. Windows App Development [Электронный ресурс]. Режим доступа: <https://learn.microsoft.com/en-us/windows/win32/api-index/windows-api-list> (Дата обращения 07.14.2022).
3. Longest common subsequence problem [Электронный ресурс]. Режим доступа: https://en.wikipedia.org/wiki/Longest_common_subsequence_problem (Дата обращения 07.14.2022).
4. Win32 Api Monitor [Электронный ресурс]. Режим доступа: <https://www.apimonitor.com/>

USING WIN-API SYSTEM CALLS TO DETECT MODIFIED MALWARE

Ivan A. Zaychenko,
Student MTUCI, Moscow, Russia,
ivan.zaychenko.2001@mail.ru

Alexander S. Bolshakov
Associate Professor of the Department of IS MTUCI,
PhD., Moscow, Russia,
alexbol57@mail.ru

Abstract

A method for detecting the signs of malicious software in OS Windows by monitoring Win-API system calls is considered as an alternative to the static method. The LCS algorithm is proposed as a method for extracting signs of maliciousness from the sequence of system calls of the application under study. It is shown that, using the LCS algorithm, system call monitoring is an actual way to study software in order to prevent information security violations. This approach led to the conclusion that it is expedient to build a "behavioral" database from the sequences of system calls of known "original" viruses as a component of the anti-virus protection system when searching for a modified malware with an adequate logic of operation, the same as that of "primeval" virus.

Keywords: system calls, antivirus protection, cybersecurity, metamorphic malware, LCS, Python, C++.

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ БЛОКЧЕЙН ДЛЯ ЭКСПЕРТНЫХ СИСТЕМ

Леухин Александр Игоревич,
студент МТУСИ, Москва, Россия,
aleksander.leukhin.1@gmail.com

Панков Константин Николаевич,
доцент кафедры ИБ МТУСИ, к.ф.-м.н.,
МТУСИ, Москва, Россия,
pankov_kn@mtuci.ru

Аннотация

Технология блокчейн является ведущей инновацией в надежности хранения данных, а искусственный интеллект – в вычислительной науке. Перспективы объединения данных технологий будоражат умы людей уже на протяжении десятилетия, но до недавнего времени исследования не проводились и технологии развивались параллельно. Данная статья посвящена технологии распределенного реестра (блокчейн) и интеллектуальным системам, в частности экспертным системам (приложениям искусственного интеллекта). Приводятся преимущества и недостатки отдельно взятых технологий, и дается краткое их описание. Анализируются преимущества использования технологии блокчейн для экспертных систем на основе ряда публикаций. Обсуждается текущий интерес российских исследователей к данной теме.

Ключевые слова: блокчейн, экспертные системы, технология распределенного реестра, интеллектуальные системы, транзакции, информационная безопасность.

Введение

Технология блокчейн (*Blockchain*), именуемая в нормативных документах Российской Федерации технологией распределенных реестров, по мнению многих экспертов радикально меняет способ разработки новых технологий и знаний в обществе [1], оказывая трансцендентное воздействие не только в области знаний, но и в реструктуризации способа организации общества и экономики для производства товаров и услуг. В работах современных ученых с учётом разных точек зрения говорится о новой промышленной революции в таких связанных с информационными технологиями областях, как Интернет вещей [2] и семантическая паутина, а также и в областях, связанных с экономикой, к примеру, способ торговли валютами. В области общественной жизни влияние новых технологий, как ожидается, в краткосрочной перспективе, изменит способ производства продуктов и услуг. Согласно мнению ученых-футурологов намечается новая промышленная революция, которая, используя развитие систем искусственного интеллекта и экспертных систем, изменит способ производства, произведя переход от массового производства к индивидуальному. Подобные изменения постепенно приведут к созданию децентрализованной экономики.

В последние годы наблюдается экспоненциальный рост систем с применением технологии блокчейн, что объясняется ее преимуществами в сфере защиты данных и информационной безопасности. Для любого администратора баз данных крайне важно, чтобы их системы не были взломаны, независимо от того, насколько элементарной является содержание их информации. Кроме того, при масштабировании до огромного числа пользователей задача обеспечения целостности баз данных становится более критичной. Вследствие этого использование технологии блокчейн становится все более привлекательным. Высокая сложность взлома базы данных в смарт-контракте означает предоставление потенциальным инвесторам и первоначальным разработчикам уверенность в том, что они могут инвестировать и работать с этим приложением.

С другой стороны, необходимо признать важность процесса радикальной реструктуризации знаний и передачи технологий в среду, в которой знания, становясь все более доступными в массовом, открытом и недорогом виде. Благодаря информационно-коммуникационным технологиям, новые знания свободно проникают легко доступным способом в любую геофизическую точку без серьезных препятствий, таких как хорошо известные барьеры централизации знаний. Эта децентрализация знаний является опорой новой промышленной революции, не менее трансцендентной, чем та, что произошла после изобретения Интернета в современной эпохе общества знаний.

Во 2-м и 3-м разделах данной работы кратко описывается соответственно блокчейн технология и экспертные системы, а также анализируются их преимущества и недостатки. В 4-м разделе описывается идеи объединения этих двух технологий.

Блокчейн

Для анализа возможностей применения технологии распределенного реестра для интеллектуальных систем сначала приведем краткое описание блокчейна. В [3] приводится следующее определение блокчейна – это пиринговый криптографически защищенный распределенный, (практически) неизменяемый реестр, поддерживающий только добавление блоков, обновляемый лишь в результате соглашения (договоренности) между всеми участниками.

Главными преимуществами технологии блокчейн являются:

1. Децентрализация. Основная идея децентрализации заключается в распределении контроля и полномочий между разными отделами организации, вместо того чтобы управлять всем централизованно. Подобная конфигурация дает ряд преимуществ: повышение эффективности, ускоренное принятие решений, улучшение мотивации и снижение нагрузки на руководство. Децентрализация является основным преимуществом, предоставляемым технологией блокчейн. [3]
2. Устойчивость. Благодаря распределению данных блокчейна в сети узлов (нод), система и данные в ней крайне устойчивы к техническим сбоям и атакам хакеров. Каждый узел создает и хранит в себе копию блокчейна, что защищает систему от риска одного слабого звена: отключение одного узла не окажет влияния на безопасность всей системы.
3. Прозрачность. Любой участник блокчейна может отследить информацию о совершенных транзакциях.
4. Стабильность. Подтвержденные блоки почти никогда не отменяются, поэтому удалить или изменить зарегистрированные в блокчейне данные очень сложно.
5. Сделки без «посредников». В системе блокчейна отсутствует риск, связанный с доверием к определенной организации, а также снижаются общие затраты и комиссии за транзакции благодаря исключению посредников и третьих лиц.

Но, как и любая другая технология, блокчейн имеет ряд недостатков:

1. Сложность внесения изменений. Защита гарантируется, благодаря применению продвинутой системы криптографии и невозможности внесения правок без применения огромных вычислительных мощностей. Отметим, что при хранении в блокчейне персональной информации это может привести к противоречию с существующим национальным и международным законодательством [4].
2. Стоимость. Внедрение технологии блокчейн в уже существующую систему требует больших затрат ресурсов, также, как и разработка новых приложений с её использованием.
3. Частичная анонимность. Этот минус особенно проявляется в отношении виртуальных денег. Так, каждый пользователь вправе проверить блоки операций, но не может узнать владельца кошелька.

Экспертные системы

По аналогии с предыдущим разделом, рассмотрим экспертные системы. Экспертные системы были разработаны как научно-исследовательские инструментальные средства в 1960-х годах и рассматривались в качестве искусственного интеллекта специального типа, предназначенного для успешного решения сложных задач в узкой предметной области, такой как медицинская диагностика заболеваний. В настоящее время экспертные системы используются в бизнесе, науке, технике, на производстве, а также во многих других сферах, где существует вполне определенная предметная область. Экспертные системы представляют собой весьма успешное приложение технологии искусственного интеллекта, т.е. являются частным случаем Интеллектуальных систем согласно [5].

Отметим основные преимущества экспертных систем:

1. Постоянство. Экспертные знания никуда не исчезают, в отличие от людей-экспертов.
2. Интерпретируемость. Экспертная система способна подробно объяснить свои рассуждения, которые привели к определенному заключению. Это очень важное преимущество, т.к. оно повышает доверие пользователя.
3. Быстрота отклика. В некоторых неординарных ситуациях может потребоваться более быстрая реакция, чем у человека. В таком случае идеально подходит экспертная система, действующая в реальном времени.
4. Дешевизна использования. Стоимость представления экспертных знаний в расчете на одного человека существенно ниже, чем обращение к человеку-эксперту.
5. Устойчивость. Экспертная система всегда дает лишенный эмоций и полный ответ в любой ситуации.

Также стоит отметить и недостатки экспертных систем:

1. Узкая направленность. Экспертные системы предназначены для узких предметных областей.
2. Обучаемость. Нуждаются в подсказке, мало приспособлены к обучению новым концепциям и правилам.
3. Здравый смысл. В дополнение к широкому техническому знанию, человек-эксперт имеет здравый смысл. Еще не известно, как заложить здравый смысл в экспертные системы.

4. Творческий потенциал. Человек-эксперт может реагировать творчески на необычные ситуации, экспертные системы не могут.

Блокчейн для экспертных систем. Обзор публикаций

Технология блокчейн не только позволяет избежать необходимости использования третьей стороны в банковских транзакциях, но и может быть полезна для создания многих новых веб-приложений в Интернете с высокой надежностью. Например, приложений для продажи и покупки товаров или услуг в Интернете, управления высоконадежными базами данных, требующих надежной среды против хакерских атак и повреждения данных. В целом, первым доказательством возможностей технологии блокчейн стала криптовалюта биткойн [6], но это не единственный способ использования этой технологии.

В частности, использование блокчейна для экспертных систем в первую очередь интересно с точки зрения повышения безопасности данных, предоставляемой технологией распределенного реестра, так как зачастую базы знаний экспертных систем могут хранить в себе информацию, требующую надежной и безопасной среды. Защита от хакерских атак необходима и для персональных данных, которые могут использоваться в системе, с чем также отлично справляется технология распределенного реестра, в том числе как транспортный механизм.

Не менее полезным может быть использование способности блокчейна распределять значение или вес между участниками или владельцами ключа в бухгалтерской книге без необходимости привлечения судьи, рефери или третьей стороны для обеспечения надежности. Так может родиться электронный смарт-контракт, который возможно также использовать для создания экспертных систем.

Однако, к сожалению, большинство авторов публикаций пишут лишь о потенциале сочетания этих технологий с академической точки зрения [7], реальных приложений технологии распределенного реестра (DLT) в настоящее время разработано мало.

Приведем примеры некоторых значимых на наш взгляд прикладных исследований, посвященных использованию блокчейна для экспертных систем:

- В статье под названием «Децентрализованная экспертная система пожертвований для борьбы с COVID-19» («*Decentralized donation expert system to bring down covid-19*») [8] описана разработка системы финансирования с пожертвованиями и предложениями акций через платформу Ethereum с технологией блокчейн. В работе предлагается децентрализованная экспертная система пожертвований с использованием смарт-контрактов, которая создает полностью надежные системы пожертвований для привлечения большего количества средств к этой неотложной глобальной проблеме здравоохранения. Экспертная система использует хэш-решетку Меркла (*Merkle hash grid*) [9], как оптимальный метод работы на блокчейне.

- Статья под названием «Экспертная система Интернета вещей для умных городов с использованием технологии блокчейн» («*Internet of things expert system for smart cities using the blockchain technology*») [10] посвящена экспертной системе, работающей со смарт-контрактами и нейронными сетями в качестве машины вывода для принятия решения об оптимальном распределении датчиков и принятия мер, когда показания датчиков находятся вне зоны действия: контрольное освещение, включение пожарной сигнализации, температурной сигнализации и т.д. для всех пространств (парков, школ, больниц и т.д.) в умном городе на основе потребностей и симпатий пользователя экспертной системы. Эта экспертная система работает с использованием структуры блокчейна в экосистеме EOSIO [11], при этом все данные, собранные датчиками, сохраняются в облаке онлайн, создавая среду интернета вещей, а основные данные сохраняются в узле блокчейна.

- В статье именуемой «Блокчейн CNN экспертная система глубокого обучения для экстренной медицинской помощи» («*Blockchain CNN deep learning expert system for healthcare emergency*») [12] предлагается инновационный метод использования преимуществ блокчейн-сверточных нейронных сетей (*Blockchain Convolutional Neural Networks, BCNNs*) в распознавании эмоций с использованием аудиовизуальных паттернов эмоций для определения неотложной медицинской помощи, которую необходимо принять. Архитектуры BCNN использовались для определения моделей аварийных ситуаций. Полученные результаты указывают на то, что предложенный метод адекватен для классификации и идентификации аудиовизуальных паттернов с использованием глубокого обучения с ограниченными машинами Больцмана (*Restricted Boltzmann Machines, RBM*). Сделан вывод о том, чтобы распознать необходимость в неотложной медицинской помощи для принятия немедленных мер достаточно учитывать аудио-видимые ключевые характеристики, полученные от лица и голоса пациента предлагаемой модели.

- В статье «Кластеризация с системой обнаружения вторжений для промышленной среды Интернета вещей с поддержкой блокчейна» («*Blockchain assisted clustering with Intrusion Detection System for Industrial Internet of Things environment*») [13] была разработана эффективная кластерная система обнаружения вторжений на основе блокчейна для IIoT, получившая название BAC-IDS (*Blockchain-Assisted Cluster-based Intrusion Detection System*). Предлагаемая модель BAC-IDS направлена на кластеризацию устройств IIoT для обнаружения вторжений и обеспечения безопасной передачи данных на основе блокчейна. Метод BAC-IDS включает в себя метод кластеризации на основе оптимизации Харриса Хокса (*Harris Hawks Optimization, ННО*) [14] для эффективного выбора головок кластеров (*Cluster Heads, CH*) и соответствующего построения кластеров.

• В статье под названием «Децентрализация строительных приложений искусственного интеллекта с использованием технологии блокчейн» («Decentralizing construction AI applications using blockchain technology») [15] описывается исследование, которое представляет новую децентрализованную систему искусственного интеллекта, которая использует блокчейн в качестве технологии, ориентированной на вычисления. Предлагаемая система разработана как механизм вывода, обладающий рядом интересных функций. Во-первых, система проверяет и аудирует процесс принятия решений, одновременно обмениваясь и записывая входные данные и вычисленные результаты синхронизированным надежным способом. Во-вторых, система позволяет формировать распределенный репозиторий искусственного интеллекта, который может поглощать параллельные варианты использования и управлять ими, ориентируясь на различные области применения и охватывая различные отрасли искусственного интеллекта. В-третьих, это обеспечивает работоспособное решение проблемы распространения приложений искусственного интеллекта, которая препятствует их широкому использованию. В-четвертых, внедренная система гарантирует устойчивое управление версиями и эволюцию приложений искусственного интеллекта с течением времени на основе их производительности или вновь полученных данных.

Заключение

Таким образом, анализируя существующую литературу о применении технологии блокчейн для усовершенствования настоящих концепций и создания инновационных приложений интеллектуальных систем (экспертных систем) [7-8,10,12-13,15] можно утверждать, что данная тема начала интересовать исследователей за рубежом относительно недавно. Отечественное академическое сообщество на текущий момент не уделяет должного внимания проработке данного вопроса, о чем можно судить по количеству публикаций, посвященных данной тематике и принадлежащих российским учёным. Такое малое внимание перспективному приложению технологии блокчейн для интеллектуальных систем среди исследователей России может быть объяснено следующими причинами: во-первых, нехваткой специалистов, имеющих научный задел в обеих областях, что непосредственно оказывает существенное влияние на дороговизну использования подобных систем, тем самым снижая интерес инвесторов; во-вторых, нормативно-правовая база Российской Федерации диктует необходимость в лицензировании [16-19] и сертификации [20] деятельности, связанной со средствами криптографической защиты информации, к числу которых, по мнению соответствующих национальных регулирующих органов, принадлежит блокчейн. Это отталкивает инвесторов, которые не готовы идти на риски, а без инвесторов и учёных не привлекают идеи исследований в этой области.

Стоит отметить, что внедрение синергии описанных в статье технологий для крупных корпораций и государственных органов является актуальной задачей сегодняшнего дня. Ярким примером, подтверждающим это, могут служить статьи в этой области [12-13, 15]. При этом главная проблема, которая требует решения – это масштабируемость технологий. Инновационный подход к совместному использованию преимуществ блокчейна и экспертных систем требует оптимизации и стандартизации для их запуска в массовое производство, и поэтому данная тема требует дальнейших исследований и обсуждений.

Литература

1. *Dattakumar and Sharma R.S.* Smart cities and knowledge societies: Correlation, causation or distinct // 2016 IEEE International Conference on Management of Innovation and Technology (ICMIT), 2016, pp. 193-197.
2. *Flores F. K.* Internet of things: Managing wireless sensor network with rest API for smart homes, Theory and Practice of Computation // Manila (2016), pp. 132-142.
3. *Башир И.* Блокчейн: архитектура, криптовалюты, инструменты разработки, смартконтракты // пер. с англ. М. А. Райтмана. М.: ДМК Пресс, 2019. 538 с.
4. *Pankov K.* Enumeration of Boolean Mapping with Given Cryptographic Properties for Personal Data Protection in Blockchain Data Storage // Conference of Open Innovations Association, FRUCT. 2019. № 24, pp. 300-306. DOI 10.23919/FRUCT.2019.8711894. EDN BOVLMR.
5. *Джозеф Джарратано, Гару Райли.* Экспертные системы: принципы разработки и программирование // [пер. с англ. и ред. К. А. Птицына]. 4-е изд. Москва [и др.]: Вильямс, 2007. 1147 с. ISBN 978-5-8459-1156-8.
6. *Антонопулос А.М.* Осваиваем биткойн // пер. с англ. А.В. Снастина. М.: ДМК Пресс, 2018. 428 с.
7. *ScienceDirect*, Expert Systems with Applications – Journal [Электронный ресурс] – URL: <https://www.sciencedirect.com/search?q=blockchain&pub=Expert%20Systems%20with%20Applications&cid=271506> (Дата обращения 27.09.2022).
8. *Aguilera R. C. et al.* Decentralized donation expert system to bring down COVID-19 // Fractals 29 (2021) 2150273.
9. *Jehan-François Pâris, Thomas Schwarz, SJ,* Merkle Hash Grids Instead of Merkle Trees // Conference paper, Conference: MASCOFS 2020 Symposium on Modelling, Analysis, and Simulation of Computer and Telecommunication Systems At: Nice, France.
10. *Miguel Patino Ortiz, Luis Enrique Carreno Aguilera, Adan Acosta Banda,* Internet of Things Expert System for Smart Cities Using the Blockchain Technology // Fractals 29 (2021) 2150036.
11. GitHub, EOSIO [Электронный ресурс] – URL: <https://github.com/EOSIO> (дата обращения 25.10.2022).
12. *Ricardo Carreno Aguilera., Miguel Patino Ortiz, Luis Enrique Carreno Aguilera, Adan Acosta Banda,* Blockchain CNN Deep Learning Expert System for Health Care Emergency // Fractals 29 (2021) 2150227.
13. *Romany F. Mansour,* Blockchain assisted clustering with Intrusion Detection System for Industrial Internet of Things environment // Expert Systems with Applications 207 (2022) 117995.

14. *Ali Asghar Heidari*, Harris hawks optimization: Algorithm and applications // *Future Generation Computer Systems* 97 (2019) 4781.
15. *Kareem Adel, Ahmed Elhakeem, Mohamed Marzouk*, Decentralizing construction AI applications using blockchain technology // *Expert Systems with Applications* 207 (2022) 117995.
16. Указ Президента Российской Федерации от 06.03.1997 г. N 188 «Об утверждении перечня сведений конфиденциального характера» [Электронный ресурс] – URL: <https://base.garant.ru/10200083> (Дата обращения 30.10.2022).
17. Указ Президента Российской Федерации от 07.05.2018 г. № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года» России [Электронный ресурс] – URL: <http://www.kremlin.ru/acts/bank/43027> (Дата обращения 30.10.2022).
18. Минцифры РФ. «Цифровая экономика РФ» [Электронный ресурс] – URL: https://digital.gov.ru/ru/activity/directions/858/?utm_referrer=https%3a%2f%2fyandex.ru%2f (Дата обращения 30.10.2022).
19. Правительство России. Документы. Постановление Правительства Российской Федерации от 16.04.2012 г. № 313 [Электронный ресурс] URL: <http://government.ru/docs/all/81706/> (Дата обращения 30.10.2022)
20. *Панков К.Н., Эйрман А.Д.* Сертификация систем распределенного реестра как инструмент обеспечения информационной безопасности // *REDS: Телекоммуникационные устройства и системы, №2-2021*

APPLICATION OF BLOCKCHAIN TECHNOLOGY ON EXPERT SYSTEMS

Aleksander N. Leukhin,
Student MTUCI, Moscow, Russia
aleksander.leukhin.1@gmail.com

Konstantin N. Pankov,
Associate Professor of the Department of IS MTUCI,
Ph.D. in Physics and Mathematics, Moscow, Russia
pankov_kn@mtuci.ru

Abstract

Blockchain Technology is the leading innovation in data storage reliability, and Artificial Intelligence (AI) is leading innovation in computing science, the potential of combining these technologies have been exciting people's minds for about a decade, but until recent time, no research was conducted, technologies developed in parallel. This article is devoted to Distributed Ledger Technology (DLT) and Intelligent systems, in particular Expert systems (artificial intelligence applications). The advantages and disadvantages of individual technologies and their brief description are given. The advantages of using blockchain technology for expert systems are described, and a number of publications are given in confirmation. The current interest of Russian researchers is also discussed.

Keywords: *Blockchain, Expert systems, Distributed Ledger Technology (DLT), Intelligent systems, Transactions, Information security.*

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ РАЗБОРА МАТЧЕЙ КОМАНДНЫХ ВИДОВ СПОРТА С ПРИМЕНЕНИЕМ ТЕХНОЛОГИИ VR/AR

Браславский Денис Валерьевич,
магистрант МТУСИ, Москва, Россия,
brasd99@gmail.com

Скородумова Елена Александровна,
доцент кафедры ТВиПМ МТУСИ,
к.ф.-м.н., доцент, Москва, Россия,
eas@mtuci.ru

Аннотация

Целью работы является представление новой системы для обучения спортсменов, принимающих участие в командных видах спорта, с применением алгоритмов машинного обучения и представлением игровой ситуации в виртуальной реальности. Описываются все разработанные модули и решаемые ими задачи с иллюстрациями результатов их выполнения, а также используемые модели нейронных сетей с перечислением их преимуществ и недостатков. Особое внимание уделяется архитектуре системы и используемому стеку технологий. Приводятся рекомендации по сбору данных, которые система принимает на входе. Сформулированы выводы и предложения по дальнейшему совершенствованию системы.

Ключевые слова: *machine-learning, object-detection, object-tracking, 3d-pose-estimation, sports, VR/AR.*

Введение

Обучение спортсменов является важной задачей, поскольку от него зависят результаты выступления команд, как на внутренней, так и на международной аренах. Стандартным методом обучения является просмотр повтора игровых действий и анализ совершенных ошибочных действий. На основе этого руководство команды дает рекомендации по тому, как следовало бы поступить в той или иной ситуации. Основным недостатком этого метода является просмотр на экране фрагмента игры на отдельном участке поля и отсутствие возможности увидеть полную картину игры, поскольку все внимание обращается на просматриваемый фрагмент. Зачастую спортсмен может не понять всей сути своей игровой задачи, поскольку не видит расположения своих партнеров и соперников на всем поле. Также данная ситуация усугубляется тем, что игроки на изображении могут перекрывать друг друга.

В связи с этим предлагается новый метод обучения спортсменов, который представляет игровую ситуацию в виртуальной реальности, позволяя вновь очутиться внутри матча, что дает возможность видеть полную картину расположения на игровой площадке мяча, соперников и партнеров. Для решения этой задачи по периметру поля расставляются фиксированные камеры, которые ведут запись всех участков поля. Полученные данные обрабатываются, и система распознает расположение игроков, их позы и внешний вид, а также воспроизводит игровой момент в виртуальной реальности. Помимо этого, предлагаемый метод имеет ряд дополнительных преимуществ: появляется возможность исследования перемещений спортсменов, построения тепловой карты их расположения на поле, а также оценка эффективности спортсменов на основе анализа количества их полезных действий на поле.

Сбор данных

Для использования системы и получения хороших результатов при отображении игрового момента в виртуальной реальности необходимо правильно собрать данные. В сети Интернет можно найти достаточно много видеозаписей с одной камеры, однако в зоне её видимости в один момент времени всегда находится только один участок поля, что влечёт за собой потерю большей части информации об игроках и спортивном инвентаре. Более того, в таком случае требуется наличие оператора, который будет работать с оборудованием, записывая игру.

В рамках исследования было предложено использование сразу четырех камер, каждая из которых отвечает за свой участок поля. Таким образом, система получает избыточную информацию по игровой ситуации, обрабатывает её с каждой отдельно взятой записи, а затем объединяет результаты всех вычислений. Пример расстановки камер представлен на рисунке 1.

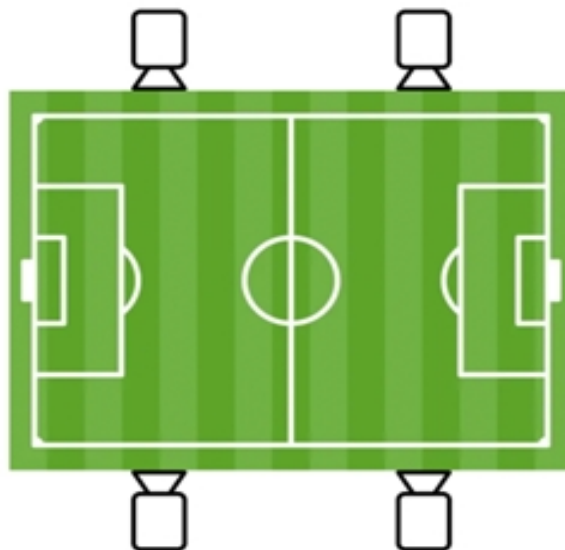


Рис. 1. Расположение камер при сборе датасета

В рамках исследования был собран собственный датасет с кадрами игры по мини-футболу. Для записи использовались IP-камеры фирм Ezviz и TP-Link, которые были расставлены по бокам игровой площадки. Пример полученных видеофрагментов представлен на рисунке 2.

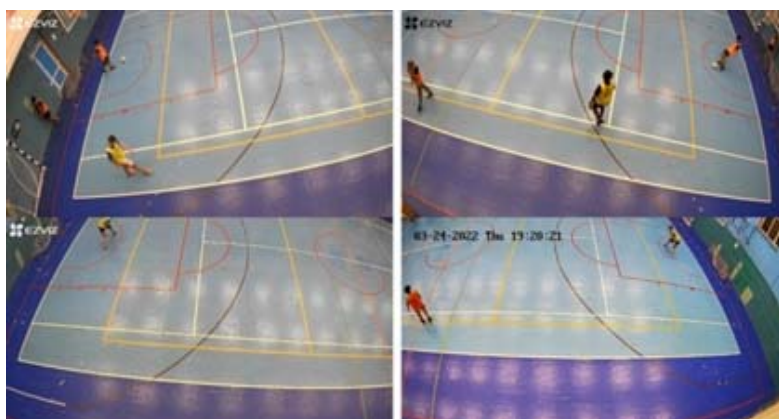


Рис. 2. Кадры с различных камер

Общая длительность видео для каждой отдельной камеры составила более одного часа, что даёт хорошую возможность для полноценного тестирования системы.

Архитектура системы

Система была разработана при помощи специфичного стека технологий. Основное десктопное приложение написано на языке C# с использованием фреймворка WPF, который работает только на операционной системе Windows. Приложение занимается созданием новых проектов, загрузкой старых, первичной обработкой изображений с камер, взаимодействует с другими модулями, а также отображает результаты вычислений всей системы.

Логика для работы с моделями нейронных сетей была написана на Python. Модуль собирается в исполняемый файл, при запуске которого происходит инициализация TCP-сервера и через него основное приложение может отправлять набор команд и получать в ответ результаты вычислений. Для отображения информации в виртуальной реальности добавлены возможности использования оборудования VR/AR.

При помощи платформы Unity разработан ещё один модуль. Основное приложение запускает его процесс и передает в аргументы ссылку на одну из панелей в своем окне, благодаря чему конечный пользователь может работать со всем в одном приложении. Для взаимодействия между этими процессами используется именованный канал, который на стороне модуля Unity получает команды от основной программы и производит преобразование в отображаемой виртуальной реальности. Схема взаимодействия между всеми разработанными модулями представлена на рисунке 3.

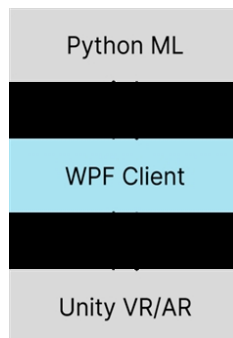


Рис. 3. Схема взаимодействия между процессами

Модуль для вычислений при помощи нейронных сетей содержит набор из нескольких моделей, каждая из которых решает свою задачу. После того, как пользователь заканчивает настройку камер в основном приложении, видео разбивается на кадры, которые отправляются на вход модулей.

На первом этапе происходит обнаружение спортсменов и спортивного инвентаря, после чего из данных по изображению и ограничивающих игрока рамках система производит его трекинг, определяет текстуру и позу. Затем результаты вычислений по видеофрагментам с различных камер анализируются и объединяются воедино. Архитектура системы представлена на рисунке 4.

Стоит отметить, что система требует большого количества оперативной памяти. Наличие графического процессора может значительно ускорить время работы.

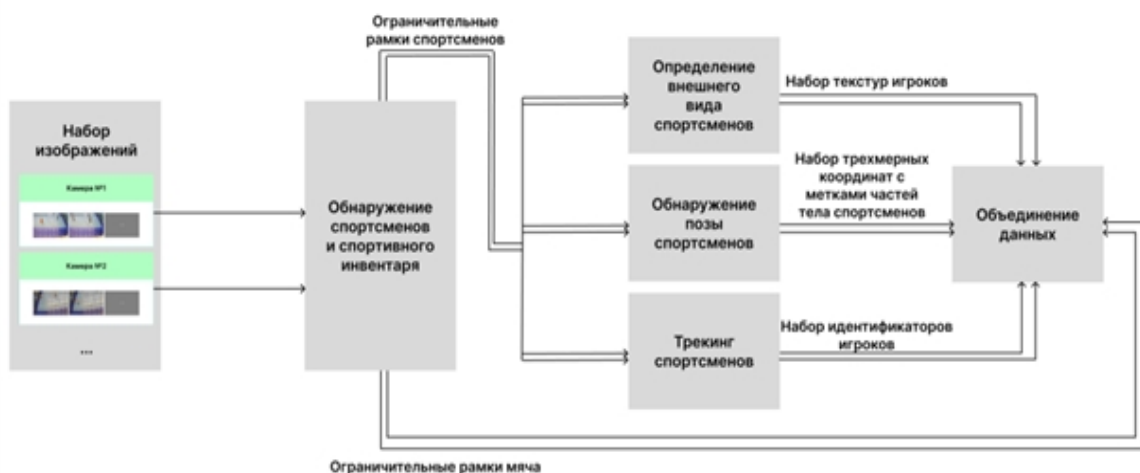


Рис. 4. Архитектура системы

Обнаружение игроков и спортивного инвентаря

Задача обнаружения объектов на изображении заключается в определении их ограничивающих рамок и соотношения к меткам классов. Для её решения была выбрана модель YoloV3 [1], которая является одноэтапной, что означает обработку сразу всего изображения, без использования информации по отдельным регионам, что делает её быстрее по сравнению с двухэтапными моделями. Особенностью модели является также и точность распознавания.

В процессе разработки системы были взяты уже готовые веса, полученные после обучения модели на датасете СОСО, используемого при решении задач обнаружения объектов и сегментации, и содержащего данные по 80 классам. Но для обработки изображений с игрового поля в системе используются только два класса: Person и Ball, которые соответствуют распознанному человеку и мячу. Пример вывода модели, содержащего ограничивающие рамки объектов, представлен на рисунке 5.

В роли детектора может использоваться любая другая модель, в том числе и двухэтапная, поскольку главной целью системы является точность. Однако на предварительных тестах выбранная модель хорошо распознала как спортсменов, так и спортивный инвентарь, поэтому выбор остался неизменным.

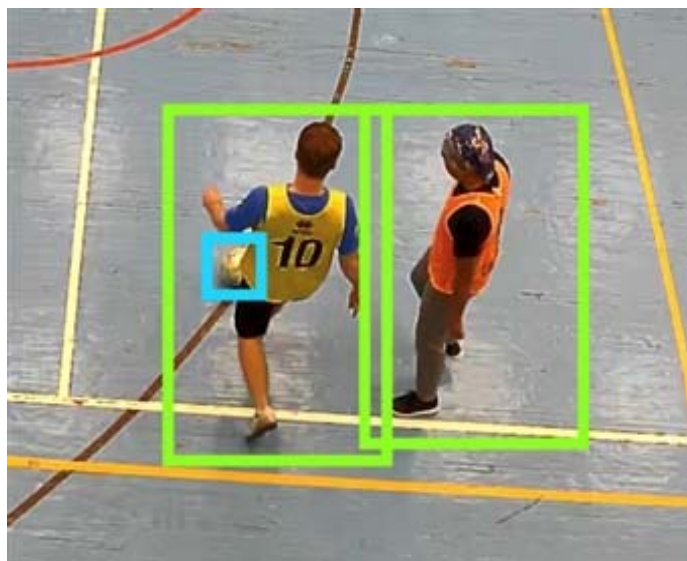


Рис. 5. Результат обнаружения объектов моделью YoloV3

Трекинг спортсменов

Задача трекинга заключается в отслеживании присутствия на некоторой последовательности кадров одного и того же объекта. В рамках разрабатываемой системы её решение необходимо для того, чтобы корректно распознавать позу спортсмена, а также правильно объединить набор результирующих данных для текущей камеры.

Выбранный алгоритм Sort [2] использует фильтры Калмана и Венгерские алгоритмы, что делает его одним из наиболее точных и популярных при решении задачи трекинга объектов на изображениях.

Иллюстрация работы алгоритма представлена на рисунке 6.

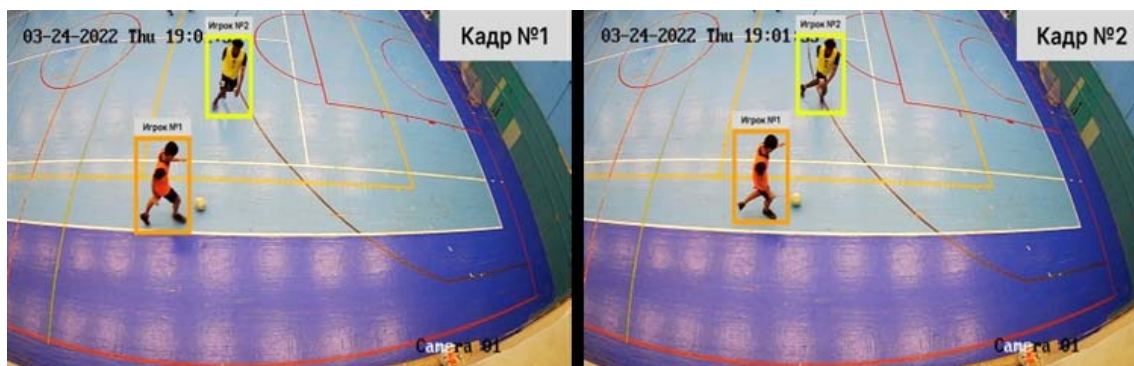


Рис. 6. Результат трекинга объектов на изображении

Большинство командных видов спорта в процессе игры используют всего один мяч. Соответственно, отсутствует необходимость в трекинге спортивного инвентаря и алгоритм Sort применяется только к спортсменам.

Определение позы спортсменов

Задача определения трехмерной позы человека сводится к определению ключевых трехмерных точек и соотношения их с метками частей тела. Для решения задачи используется модель VIBE [3], особенностью которой является частичная совместимость с SMPL [4] моделью, что делает её крайне удобной при отображении данных в виртуальной реальности. На вход модели подается изображение, а также информация об ограничивающих рамках. Результатом вычисления является набор сведений о расположении частей тела, параметров камеры, с которой велась съемка. Пример визуализации вывода метода представлен на рисунке 7.

Стоит отметить, что область распознавания трехмерной позы человека находится на начальной стадии, но уже с существующими моделями можно добиваться неплохих результатов. Однако позы моделей, построенных по данным, полученным из VIBE, все же отличаются от оригиналов небольшим смещением частей тела.



Рис. 7. Визуализация результатов определения позы спортсменов

Определение внешнего вида спортсменов

Важнейшей задачей является определение внешнего вида спортсмена, поскольку при воспроизведении игровой ситуации необходимо различать как партнеров по команде, так и соперников. Для её решения используется модель DensePose [5], которая принимает на вход изображение, ограничительные рамки человека и в ответ выдает текстуру, представленную в формате Atlas. Пример полученной текстуры и оригинального изображения представлен на рисунке 8.

В процессе выполнения модель DensePose занимает большое количество оперативной памяти. Также формат Atlas несовместим с SMPL моделью, поэтому возникает необходимость в дополнительных преобразованиях. Однако результат её работы позволяет придать моделям вид, похожий на оригинального человека из изображения (рис. 8).

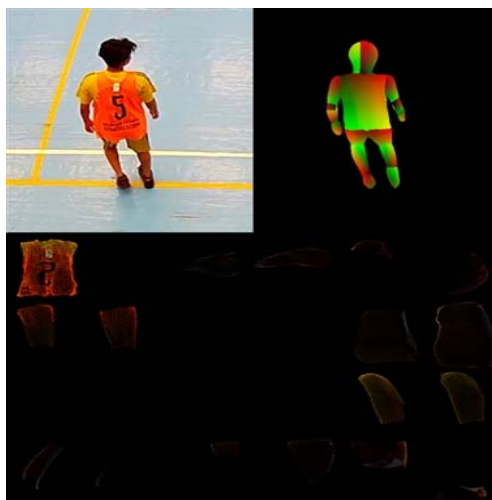


Рис. 8. Получение текстуры из изображения спортсмена

Отображение данных на плоскости поля

Из модуля, решающего задачу обнаружения спортсменов и спортивного инвентаря на изображении, известны ограничивающие рамки всех объектов, но одного этого недостаточно для того, чтобы правильно расположить модели на плоскости поля. Для решения этой проблемы система предлагает предварительную настройку камер и соотносящихся к ним видеофрагментам игры. Поскольку камеры находятся в фиксированном положении, то перед запуском вычислений можно определить матрицы гомографии, благодаря которым в дальнейшем будет проходить преобразование из точек, расположенных на плоскости фрагментов, на плоскость поля.

Благодаря этому на выходе модуля известны достоверные расположения игроков и спортивного инвентаря. Для определения матрицы гомографии необходимо задать соответствие между четырьмя точками на двух изображениях. Пример расстановки ключевых точек представлен на рисунке 9.

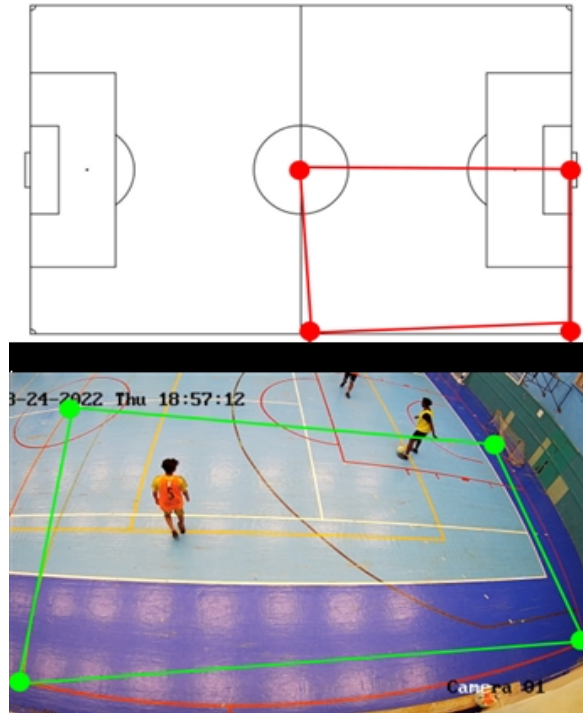


Рис. 9. Сопоставление четырех точек на плоскости фрагмента и плоскости поля

Связь между точками плоскости изображения и плоскости поля можно определить по формуле:

Результат вычислений по матрице гомографии для одного изображения представлен на рисунке 10.

$$x'_i = \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}}, y'_i = \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}}, \quad (1)$$

где h_{ij} – элемент матрицы гомографии.

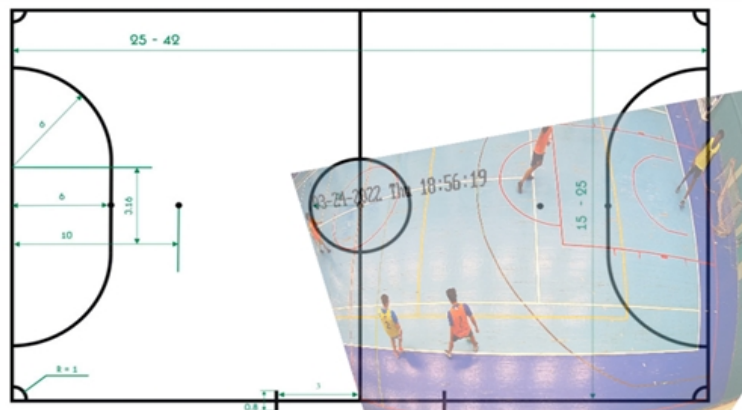


Рис. 10. Наложение точек изображения на плоскость поля

После получения ограничительных рамок объектов их можно отобразить на плоскости поля. Для этого берется середина нижней границы и производится четырехточечное преобразование, по которому для спортсмена или мяча определяются координаты расположения на спортивной площадке.

Объединение данных с различных камер

В процессе игры возникают ситуации, когда игрок мог появиться в зоне захвата одной камеры, а затем попасть в зону видимости другой. Для системы крайне важно понимать, что это один и тот же человек, поскольку от этого зависит вычисление метрик его эффективности. Это типичная задача реидентификации человека на изображениях, для решения которой было предложено использование алгоритма AlignedReId [6], который достигает крайне высокой точности на датасете SUNK03, содержащем 14097 изображений с 1467 разными людьми. На вход подается набор обнаружений игроков с различных камер, производится поиск соответствия между ними. Пример вывода модуля представлен на рисунке 11.

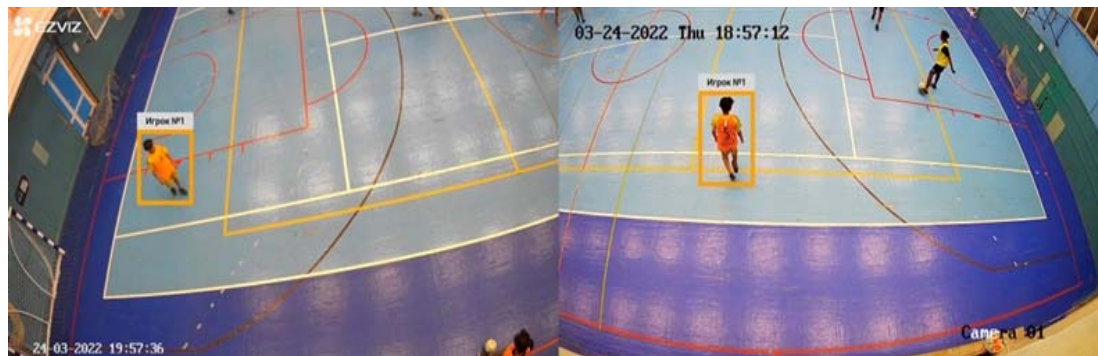


Рис. 11. Вывод модуля для реидентификации спортсмена на изображениях

В результате системе становятся известны данные о появлении спортсмена на различных камерах, что дает возможность улучшить точность его отображения в виртуальной реальности. Ранее были вычислены текстуры для различных изображений. Использование данных обо всех появлениях спортсмена можно использовать для улучшения внешнего вида модели.

На рисунке 12 представлены результаты получения текстур для спортсмена на каждом изображении, а также объединенная текстура, содержащая в себе все остальные.



Рис. 12. Результат объединения текстур

В процессе объединения может возникнуть искажение результирующей текстуры. В случае вышеописанного примера видно, как номер со спины спортсмена получает некоторое зашумление.

Представление игровой ситуации в виртуальной реальности

Для отображения игроков в виртуальной реальности используется SMPL модель человека, которая содержит в себе информацию о метках классов частей тела, что значительно упрощает преобразование данных, полученных с модуля обнаружения трехмерной позы человека. Однако одного этого мало, координаты находятся в формате ось-угол, их необходимо преобразовать в формат матрицы вращения по формуле поворота Родригеса:

$$\vec{R}_2(\vec{e} * \vec{R}_1)(1 - \cos(\chi))\vec{e} + (\vec{e} \times \vec{R}_1)\sin(\chi) + \vec{R}_1\cos(\chi), \quad (2)$$

где \vec{R}_1 – исходный вектор, \vec{R}_2 – результирующий вектор, \vec{e} – единичный вектор оси поворота, χ – угол поворота. После вычислений корректируется поза модели в виртуальной реальности, результат показан на рисунке 13.



Рис. 13. Отображение позы спортсмена в трехмерном пространстве

Текстуры, полученные из модуля для определения внешнего вида спортсмена, имеют формат Atlas, который несовместим с SMPL моделью, поэтому система автоматически переводит их в другой формат, который во всех источниках называется нормальным. В процессе конвертации происходит минимальная потеря данных.

Результат преобразования показан на рисунке 14.

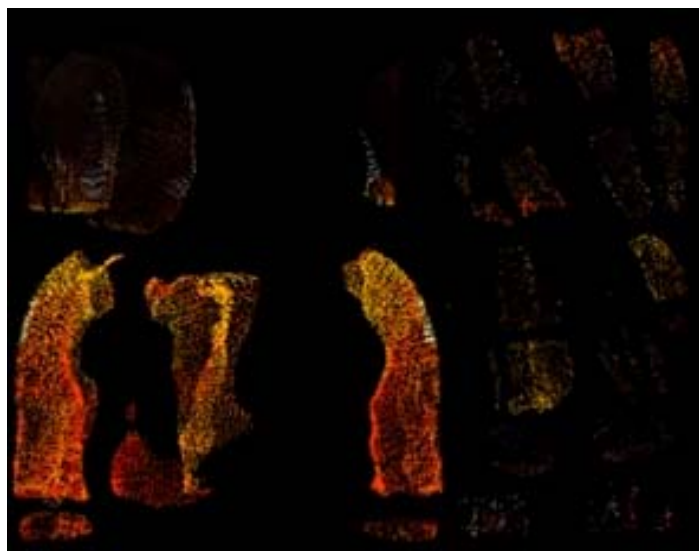


Рис. 14. Результат конвертации текстуры в нормальный вид

Основная задача переноса текстуры – это отображение игровой формы спортсмена, с чем отлично справляется модель DensePose, выделяя базовую расцветку футболки и шорт.

По результатам обработки всех данных текстура применяется к модели, иллюстрация показана на рисунке 15.



Рис. 15. Отображение текстуры на SMPL-модели

Заключение

Предложенная система действительно может стать новым инструментом для обучения спортсменов. По результатам обработки собранного датасета обнаружилась уязвимость модели VIBE, которая заключается в том, что если запись ведется сверху, а спортсмен стоит под камерой, сеть считает, что он двигается в полуприседе.

Для того чтобы избежать этой проблемы, было предложено вести съемку дальних сторон, в которых видно людей в полный рост. Используемый в системе детектор YoloV3 работает с высокой точностью, хорошо распознавая, в том числе и спортивный инвентарь.

Основной акцент при разработке решения делается не на скорость, а на точность преобразования игровой ситуации в виртуальную реальность. Предполагается, что после настройки камер приложение приступает к обработке данных, которая может занимать некоторое время. Вывод системы сохраняется в отдельный файл проекта, что позволит легко и быстро запускать уже готовые результаты для работы.

Особый акцент в текущий момент делается на модель DensePose, которая оказалась крайне требовательной к ресурсам, происходит поиск вариантов оптимизации алгоритма. Также при использовании метода AlignedReId существует проблема, которая заключается в том, что спортсмены играют в одинаковой форме, что усложняет поиск уникальных признаков на изображении человека.

Литература

1. Redmon J., Farhadi A. *Yolov3: An incremental improvement* //arXiv preprint arXiv:1804.02767. 2018.
2. Wojke N., Bewley A., Paulus D. Simple online and realtime tracking with a deep association metric //2017 IEEE international conference on image processing (ICIP). IEEE, 2017. С. 3645-3649.
3. Kocabas M., Athanasiou N., Black M. J. *Vibe: Video inference for human body pose and shape estimation* // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. С. 5253-5263.
4. Loper M. et al. SMPL: A skinned multi-person linear model //ACM transactions on graphics (TOG). 2015. Т. 34. №. 6. С. 1-16.
5. Güler R. A., Neverova N., Kokkinos I. *Densepose: Dense human pose estimation in the wild* //Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. С. 7297-7306.
6. Zhang X. et al. *Alignedreid: Surpassing human-level performance in person re-identification* //arXiv preprint arXiv:1711.08184. 2017.

**DEVELOPMENT OF AN AI-SYSTEM FOR ANALYZING MATCHES
OF TEAM SPORTS USING VR/AR TECHNOLOGY**

Denis V. Braslavskiy,
Graduate MTUCI, Moscow, Russia,
brasd99@gmail.com

Elena A. Skorodumova,
Associate Professor of the Department of PT&AM MTUCI,
Ph.D. in Physics and Mathematics, Moscow, Russia,
[eas@mtuci.ru](mailto: eas@mtuci.ru)

Abstract

The aim of the work is to present a new system for training athletes taking part in team sports, using machine learning algorithms and representing the game situation in virtual reality. All the developed modules and the tasks they solve are described with illustrations of the results of their implementation, as well as the models of neural networks used with a list of their advantages and disadvantages. Particular attention is paid to the architecture of the system, the technology stack used, and recommendations are given for collecting data that the system accepts as input. Conclusions and suggestions for further refinement of the system are made.

Keywords: machine-learning, object-detection, object-tracking, 3d-pose-estimation, sports, VR/AR.

ОБЗОР СУЩЕСТВУЮЩИХ СИСТЕМ ОТСЛЕЖИВАНИЯ СПУТНИКОВ

Чекина Дарья Александровна,
студент МТУСИ, Москва, Россия,
darachekina@gmail.com

Пронина Евгения Дмитриевна,
старший преподаватель кафедры СиСРТ
МТУСИ, Москва, Россия,
e.d.pronina@mtuci.ru

Аннотация

Спутники типа *CubeSat* являются быстроразвивающейся областью в космической связи благодаря их простоте в изготовлении и дешевизне. Всё больше отраслей используют малые спутники для сбора и анализа данных, поэтому программы по отслеживанию спутников являются важной частью в разработке и изучении спутниковых технологий. Проведён сравнительный анализ нескольких самых популярных программ для отслеживания спутников, рассмотрены их плюсы и минусы. Сделан вывод о том, что есть необходимость в создании новой программы, учитывающей минусы существующих программ.

Ключевые слова: спутники типа *CubeSat*, траектория полета, зона покрытия, орбита, спутниковые технологии.

Введение

Приступив к анализу орбит и траекторий полётов спутников типа *CubeSat*, можно неожиданно столкнуться с такой проблемой, как отсутствие адекватных программ для моделирования вышеуказанных параметров. В интернете существует много программ для отслеживания спутников в режиме реального времени, расчёта некоторых параметров полёта, а также во многих из этих программ есть большая база спутников, включая спутники типа *CubeSat*. Однако, при более детальном их изучении, становится ясно, что программы не умеют моделировать траекторию полётов малых спутников и учитывать их срок службы.

Стандартный спутник типа *CubeSat* имеет размеры $10 \times 10 \times 10$ см³ [1] и массу 1 кг [2] и данная конфигурация носит название *1U*. Спутники типа *CubeSat* в основном собираются в форм-факторах *1U* и *3U*, но также часто могут иметь размеры *6U* (12 кг), *12U* (24 кг) и даже *26U* (54 кг) [3].

Наноспутники являются экономически эффективным средством для проведения научных и технологических исследований в космосе. В последние годы наблюдается огромный рост интереса к *Cube*-спутникам (*CubeSats*) среди космического сообщества, включая космические агентства, промышленность и научные круги. На этот всплеск интереса повлияли два фактора: во-первых, недорогой доступ к космосу в качестве вторичной полезной нагрузки с целью демонстрации технологий, проверки научной концепции, коммуникации и образования; во-вторых, использование готовых коммерческих технологий (*COTS*) в архитектуре проектирования. Эти два фактора привели к значительному снижению общей стоимости миссии *Cube*-спутников [4].

Многие космические агентства предоставляют возможности запуска спутника типа *CubeSat* академическим кругам посредством различных запросов предложений, в результате чего университеты по всему миру запускают свои собственные малые спутники и предоставляют студентам ценный практический опыт в области проектирования космических систем. Например, в Москве разработкой спутников типа *CubeSat* занимаются МГУ им. М.В. Ломоносова, МГТУ им. Н.Э. Баумана и МТУСИ.

На текущий день *CubeSat* используются только для околоземных миссий, однако уже было предложено несколько дальних околоземных и межпланетных миссий (*INSPIRE* и *MarCO*) [5,6]. Операции малых спутников ограничены в основном из-за их небольшого размера, что уменьшает их бортовые возможности (мощность, мобильность и полезную нагрузку), что приводит к снижению срока службы и дальности полета (после развертывания). Анализ влияния основных факторов, таких как высота орбиты, влияние поперечного сечения, влияние атмосферы, на срок службы *Cube*-спутников был проведён в предыдущей статье [7]. Исходя из полученных данных, оказалось, что срок жизни таких спутников довольно мал и составляет приблизительно от полугода до двух лет.

Однако программы, следящие за движением спутников, не учитывают данный фактор, и можно заметить, как спутник типа *CubeSat* летает долгие годы даже после окончания срока службы по той же самой траектории, хотя в реальности уже давно покинул свою орбиту.

Ниже приведен сравнительный анализ нескольких самых популярных программ для отслеживания спутников, рассмотрены их плюсы и минусы.

Программы для отслеживания спутников

Программа по запросу

Самой первой программой по запросу «Отслеживание спутников в режиме онлайн» в списке является программа «Спутниковая карта онлайн в реальном времени» (рис. 1) [8].

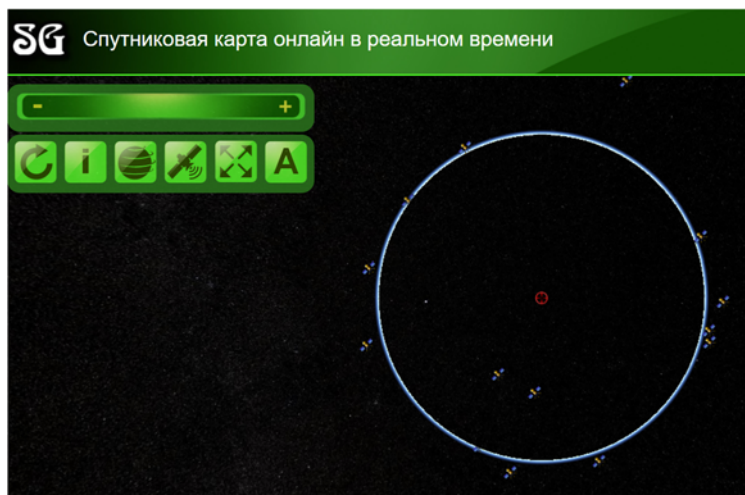


Рис. 1. Спутниковая онлайн-карта

Из минусов пользования программой можно выделить следующее:

1. Интерфейс не является интуитивно понятным.
2. Нет зоны покрытия спутников: зона покрытия является одним из важнейших параметров, который описывает характеристики спутников. Без ЗП данные о спутнике являются не информативными и носят скорее любительский характер, не представляя научному сообществу особого интереса.
3. Нет моделирования времени: невозможно произвести расчёт, когда спутник окажется над нужным местоположением через некоторое количество времени. Также, глядя на спутник, невозможно понять, в какую сторону он летит.
4. Сайт обновлялся последний раз в 2014 году, что говорит о некорректной базе спутников: новые спутники, выводимые на орбиту, не отображены в списках, а спутники, выведенные из использования, всё ещё отслеживаются.
5. Вполне вероятно, что форма Земли описана геометрически не совсем корректно, поскольку на сайте изображена Земля идеально круглой формы, в то время как реальная форма Земли - эллипсоид или геоид.

К плюсам данной программы можно отнести:

1. Достаточно большой выбор спутников даже для последнего обновления в 2014 году.
2. При выборе спутника можно выбрать разные группировки: военные спутники, инженерные, исследовательские и т.д., а также отследить крупный космический мусор (обломки разных спутников (рис. 2)).

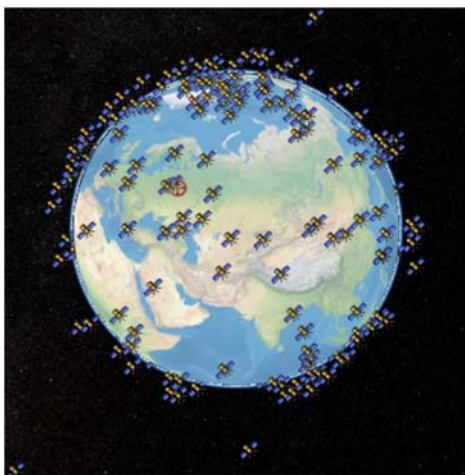


Рис. 2. Обломки спутника IRIDIUM 33

Программа *Orbitron*

Далее рассмотрим *Orbitron* - систему слежения за спутниками, предназначенную для радиолюбителей и любителей визуальных наблюдений. Ее также применяют профессионалы-метеорологи, пользователи спутниковой связи, люди, увлекающиеся поиском НЛО и астрологи.

Программа показывает положения спутников в любой заданный момент (как в реальном времени, так и в режиме симуляции). Программа бесплатна, и считается одной из самых простых в обращении, и одновременно одной из самых мощных программ слежения за спутниками, по мнению тысяч ее пользователей со всего мира [9]. Интерфейс программы показан на рисунке 3.

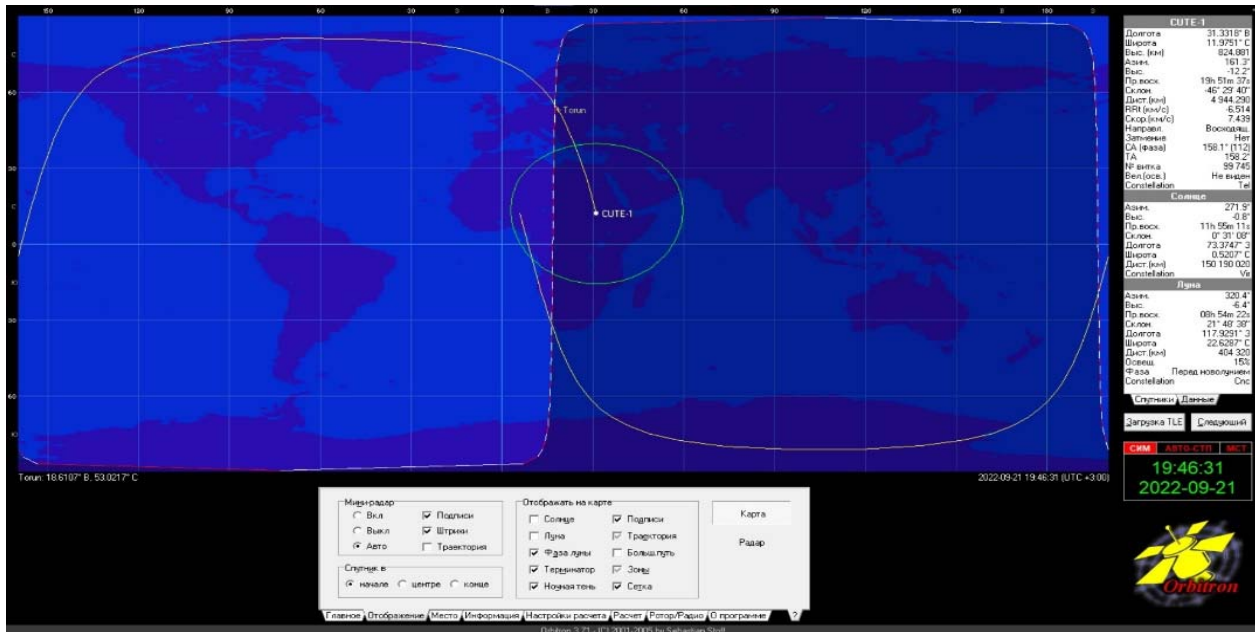


Рис. 3. Интерфейс программы *Orbitron*.

Для начала следует отметить неоспоримые плюсы данной программы, и в первую очередь отметить, что программа является единственной в таком роде. Во-вторых, в программе можно выполнять моделирование во времени, т.е. смотреть, в какой точке будет спутник через несколько часов или минут, и линией выделена траектория движения спутника. И третий плюс, который обязательно нужно отметить – то, что *Orbitron* умеет рассчитывать различные параметры, например, геометрические соотношения для спутника (азимут, угол места и т.д.). Эти же параметры умеют рассчитывать и другие программы, например, *Satellite Antenna Alignment*, но в данной программе нет визуализации орбит спутников.

Но даже в этой программе есть минусы, на которые стоит обратить внимание:

1. Техническая поддержка программы *Orbitron* закончилась в 2003 году, и последнее обновление, которое можно найти в открытом доступе, было в 2005 году. В связи с обновлением операционных систем, у многих возникают проблемы с нахождением необходимого спутника для работы.

2. Нет документации (инструкции) по работе с программой. Многие функции, например, расчёт «прохождение и вспышки» непонятны человеку, без специального обучения работы с программой и образования в данной области (рис. 4).

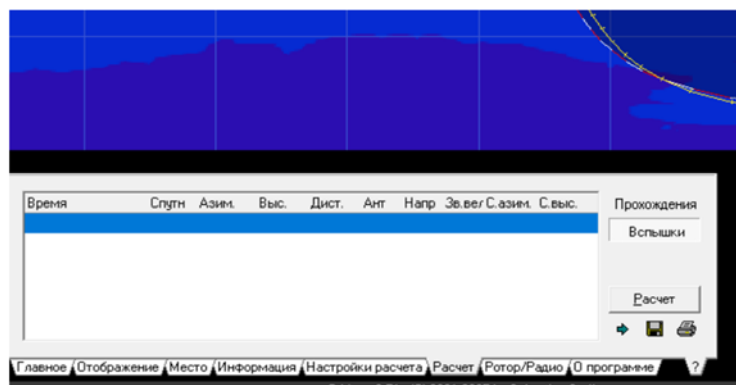


Рис. 4. Вкладки программы

3. Не учитываются сроки службы спутников, и, следовательно, траектории снижения (ухода с орбиты) спутников. Даже через много лет спутники находятся на своих орбитах на тех же высотах, что является весьма неинформативным для спутников типа *CubeSat*, которые, как уже было сказано выше, имеют срок службы от полугода до двух лет, и далее существование их на орбите вводит пользователя в заблуждение.

Программа *WeatherRadarLive*

Следующая рассматриваемая программа - *WeatherRadarLive* [10]. Следует отметить, что это сайт с различными областями для исследований. Можно выбрать следующие функции: радар погоды, природные силы, флайтрадар и спутниковую карту (рисунок 5). Так как в статье проводится сравнение программ отслеживания спутников, то следует выбрать вкладку «Спутниковая карта».



Рис. 5. Функционал сайта *WeatherRadarLive*

Основным минусом данного сайта является малый выбор отслеживаемых спутников, а точнее, представлены только спутники *Starlink* компании *SpaceX* (рис. 6). Вторым минусом можно отметить крайне малое количество спутников, отслеживаемых в режиме реального времени. На сайте представлено всего семь спутников и Международная космическая станция (МКС), (рис. 7).

К другим минусам сайта можно отнести следующее:

- в отличие от того же *Orbitron*, отсутствует расчёт геометрических параметров;
- отсутствует моделирование во времени, нет возможности правильно спрогнозировать подспутниковую точку, просто воспользовавшись данным сайтом;
- не отображена зона видимости.

И главный минус, по мнению авторов статьи, заключается в том, что на сайте *WeatherRadarLive* не представлена база спутников типа *Cubesat*.

Однако следует выделить и плюсы этого сайта:

- как было упомянуто выше, на сайте можно выбрать не только систему отслеживания спутников, но и другие функции;
- на каждой вкладке есть небольшая теория, позволяющая получить минимальные знания и умения для работы с выбранным параметром;
- каждая страна подписана на своём языке, что создаёт приятную атмосферу при работе с сайтом;
- траектория каждого спутника на спутниковом радаре выделена своим цветом, что позволяет легко отследить перемещение выбранного спутника.

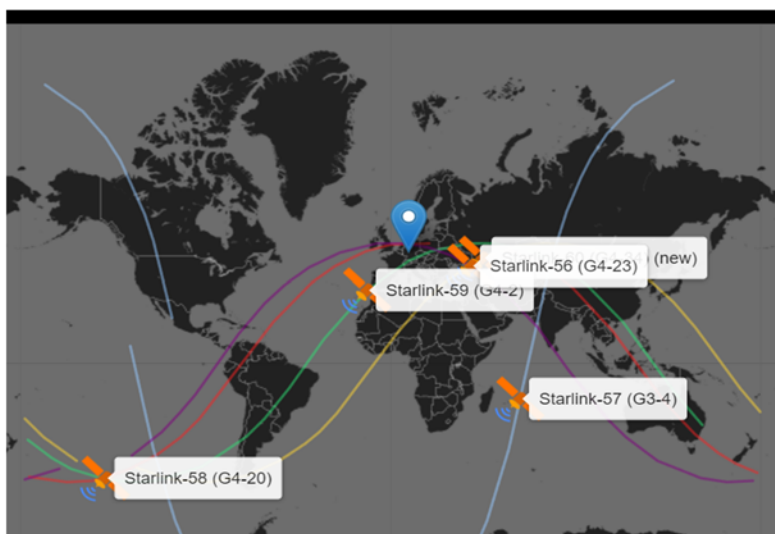


Рис. 6. Спутниковый радар для спутников *Starlink* компании *SpaceX*

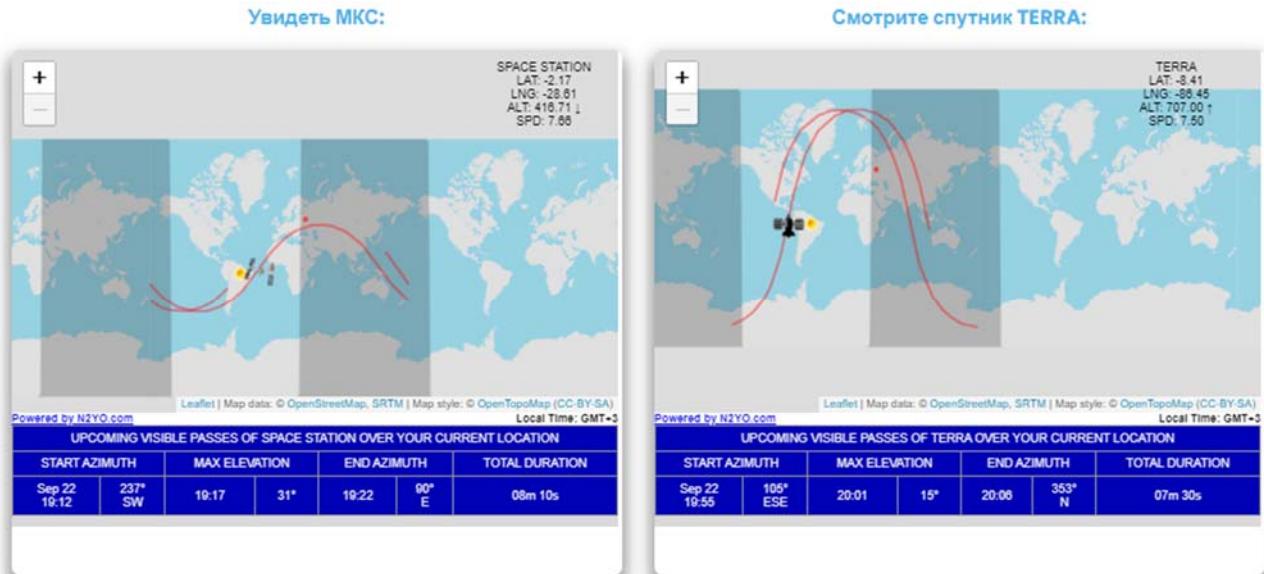


Рис. 7. Спутниковая карта (МКС и спутник *TERRA*)

Мобильное приложение *Satellite Tracker*

Для дальнейшего анализа программ для отслеживания спутников стоит рассмотреть продукт от *Vivo Technology* [11]. *Vivo Technology* – компания по разработке мобильных приложений, направленных на изучение и популяризацию космоса. Приложение *Satellite Tracker* разработано для наблюдения за искусственными спутниками, например, *Starlink* или МКС, пересекающими ночное небо [12]. Оно может показывать точное положение различных спутников (спутников разных стран, конфигураций и назначений) над Землей и отправлять уведомления, когда интересующий спутник будет проходить над выбранным местоположением. Интерфейс приложения показан на рисунке 8.

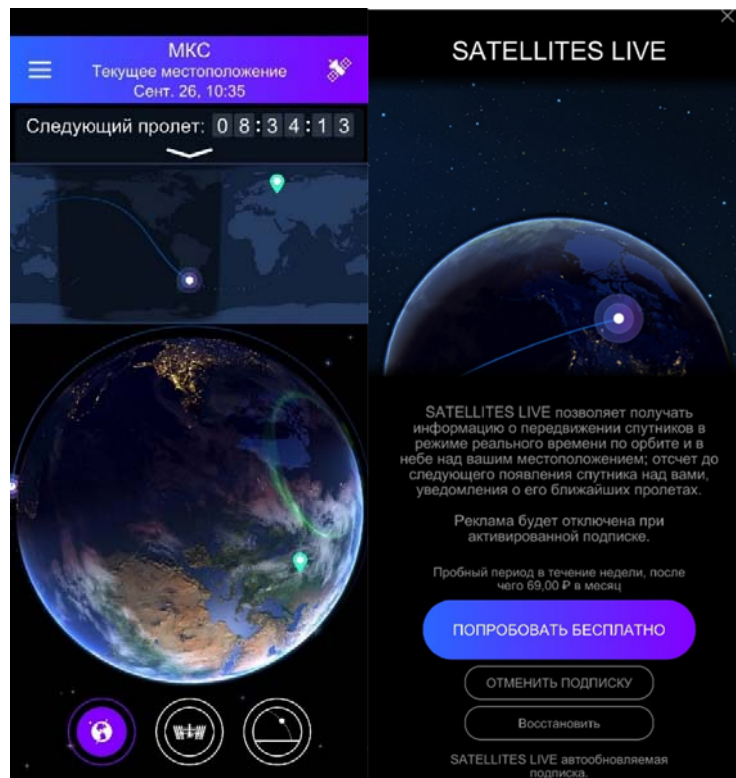


Рис. 8. Интерфейс приложение *Satellite Tracker*

Плюсы приложения *Satellite Tracker*:

1. Доступность в *App store* и *Google play*;
2. Красивый и интуитивно понятный интерфейс;
3. Возможность наглядного отслеживания аппарата с помощью изменения положения телефона;
4. Есть функция получения уведомлений в момент, когда аппарат пролетает над приемным устройством;
5. Большое количество спутников разного вида.

К минусам приложения следует отнести:

1. Платная подписка: бесплатная подписка позволяет отслеживать только МКС;
2. Нет возможности моделирования траектории полета: есть только функция отслеживания;
3. Нет поддержки в виде сайта: функционал доступен только через приложение;
4. Нет возможности перейти в *2D* режим: приложение направлено на визуальную составляющую, а не на реальные расчеты;
5. Нет зоны покрытия спутника: можно только отслеживать спутник, но нет возможности получить какие-либо технические или расчетные данные;
6. Нет границ стран на карте, что усложняет восприятие точного местоположение спутника.

Все рассмотренные программы находятся в свободном доступе и относятся к категории любительских, ими могут воспользоваться студенты или начинающие радиолюбители.

Однако существуют и более совершенные программы для профессиональных исследований. Например, программное обеспечение для цифрового проектирования миссий и системного анализа *Ansys STK (Satellite Tool Kit)* [13]. Данная программа позволяет проектировать системы космических полетов, радиочастотных систем и многое другое.

Профессиональных программ для отслеживания спутников разработано достаточно большое количество, но стоит отметить, что все они либо в закрытом доступе, т.е. разработаны на базе исследовательских институтов и находятся только в их пользовании. В отдельных случаях к ним возможен платный доступ.

Заключение

По результатам приведённого анализа программ отслеживания спутников можно сделать вывод, что каждой из программ свойственны как плюсы, так и минусы. В целом ни одна из программ, находящихся в свободном доступе, не отвечает всем необходимым требованиям, а именно:

1. Правильная физическая форма Земли: для облегчения программирования редко прописывается реальная форма Земли, происходит ее упрощение до сферы;
2. Интуитивно-понятный интерфейс;
3. Регулярное обновление: почти все программы, направленные на проектирование траектории спутника не обновлялись уже продолжительное время;
4. Возможность моделирования: в большей части программ отсутствует возможность моделирования полета спутника;
5. Возможность делать *2D* и *3D* вид: при визуализации траектории важна возможность представления Земли в *2D* и *3D* моделировании для лучшего восприятия полученных данных пользователем;
6. Возможность одновременно запускать несколько спутников: для анализа полезно иметь возможность наглядного сравнения нескольких спутников.
7. Отдельное окно для технических данных и результатов расчетов;
8. Возможность моделирования теоретического спутника: в открытом доступе отсутствует возможность расчета теоретического спутника, есть возможность работы только с существующими экземплярами;
9. Расчет жизнедеятельности спутника: все расчеты проводятся в идеальных условиях, и расчет ухода спутника с орбиты отсутствует, что особенно важно при проектировании спутников типа *CubeSat*;
10. Зона покрытия: чаще всего отсутствует расчет и визуализация зоны покрытия спутника, которая является одним из важнейших параметров для работы со спутниковыми технологиями.
11. Большинство пользователей программ в свободном доступе являются студентами, у которых нет доступа к профессиональным ресурсам. Некоторые из перечисленных минусов являются критичными для анализа и полного понимания студентами работы спутниковой системы связи. При этом всё большее число студентов и аспирантов проявляют заинтересованность в программах, полностью удовлетворяющих их запросам

Литература

1. *Barnhart D.J., Vladimirova T., Sweeting M.N.* Very-small-satellite design for distributed space missions. *J. Spacecr. Rocket* 2007, no. 44, pp. 1294-1306.
2. *Toorian A., Blundell E., Suari J.P., Twiggs R.* CubeSats as responsive satellites. *Aerosp. Eng.* 2005, no. 805, pp. 756-6479.
3. *Hevner R., Holemans W., Puig-Suari J., Twiggs R.* An advanced Standard for CubeSats // *Proceedings of the 25th Annual AIAA/USU Conference on Small Satellites*, Logan, UT, USA, 8-11 August 2011.
4. *Woellert K., Ehrenfreund P., Ricco A.J., Hertzfeld H.* Cubesats: Cost-effective science and technology platforms for emerging and developing nations. *Adv. Space Res.* 2011, no. 47, pp. 663-684.

5. Klesh A., Baker J., Castillo-Rogez J., Halatek L., Murphy N., Raymond C., Sherwood B., Bellardo J., Cutler J., Lightsey G. Inspire: Interplanetary NanoSpacecraft Pathfinder // Relevant Environment. In Proceedings of the AIAA SPACE 2013 Conference and Exposition, San Diego, CA, USA, 10-12 September 2013.
6. Schoolcraft J., Klesh A., Werne T. MarCO: Interplanetary Mission Development on a CubeSat Scale. In Space Operations: Contributions from the Global Community; Springer: Daejeon, Korea, 2017, pp. 221-231.
7. Чекина Е.А., Пронина Е.Д. Факторы, влияющие на срок службы спутников типа CubeSat // Сборник трудов молодежного научного форума МТУСИ. 2022. С. 49-59.
8. Spacegid.com – интерактивный гид в мире космоса [Электронный ресурс] – Режим доступа: <https://spacegid.com/>, свободный – (дата обращения 15.10.2022).
9. Страница программы Orbitron [Электронный ресурс] – Режим доступа: <http://www.cqham.ru/orbitron.htm>, свободный – (дата обращения 15.10.2022).
10. Погодный радар LIVE [Электронный ресурс] – Режим доступа: <https://weather-radar-live.com/>, свободный – (дата обращения 15.10.2022).
11. Vito Technology [Электронный ресурс] – Режим доступа: <https://vitotechnology.com/>, свободный – (дата обращения 15.10.2022).
12. ISS, Starlink [Электронный ресурс] – Режим доступа: Tracker App <https://vitotechnology.com/apps/satellite-tracker>, свободный – (дата обращения 15.10.2022).
13. Ансис СТК | Программное обеспечение для разработки цифровых миссий [Электронный ресурс] – Режим доступа: <https://www.ansys.com/products/missions/ansys-stk>, свободный – (дата обращения 07.11.2022).

OVERVIEW OF EXISTING SATELLITE TRACKING SYSTEMS

Daria A. Chekina,
Student MTUCI, Moscow, Russia,
darachekima@gmail.com

Evgeniya D. Pronina,
Senior Lecturer of the Department of S&SRT
MTUCI, Moscow, Russia,
e.d.pronina@mtuci.ru

Abstract

CubeSat-type satellites are a rapidly developing field in space communications due to their ease of manufacture and cheapness. More and more industries are using small satellites to collect and analyze data, so satellite tracking programs are an important part in the development and study of satellite technologies. A comparative analysis of several of the most popular satellite tracking programs is carried out, their pros and cons are considered. It is concluded that there is a need to create a new program that takes into account the disadvantages of existing programs.

Keywords: *CubeSat satellites, flight path, coverage area, orbit, satellite technologies.*

ПРИМЕНЕНИЕ ЦИФРОВЫХ ДВОЙНИКОВ В ПРОМЫШЛЕННОМ ПРОИЗВОДСТВЕ

Панов Алексей Ильич,
студент ИИТММ ННГУ им. Н.И. Лобачевского,
Нижний Новгород, Россия,
aleks.npc459@mail.ru

Аннотация

С развитием науки и техники индустрия высоких технологий также быстро развивается, и продолжают появляться различные технологии нового времени, и цифровые двойники – одна из них. Будучи совершенно новой интерактивной технологией, технология цифровых двойников может хорошо справляться с взаимодействием между реальным и виртуальным миром. Они интегрированы с другими технологиями и применяются во многих областях, таких как умные фабрики в промышленном производстве, цифровая модель жизни в медицине, строительство умных городов, гарантия безопасности в аэрокосмической сфере, шопинг в коммерческой сфере и так далее. Цель данной статьи – дать возможность понять статус применения технологии цифровых двойников. В то же время внедрение основных технологий, связанных с цифровыми двойниками, перемежается с внедрением приложений.

Ключевые слова: цифровые двойники. виртуальная реальность, промышленность, фабрика, производство

Введение

История разработки цифровых двойников

Концепция цифровых двойников (*Digital twins, DT*) впервые появилась в американской спасательной миссии «Аполлон-13» в 1970 году. После запуска космического корабля в космос внезапно взорвался кислородный баллон, и ситуация стала критической. Чтобы спасти астронавтов, национальное управление по авиации и исследованию космического пространства (*National aeronautics and space administration, NASA*) использовало цифровую модель «Аполлона-13» для его имитации. После многих попыток они выработали решение и отправили его астронавтам на космическом корабле. Этот инцидент можно назвать самой ранней практикой применения цифровых двойников. Прототип технологии *DT* был также представлен Майклом Гривсом в 2002 году для презентации промышленности в Мичиганском университете по созданию управления жизненным циклом продукта (*Product Lifecycle Management, PLM*). В модели упоминаются как концепции реального, так и виртуального пространства, в которых уже есть все элементы *DT* [1]. Затем, в 2006 году, модель зеркального отображения информации была расширена. В модели была предварительная идея, как перенести реальные объекты в виртуальное пространство в цифровом двойнике, но она не предполагала взаимодействие реального объекта с виртуальным, и не указывала, какую технологию необходимо использовать для достижения этой цели (рис. 1).

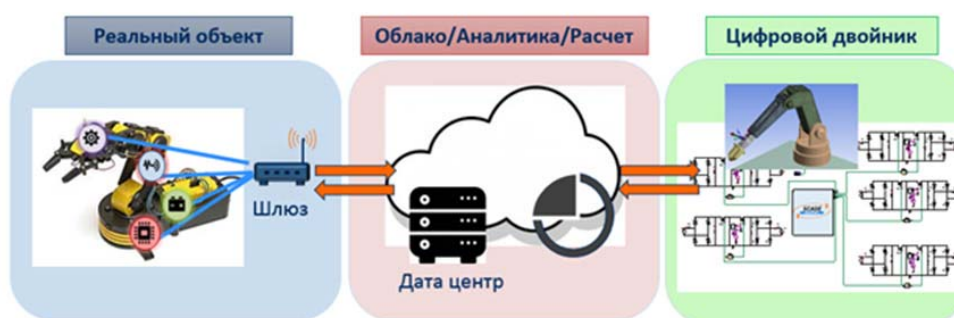


Рис. 1. Пример схемы взаимодействия цифрового двойника с реальным объектом

В 2010 году цифровой двойник был официально предложен Джоном Уиксом в отчете о дорожной карте *NASA* за 2010 год, что ознаменовало официальное рождение технологии *DT*. В 2012 году *NASA* начало использовать термин «цифровые двойники» в своей технологической дорожной карте. С 2014 года некоторые известные компании, такие как *Siemens, Dassault, PTC, ESI* и *ANSYS*, используют термин в своих маркетинговых кампаниях.

Определение – цифровой двойник (DT)

DT – это технология, которая объединяет несколько сфер. В настоящее время *DT* не имеет единого определения, так как все еще находится в постоянном развитии.

В 2012 году *NASA* определило *DT* как комплексную мультифизическую, многомасштабную систему вероятностного моделирования транспортных средств или систем.

В 2017 году университет оборонных закупок определил *DT* следующим образом: интегрированная мультифизика, многомасштабное вероятностное моделирование с использованием наилучших доступных моделей, информации датчиков и входных данных для отражения и прогнозирования функционирования, активности и производительности соответствующего физического двойника.

В 2019 году Старк Дамерау дал своё определение *DT*. *DT* – это цифровое представление, которое содержит описание характеристик выбранного объекта или его системы продуктов и услуг, и получает атрибуты, условия и поведение объекта с помощью моделей, информации и данных на одном или даже нескольких этапах жизненного цикла.

В 2020 году технология *DT* была также определена Китайским институтом стандартизации электронных технологий. *DT* означает полное использование таких данных, как физическая модель, показатели датчиков, история операций, а также интеграцию междисциплинарного, мультифизического, многомасштабного и многовероятностного процесса моделирования для завершения отображения в виртуальном пространстве, тем самым отражая полный срок службы соответствующего физического оборудования [2,3].

Таким образом, определение *DT* многогранно. Помимо этого предполагается, что искусственный интеллект становится неотъемлемой частью *DT*. Исследование, где могут быть применены алгоритмы искусственного интеллекта также является будущим направлением развития.

Характеристики цифровых двойников

DT создает высокоточную модель, соответствующую физическому миру, в виртуальном пространстве в форме оцифровки. Благодаря непрерывной обратной связи по замкнутому контуру информационного взаимодействия и слиянию данных с физическим миром можно моделировать поведение объектов в реальных сценариях.

Как видно из определения, *DT* обладают следующими характеристиками:

1) *Унификация*. Все данные в течение жизненного цикла физической системы хранятся на цифровой магистрали для централизованного и унифицированного управления, что делает двустороннюю передачу данных более эффективной.

2) *Интегрированность*. Для сложных систем *DT* интегрирует все подсистемы, что является основой для высокоточного моделирования, в то время как мониторинг данных в реальном времени может еще больше обогатить и улучшить модель, позволяя модели содержать все знания о системе.

3) *Динамичность*. Данные датчиков, описывающие среду или состояние физической системы, могут использоваться для динамического мониторинга модели, обновленные модели могут динамически направлять фактическую работу, а взаимодействие в реальном времени между физическими системами и цифровыми моделями позволяет моделям расти и развиваться на протяжении всего их жизненного цикла [4].

Характеристики *DT* позволяют ему изменять процесс выполнения традиционных проектов.

Традиционные процессы разработки проектов обычно начинаются с требований, затем внедряются, тестируются и после этого повторяются по всему процессу, часто вынуждая взаимодействовать различные отделы. Однако всегда возникают проблемы с информационным взаимодействием между различными подразделениями, что приводит к неэффективному проектированию. *DT* может управлять всеми видами информации унифицированным способом, а различные отделы могут получать доступ к данным из *DT* и добавлять их в любое время, что позволяет отделам работать параллельно. При традиционном проектировании системы сначала необходимо изготовить физическое оборудование, а затем проверить систему. Виртуальное моделирование *DT* позволяет проводить виртуальное тестирование до изготовления физических устройств.

Традиционный подход к моделированию заключается в прогнозировании окружающей среды и нагрузки, которые могут возникнуть в будущем, на основе исторических данных, и использовании этой модели для руководства производством последующих продуктов. Ошибки неизбежно будут возникать в процессе реального производства, и на этапе проектирования трудно имитировать реальную среду. На протяжении всего жизненного цикла физической системы *DT* использует данные мониторинга в реальном времени для динамического обновления модели, так что модель может не только отражать состояние устройства, но и отслеживать изменение системного окружения в сложной среде.

Именно *DT* обладают этими преимуществами, поэтому они могут быть применены к множеству реальных сценариев [5].

Применение цифровых двойников в промышленном производстве

Современная промышленность вступила в процесс развития всесторонней цифровизации [6]. В производственных системах, направленных на повышение производительности, качества и эксплуатационных характеристик продукции, современная промышленность все больше зависит от цифрового проектирования и производственных процессов.

В процессе четвертой промышленной революции одной из проблем, которые необходимо решить, является построение канала связи между физическим и цифровым мирами. В этом и заключается роль цифровых двойников. Цифровые технологии не только предоставляют возможность для ускоренных инноваций в промышленном производстве, но и значительно меняют традиционную модель промышленной системы и придают новый импульс традиционной производственной системе. Сегодня, при развитии промышленного интернета вещей, промышленный искусственный интеллект постепенно приобретает форму, и *DT* являются ключом к промышленной модернизации. В настоящее время эта технология в основном используется в областях с высокой добавленной стоимостью, таких как ракетостроение, системы добычи нефти и газа. Согласно анализу технологических тенденций Gartner, технология *DT* достигнет зрелости в ближайшие 5-8 лет [7].

С ростом популярности 5G и промышленных платформ интернета вещей глобальная технология *DT* неизбежно будет использоваться в больших масштабах. Сейчас промышленность вступает в эпоху индустрии 4.0, и технология *DT* будет играть очень важную роль в процессе этого перехода. Применение *DT* в индустрии 4.0 является горячей темой исследований и неотложной проблемой, требующей решения. Урбан Август, старший вице-президент *Siemens*, говорит, что *DT* – это важнейшая технология для перехода от традиционных предприятий к цифровым [8]. Структурная схема цифрового предприятия изображена на рисунке 2.

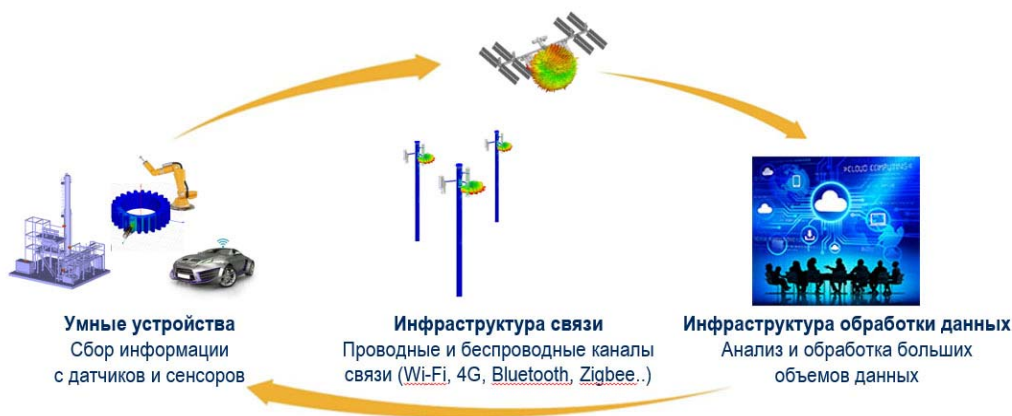


Рис. 2. Структурная схема цифрового предприятия

Как показано на рисунке 3, интеллектуальная фабрика, основанная на *DT*, играет важную роль в каждом звене жизненного цикла производства продукта:

1) Разработчики продуктов могут полагаться на *DT* для проведения моделирования и анализа надежности продукта, а затем вносить улучшения в проект и улучшать удобство использования продукта.

2) Для производителей и валидаторов надежная оценка может быть проведена с использованием *DT* и комбинирования продуктов в натуральном выражении.

3) Предприятие может полагаться на *DT* при анализе безопасности продукта, прогнозом техническом обслуживании и другой работе. Когда изделие утилизируется, компания может полагаться на *DT* для проведения анализа остаточного ресурса и оценки повторного производства каждой системы изделия, что поможет в разработке продуктов следующего поколения.

4) Для пользователей цифровые двойники могут использоваться для прогнозирования доступности продукта и получения более полного представления о продукте. В то же время система может быть использована более удобно в эксплуатации.

Что касается применения *DT* в промышленности, многие ученые также провели соответствующие исследования. *Siemens* использует *DT* на своих заводах для отладки оборудования [9]. Моделирование сложных систем с помощью *DT* позволяет как можно большему числу ключевых сотрудников участвовать в разработке и развертывании проекта, а также упрощает настройку и адаптацию устройств, повышая гибкость развертывания. Что касается моделирования, то был предложен подход к моделированию киберфизической информационной системы (*Cyber-physical production system, CPPS*), основанный на *AutomationML*. Это обеспечивает безопасный и эффективный метод моделирования данных для фабрики. Роль *AutomationML* заключается в инкапсуляции различных производственных и отраслевых сервисов и отделений, а затем интеграции *DT* в *CPPS* для предоставления услуг заводам. *CCPS* – это многомерная сложная система, которая объединяет вычислительную, сетевую и физическую среды и используется для управления сложными системами.

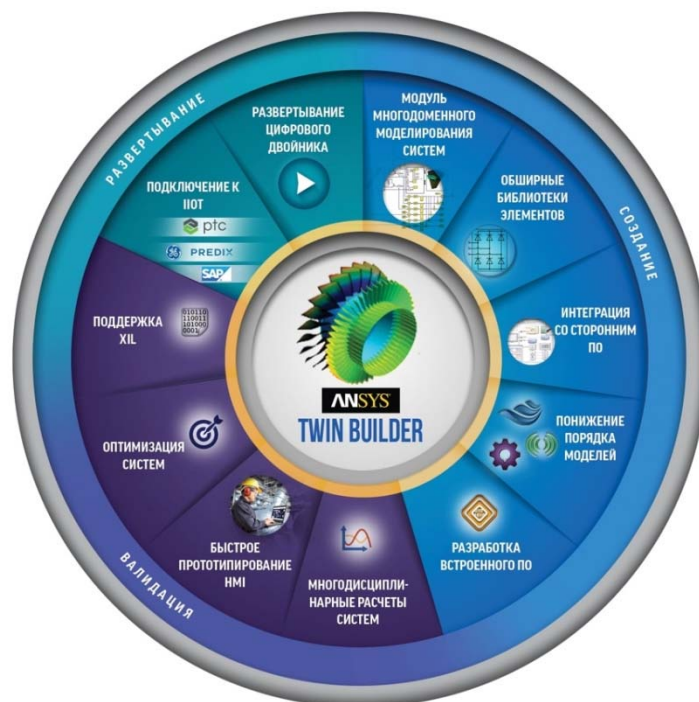


Рис. 3. Пример жизненного цикла продукта умной фабрики на основе ANSYS TWIN BUILDER

Например, компания *Schuler* начала внедрять *DT* в своё производство. Вице-президент *Schuler* предложил решение для развертывания *DT* в *Schuler*, и это дало первые результаты. С 2016 года станки производственной линии *Schuler press* оснащены датчиками для сбора данных со станков. Это означает, что появилась возможность детально отслеживать состояние системы и процесс работы. Кроме того, функция интеллектуальной диагностики, реализованная на основе алгоритма машинного обучения, может проводить ретроспективный анализ для определения причины сбоя при выходе из строя производственной линии.

Из текущих исследований следует, что в будущем *DT* станет основной технологией умных фабрик. Технология *DT* может оптимизировать исходную модель после её применения и вносить изменения в модель обслуживания продукта, чтобы полностью удовлетворить потребности потребителей в ресурсах данных *DT*.

Тем не менее, существует также значительный разрыв в степени развития *DT* в разных странах. До популярности цифровых двойников в мировой индустрии еще далеко. Кроме того, промышленный интернет вещей требует высокого уровня надежности сети и времени его реализации.

На данный момент выдвигается очень мало гипотез о том, как обеспечить надежность цифровой модели. Возможно также использование криптографических характеристик блокчейна, чтобы обеспечить безопасность данных программы *DT*.

Перспективы развития цифровых двойников

Из-за быстрого развития науки и техники *DT* по-прежнему является концепцией, находящейся в постоянном процессе совершенствования. В настоящее время все еще существует много проблем, которые необходимо решить в продвижении *DT* для практического применения. В их числе: точность моделирования, защита данных, надёжность сети и другие ключевые аспекты управления сложными системами. Необходимо также создание открытой и коммуникационной платформы для индустрии производства *DT*, содействие интеграции *DT* с другими технологиями, чтобы *DT* лучше мог использовать данные моделируемых объектов. Наиболее важной и трудной задачей является разработка международных единых стандартов *DT* для всех отраслей промышленности. Только после решения этих проблем можно будет популяризировать *DT* в реальном производстве и ускорить процесс оцифровки экономики и социальной сферы.

В настоящее время промышленное производство начало демонстрировать склонность к инновациям и цифровизации [10]. *Schuler*, *Siemens* и другие крупные компании использовали *DT* в реальном производстве и добились определенных результатов. Хотя технология *DT* еще не достигла зрелости, но с углублением исследований преимущества *DT* для промышленного производства предсказуемы. В процессе разработки промышленность в полной мере использует цифровые технологии, которые могут повысить эффективность производства и оптимизировать качество продукции, в то же время обеспечивая хорошие эксплуатационные характеристики продукта. В наши дни цифровой производственный процесс и цифровой дизайн широко используются в промышленном производстве.

Цифровые технологии могут ускорить технологические инновации в промышленности, максимально оптимизировать традиционную модель промышленного производства и сделать систему промышленного производства более интеллектуальной.

С наступлением эры 5G промышленный интернет вещей также получил широкое распространение, так что промышленные предприятия по всему миру будут обращать внимание на технологию *DT* и применять ее. Технология *DT* может помочь людям по-новому взглянуть на ситуацию с точки зрения цифровых технологий, пересмотреть недостатки традиционных корпоративных моделей и создать пространство для инноваций. Внедрение технологии *DT* повлечёт за собой глубокую интеграции технологий моделирования и большие данные, интернета вещей, искусственного интеллекта и др. Применение *DT* в промышленном производстве может революционизировать разработку промышленной продукции, производственные процессы, техническое обслуживание оборудования и т.д.

Текущая тенденция развития *DT* в промышленности состоит в следующем:

1) Сбор и обработка данных промышленного производственного оборудования.

Для того чтобы реализовать интеллектуальное развитие промышленности, необходимо сначала обеспечить интеллектуальность промышленного производственного оборудования, то есть обеспечить исправную работу оборудования. Должны осуществляться контроль состояния оборудования в режиме реального времени и прогнозирование возможных сбоев, чтобы снизить затраты на техническое обслуживание.

2) Интеллектуализация производственного процесса.

Интеллектуальное производство заключается в обеспечении безопасности и надежности производственного процесса, а также в максимально возможном снижении затрат на рабочую силу. На производстве следует в полной мере использовать интеллектуальные системы и оборудование, чтобы добиться автоматизации работы оборудования и беспилотной эксплуатации. Вкратце, это использование *DT* для мониторинга производственного процесса, обеспечения нормальной работы производства и требуемого качества продукции.

3) Интеллектуальный контроль качества продукции (рис. 4).

В настоящее время для процесса контроля качества продукции необходимо затрачивать много человеческих и материальных ресурсов. *DT* упрощает процесс контроля качества и, в сочетании с искусственным интеллектом, позволяет своевременно обнаруживать продукты с проблемами качества, предлагать решения в соответствии с классификацией продукта, анализом причин проблемы, а затем, предлагать меры по обеспечению требуемого качества производимой продукции.



Рис. 4. Пример схемы интеллектуального контроля качества с применением технологии *DT*

Заключение

В данной статье рассматривается практическое применение *DT* в области промышленного производства. В промышленном производстве *DT* имеют большие преимущества, которые заключаются в конструировании моделей, управлении производством и т.д. В частности, у *DT* есть естественное преимущество в сфере технического обслуживания оборудования и мониторинга производства в реальном времени. Цифровая трансформация – это единственный путь для будущего развития мировой экономики. В будущем, для широкого внедрения технологии *DT* в различные отрасли промышленности необходимо располагать точными и эффективными приложениями *DT*, чтобы завоевать доверие предприятий. Кроме того, они должны быть достаточно надежными, чтобы технология *DT* позволяла оцифровать предприятия производственного сектора.

Литература

1. Тетеркин М.А., Анисимов А.Р., Томильченко Л.Р. и др. Анализ сценариев использования технологии виртуализации / Инновации и инвестиции. 2022. № 5. С. 136-141. EDN VLYKPT.
2. Rivera L.F., Müller H.A., Villegas N.M., Tamura G., Jiménez M. On the Engineering of IoT-Intensive Digital Twins Software Systems. In: Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops (ICSEW'20). Association for Computing Machinery, New York, 2020, pp. 631-638.
3. Agalinos K., Ponis S.T., Aretoulaki E., Plakas G., Efthymiou O. (Discrete event simulation and Digital Twins: review and challenges for logistics. Procedia Manufacturing 51, 2020, pp. 1636-1641.
4. Assawaarayakul C., Srisawat W., Ayuthaya S.D.N., Wattanasirichaigoon S. Integrate Digital Twins to exist production system for industry 4.0. 2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON). 2019.
5. August U. Digital Twins Technology for More Efficiency. ATZ Prod Worldw 6:54. 2019.
6. Barat S. et al. An actor based simulation driven Digital Twins for analyzing complex business systems. 2019 Winter Simulation Conference (WSC). IEEE, 2019.
7. Fuller A. et al. Digital Twins: Enabling technologies, challenges and open research. IEEE Access 8: 108952–108971. 2020.
8. Гончаров А.С., Маклаков В.М. Цифровой двойник: обзор существующих решений и перспективы развития технологии. elibrary.ru, 2018 // Сборник трудов Всероссийской научно-практической конференции. Дата обращения: 28 апреля 2019.
9. Баксанский О.Е. Виртуальная реальность и виртуализация реальности // Концепция виртуальных миров и научное познание. СПб.: РХГИ, 2000. С. 292-305.
10. Барлоу Дж.П. Декларация независимости киберпространства. Перевод Ехарного 10.09.2004.

APPLICATION OF DIGITAL TWINS IN INDUSTRIAL PRODUCTION

Alexey I. Panov.

*Student of the IITM&M of UNSN of N.I. Lobachevsky,
Nizhny Novgorod, Russia
aleks.npc459@mail.ru*

Abstract

With the development of science and technology, the high-tech industry is developing rapidly, and various new technologies continue to appear, and digital twins are one of them. Being a completely new interactive technology, digital twin technology can handle the interaction between the real and virtual world well. They are integrated with other technologies and are used in many areas, such as smart factories in industrial production, a digital model of life in medicine, the construction of smart cities, security guarantees in the aerospace sector, shopping in the commercial sphere and so on. The purpose of this article is to provide an opportunity to understand the status of the use of digital twin technology. At the same time, the introduction of basic technologies related to digital twins is interspersed with the introduction of applications.

Keywords: *digital twins, virtual reality, industry, factory, production.*

«Цифровые технологии радиосвязи и телерадиовещания»

АНАЛИЗ ЭФФЕКТИВНОСТИ БЕЗРЕФЕРЕНСНЫХ МЕТРИК ПРИМЕНительно К ОЦЕНКЕ КАЧЕСТВА ВИДЕО ПРИ ПОТОКОВОЙ ПЕРЕДАЧЕ

Власюк Игорь Викторович,
доцент кафедры ТиЗВ МТУСИ,
к.т.н., доцент, Москва, Россия,
ru3dlp@yandex.ru

Киселева Александра Сергеевна,
магистрант МТУСИ, Москва, Россия,
alexandra.080600@gmail.com

Аннотация

Безреференсная оценка качества видеопоследовательностей является сложной задачей в настоящее время по причине того, что не все безреференсные метрики коррелируют с субъективной оценкой качества. В данной статье представлен обзор наиболее известных на данное время безреференсных метрик: BRISQUE, METAIQA, NIQE, PIQE и RANKIQA. В обзоре кратко описаны алгоритмы, используемые метриками. Целью данной работы является сравнительный анализ безреференсных метрик. Сравнение проводится по корреляции значений безреференсных метрик с субъективной оценкой, зависимости этих значений от разрешения видео, а также времени, затрачиваемого на анализ.

Ключевые слова: безреференсные метрики оценки качества видео, субъективные методы оценки качества видео, анализ качества видео.

Введение

Растущие тенденции обмена мультимедийным контентом и потоковым видео через социальные сети и видеохостинги вызвали необходимость распространения видео с высоким качеством восприятия. В большинстве таких случаев отсутствует доступ к оригиналу, следовательно, невозможно оценить качество полученного видео по отношению к эталонному. Поэтому существует острая необходимость в эффективном алгоритме безреференсной оценки качества видеоконтента.

За последние несколько десятилетий безреференсные метрики стали популярной темой для исследований в связи с распространением мультимедийного контента. Целью безреференсной оценки качества видео является предсказание качества восприятия цифрового видео без какой-либо информации об эталонном видео.

Процедура проведения оценки качества восприятия с помощью безреференсных метрик

Общее качество видеоконтента нельзя оценить только по нескольким параметрам, таким как яркость, контрастность или резкость. На четком изображении могут присутствовать элементы, которые могут быть восприняты алгоритмом как шум, что снижает качество. Для оценки качества независимо от типа искажений требуется стандартизированная процедура оценки. Результаты процедуры должны соответствовать человеческому восприятию качества изображения.

Целью проведения оценки качества восприятия (QoE) является измерение качества, которое соответствует восприятию зрительной системы человека (HVS) [1]. Эффективность любой модели лучше всего оценивать по ее корреляции с субъективными суждениями человека о качестве, поскольку человек является конечным получателем визуального сигнала. Наиболее очевидным способом оценки качества является субъективное исследование, когда для сбора статистических данных проводятся опросы людей-наблюдателей в лабораторных условиях [2] или краудсорсингового эксперимента [3]. В большинстве случаев оценка качества дается в виде усредненной оценки мнений (MOS), которая представляет собой среднее арифметическое всех мнений о качестве данных отдельными наблюдателями. Но эта процедура оценки качества может быть очень длительной и организационно сложной, что делает процедуру неэффективной при необходимости частого сбора данных. Однако не обязательно самим собирать эти данные, существуют общедоступные базы данных [15], например, такие как LIVE-NFLX-II [4] и AVT-VQDB-UHD-1 [5].

Объективные оценки направлены на построение вычислительных, математических моделей, которые позволяют оценить качество цифровых изображений, воспринимаемое человеком [13]. В их число входят безреференсные метрики, позволяющие оценить качество восприятия без необходимости использования эталонного видео.

Безреференсная метрика позволяет оценить качество видео с помощью алгоритма, в котором единственной входной информацией является искаженное видео, качество которого оценивается. Измерение такого показателя качества, как оценка содержания в видео шума или искажений, является сложной задачей при отсутствии эталонного видео, поэтому необходимо смоделировать показатели эталонного изображения, природу зрительной системы человека и влияние искажений на видео самостоятельным способом [14].

Метрика **Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)**

Метрика BRISQUE имеет очень низкую вычислительную сложность, что делает ее подходящей для приложений, работающих в реальном времени [8].

BRISQUE использует модель извлечения статистики естественной сцены (NSS) с локально нормализованными коэффициентами яркости и оценивает "естественность" с помощью параметров модели. Под естественной сценой понимается любое изображение, полученное с помощью оптической камеры без искусственной обработки. Параметры этой модели дополнительно количественно определяют естественность изображения [10].

Метрика BRISQUE ввиду последовательности алгоритма не может быть оптимизирована и распараллелена, и напрямую зависит от тактовой частоты процессора и наличия блока FPU. Скорость обработки видеопоследовательности прямо пропорциональна размеру кадра и количеству кадров. Сложность кадра не влияет на время, затрачиваемое на обработку.

Модель NSS

Распределение интенсивности пикселей естественных изображений отличается от распределения искаженных изображений. Эта разница в распределениях более заметна, когда нормализуется интенсивность пикселей и по ним вычисляется распределение. В частности, после нормализации интенсивность пикселей естественных изображений соответствует распределению Гаусса, в то время как интенсивность пикселей неестественных или искаженных изображений не подчиняется этому закону. Таким образом, отклонение распределения от идеальной колоколообразной кривой является мерой степени искажения изображения.

Нормализация коэффициентов MSCN

Существует несколько различных способов нормализации изображения. Один из таких способов называется нормализацией MSCN.

Хорошо известно, что естественные изображения обладают определенной закономерностью с точки зрения статистического распределения их интенсивности [6]. Эта закономерность изменяется под воздействием деградации изображения. Возможность измерения таких отклонений позволяет построить методы оценки качества изображения, способные идентифицировать эти ухудшения.

Прямым механизмом количественной оценки локальных нарушений в статистике интенсивности изображения является анализ распределения MSCN в окрестности заданного пикселя x изображения $I(x)$. Эти коэффициенты вычисляются по следующей формуле:

$$MSCN(x) = \frac{I(x) - \mu(x)}{\sigma(x) + \varepsilon} \quad (1)$$

где $\mu(x)$ – оценка среднего значения интенсивности изображения в окрестности x ; $\sigma(x)$ – оценка дисперсии интенсивности изображения в окрестности x ; ε – константа, которая используется для предотвращения возникновения неустойчивости, когда знаменатель стремится к нулю.

Коэффициенты MSCN обладают характерными статистическими свойствами, которые изменяются в присутствии искажений [7]. Моделирование этих коэффициентов позволяет предсказать тип и восприимчивость искажений. Каждое искажение изменяет статистику естественной сцены уникальным образом.

Чтобы количественно оценить статистику естественных изображений и то, как она меняется при искажении, используется асимметричное обобщенное распределение Гаусса (AGGD). Для каждого изображения оцениваются 2 параметра $(\alpha, (\sigma_l^2 + \sigma_r^2)/2)$ и формируется первый набор характеристик.

Хотя коэффициенты MSCN достаточно однородны, знаки коэффициентов имеют периодическую структуру. Чтобы оценить эту зависимость между соседними значениями, моделируется односторонняя связь через попарные произведения между соседними коэффициентами MSCN на расстоянии 1 пикселя по четырем ориентациям – горизонтальной, вертикальной, главной диагонали и побочной диагонали.

Метрика Perception based Image Quality Evaluator (PIQE)

Метрика PIQE оценивает степень искажений, присутствующих в изображении, исключительно на основе локальных характеристик изображения на уровне блоков. Метод основан на принципах человеческого восприятия искажений. Зрительное внимание человека направлено на значимые точки изображения или пространственно-активные области. Чтобы оценить, искажен блок или нет, ему присваивается соответствующее значение. Далее локальное качество на уровне блока складывается в общее качество изображения, воспринимаемое человеком.

На первом этапе из полутонового изображения извлекаются признаки NSS по формуле (1). Входное изображение сегментируется на непересекающиеся блоки B_k размером $n \times n$, исключая блоки находящихся на границах изображения со всех четырех сторон. Коэффициенты MSCN используются для маркировки данного блока как равномерного (U) или неравномерного (SA). Критерий для маркировки блока приведен ниже:

$$B_k = \begin{cases} U & v_k < T_U \\ SA & v_k \geq T_U \end{cases} \quad (2)$$

где v_k – дисперсия коэффициентов MSCN при $k \in 1, 2, \dots, N_B$; N_B – общее количество блоков размером $n \times n$.

В результате эмпирических наблюдений устанавливается порог T_U . Поскольку вычисляется дисперсия нормализованных коэффициентов, которая изменяется от 0 до 1, она рассматривается как процент. Эмпирические наблюдения также хорошо коррелируют с восприятием, следовательно, дисперсия коэффициентов MSCN будет низкой для блоков с низкой пространственной активностью.

Блоки B_k анализируются по двум типам критериев искажения:

1. критерий заметного искажения;
2. критерий аддитивного белого гауссовского шума.

Искаженному блоку присваивается значение искажения в зависимости от типа искажения, и определяются критерии для оценки того, подвержен ли блок искажениям или нет. Если блок искажен, то следующим шагом оценивается степень искажения, вносимого этим блоком в общую оценку.

Далее оценка качества изображения на основе восприятия определяется следующим образом:

$$PIQE = \frac{(\sum_{k=1}^{N_{SA}} D_{sk}) + \epsilon}{(N_{SA} + \epsilon)} \quad (3)$$

где N_{SA} – количество пространственно активных блоков в кадре; D_{sk} – коэффициент искажения для каждого блока; ϵ – константа, которая используется для предотвращения возникновения неустойчивости, когда знаменатель стремится к нулю (это возможно в случаях сильного размытия, когда изображения слишком сглажено).

Благодаря особенностям алгоритма PIQE, он может быть реализован с использованием параллельных вычислений. Однако этап перевода видео в оттенки серого и разбиение изображения на блоки алгоритма PIQE влияет на время обработки.

Метрика Natural Image Quality Evaluator (NIQE)

Алгоритм NIQE вычисляет оценку качества изображения с помощью заранее обученной модели.

Метрика отражает степень искажений в кадре и основана на построении группы статистических признаков на основе модели NSS в пространственной области [3]. Эти признаки получены из набора естественных, неискаженных изображений.

Характеристики моделируются как многомерные гауссовские распределения. В частности, значение NIQE можно рассматривать как степень отклонения от естественности цифрового изображения.

Метрика Meta-learning based Image Quality Assessment (MetaIQA)

Метрика MetaIQA оценивает качество видео на основе мета-обучения [9]. Мета-обучение – это алгоритм, позволяющий получить данные путем обучения на основе других уже существующих алгоритмов подобно человеку, который способен обучиться новой задаче базируясь лишь на уже имеющихся знаниях, косвенно связанных с ней.

Оценка качества метрикой MetaIQA состоит из двух этапов [9]:

1. Создание мета-обучающего набора данных на основе данных о специфике различных видов искажений. Для создания предварительной модели качества используется двухуровневый метод градиентного спуска;
2. Получение модели качества для неизвестных искажений на базе предварительной модели качества.

Метрика хорошо адаптируется к неизвестным искажениям благодаря своей предварительно обученной модели.

Метрика Rankings for No-reference Image Quality Assessment (RankIQA)

Метрика RankIQA оценивает степень искажений, присутствующих в изображении, на основе ранжирования [11]. Создаются искаженные изображения на основе эталонного, и после этого они ранжируются в соответствии с их качеством.

Имея набор эталонных изображений, можно исказить изображения для создания базы ранжированных изображений. Например, к эталонному изображению можно применить различные уровни гауссовского размытия. Полученный таким образом набор данных можно легко ранжировать, поскольку известно, что добавление гауссовского размытия ухудшает оценку качества. В наборе ранжированных изображений нет абсолютных оценок качества, но известно, какое из них более качественное.

Сравнительный анализ оценки видеопоследовательностей безреференсными метриками

В качестве тестовых видеопоследовательностей использовалась база данных качества видео LIVE-NFLX-II [4]. Для анализа были взяты следующие видеопоследовательности: «Eridian Conversation», «Skateboarding», «MeridianDriving» и «Sparks». После объединения видеопоследовательностей был получен видеофайл с разрешением 1920x1080, длительностью 2:14, количеством кадров 4037.



Рис. 1. Иллюстрации используемых видеопоследовательностей:
а) Eridian Conversation, б) Skateboarding, в) MeridianDriving, г) Sparks

Выше приведенная последовательность была загружена на видеохостинг «Youtube», предоставляющий пользователям услуги хранения, доставки и показа видео [12Ошибка! Источник ссылки не найден.], и выгружена с него на персональное компьютерное устройство для анализа качества видеопоследовательности метриками BRISQUE, MetaIQA, NIQE, PIQE и RankIQA.

С полученными данными была произведена нормировка значений. Нормирование переменных позволяет привести их в один диапазон для одинакового вклада в анализ. Для нормировки метрик в диапазон значений от 0 до 1 была использована следующая формула:

$$x_{\text{норм}} = \frac{x_i - x_{\text{мин}}}{x_{\text{макс}} - x_{\text{мин}}} \quad (4)$$

где $x_{\text{норм}}$ – i -е нормализованное значение в наборе данных; x_i – i -е значение в наборе данных; $x_{\text{макс}}$ – минимальное значение в наборе данных; $x_{\text{мин}}$ – максимальное значение в наборе данных.

Графики оценки качества видеопоследовательностей метриками приведены на рисунках 2-10 (на рисунках 2-5 отображены данные всех метрик по отдельно взятому разрешению, а на рисунках 6-10 – данные отдельно взятой метрики для 4 разрешений).

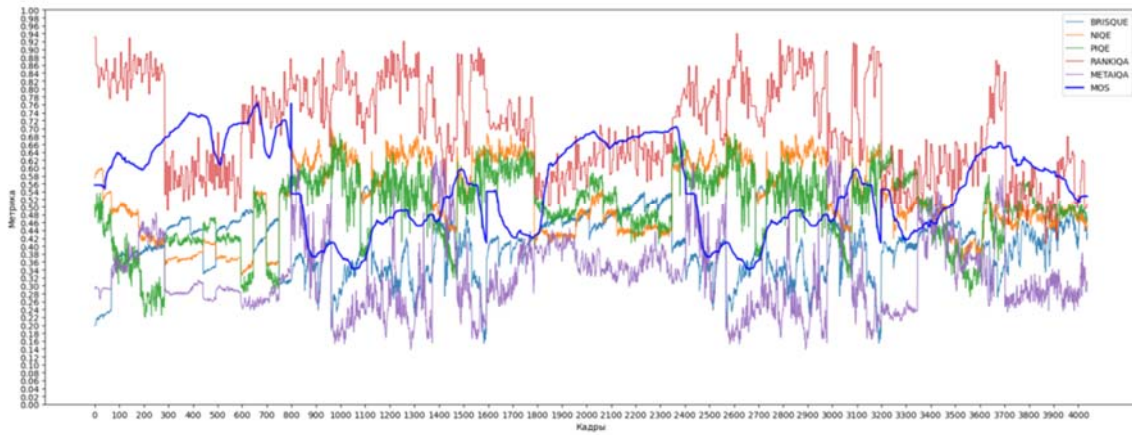


Рис. 2. График оценки качества оригинальной видеопоследовательности в разрешении 1920x1080 метриками BRISQUE, MetaIQ, NIQE, PIQE, RankIQ и MOS от LIVE-NFLX-II [4]

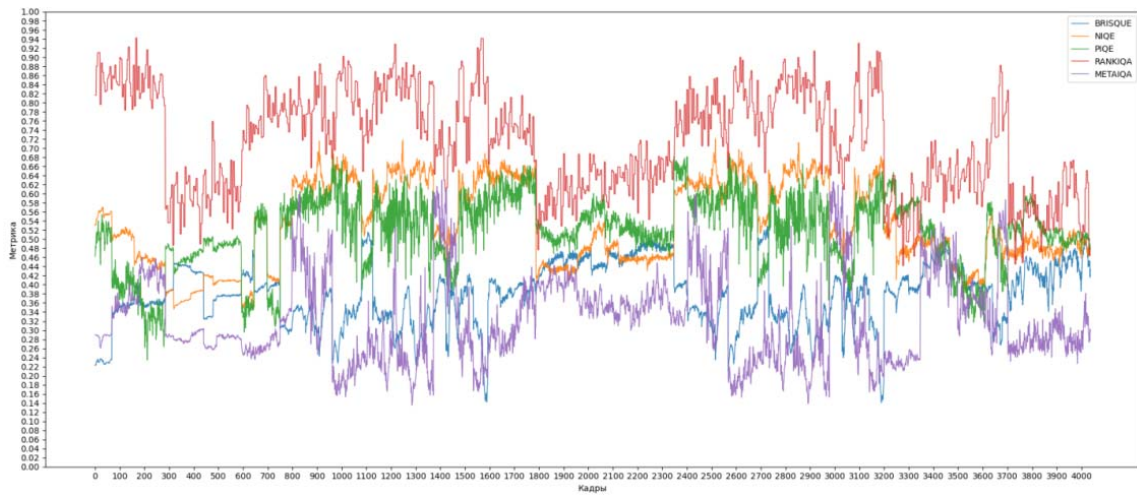


Рис. 3. График оценки качества видеопоследовательности, полученной из видеохостинга, в разрешении 1920x1080 метриками BRISQUE, MetaIQ, NIQE, PIQE, RankIQ

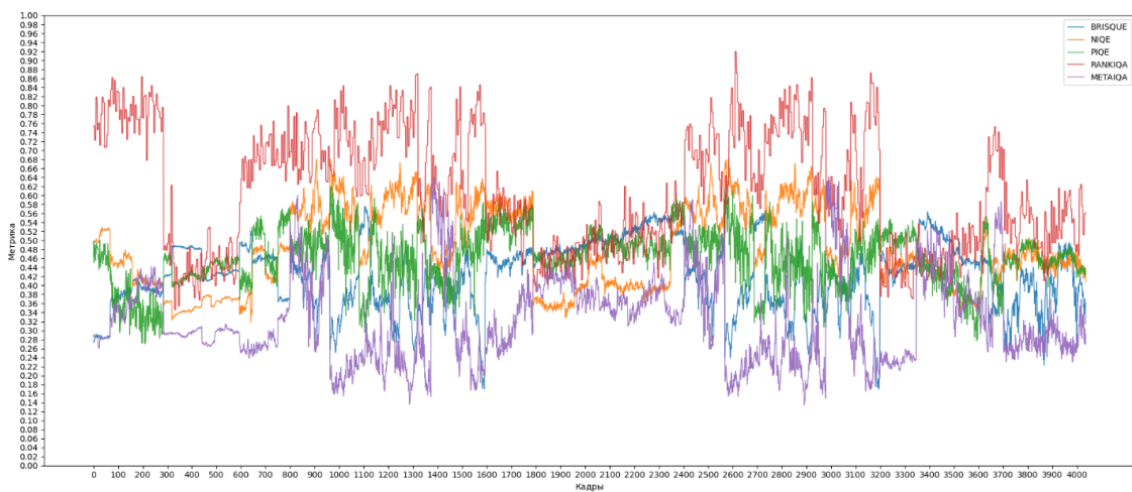


Рис. 4. График оценки качества видеопоследовательности, полученной из видеохостинга, в разрешении 1280x720 метриками BRISQUE, MetaIQ, NIQE, PIQE, RankIQ

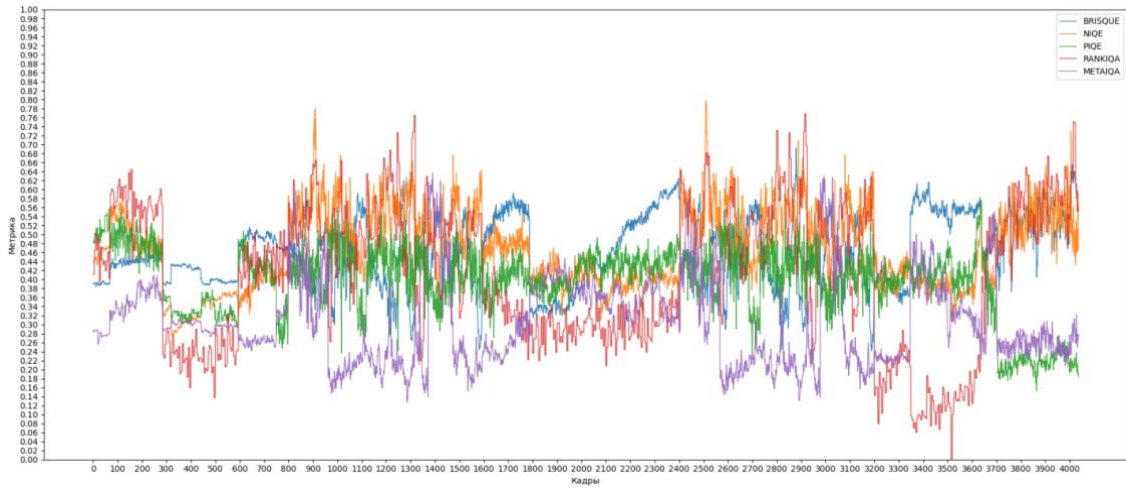


Рис. 5. График оценки качества видеопоследовательности, полученной из видеохостинга, в разрешении 640x360 метриками BRISQUE, MetaIQ, NIQE, PIQE, RankIQ

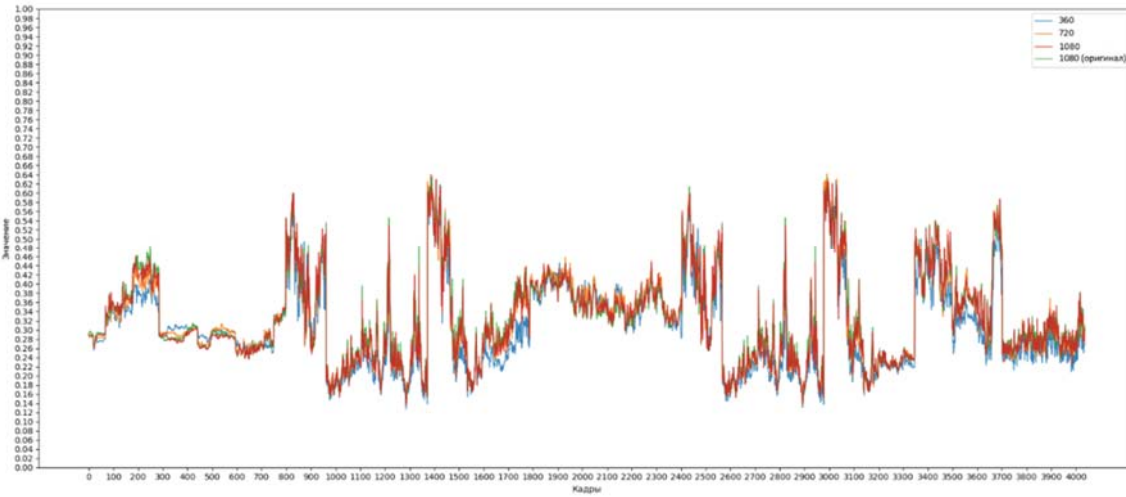


Рис. 6. График оценки качества видеопоследовательности с разрешениями 1920x1080, 1280x720, 640x360 метрикой MetaIQ

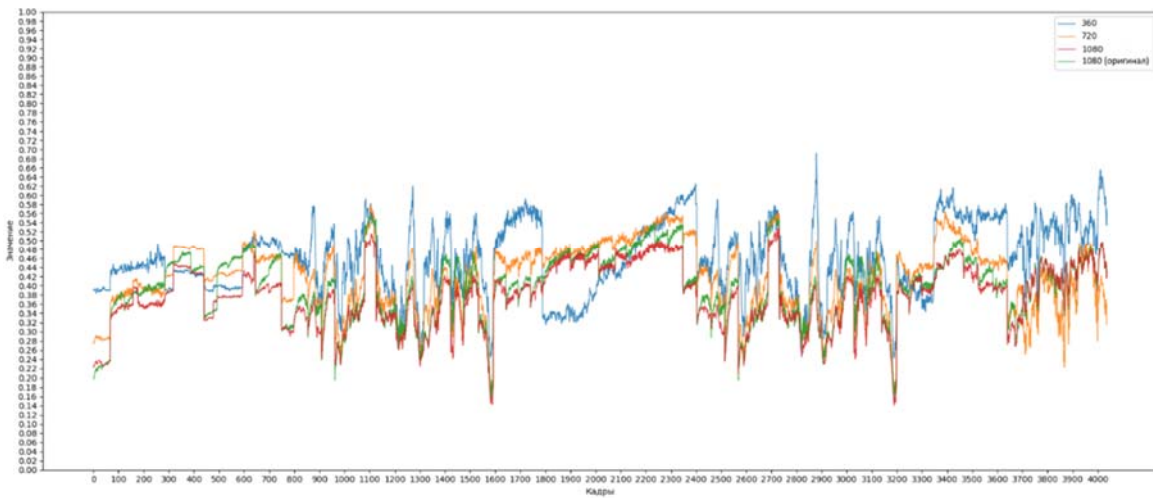


Рис. 7. График оценки качества видеопоследовательности с разрешениями 1920x1080, 1280x720, 640x360 метрикой BRISQUE

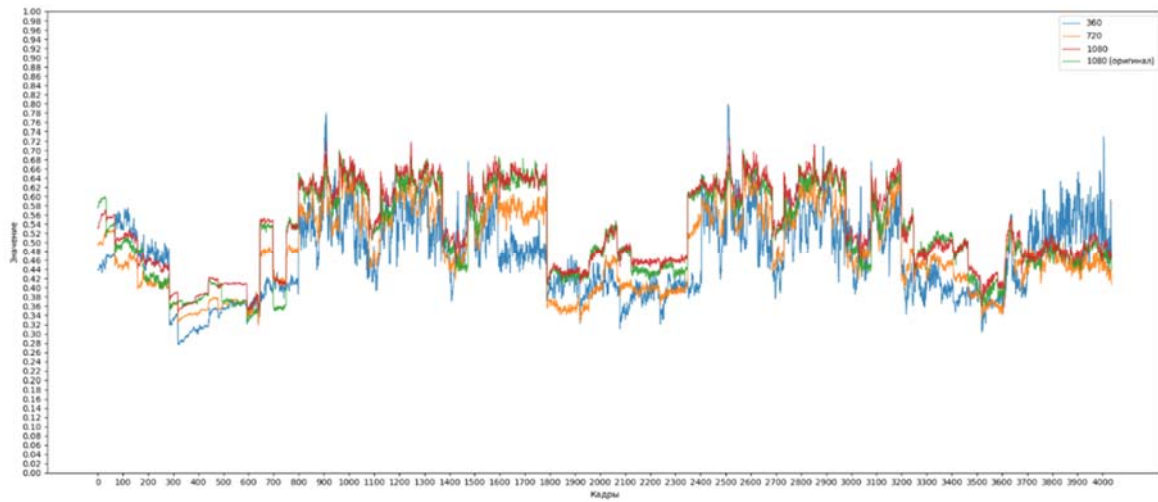


Рис. 8. График оценки качества видеопоследовательности с разрешениями 1920x1080, 1280x720, 640x360 метрикой NIQE

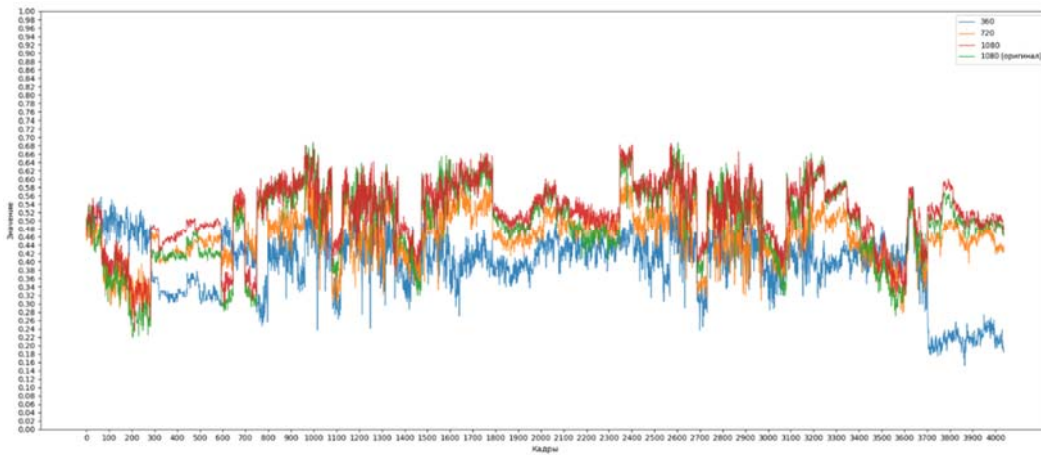


Рис. 9. График оценки качества видеопоследовательности с разрешениями 1920x1080, 1280x720, 640x360 метрикой RIQE

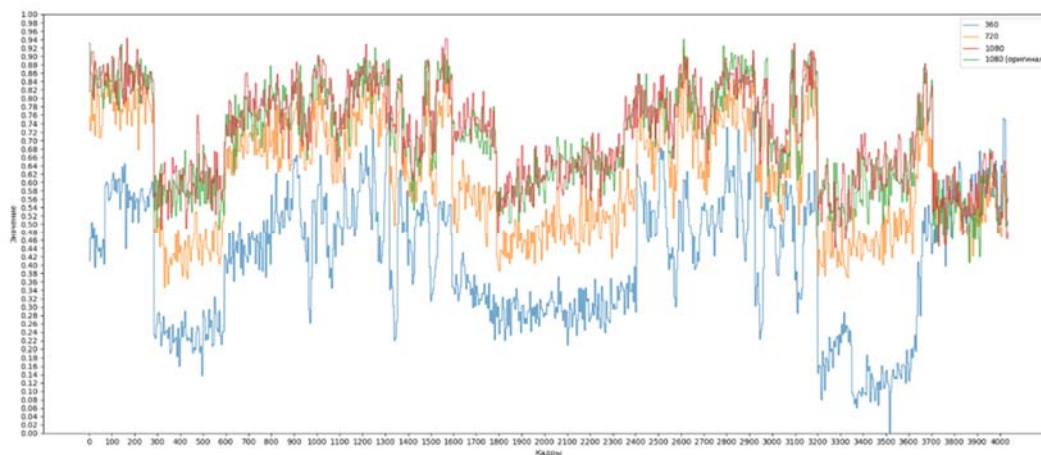


Рис. 10. График оценки качества видеопоследовательности с разрешениями 1920x1080, 1280x720, 640x360 метрикой RankIQa

Обработка результатов

1. Расчет коэффициента корреляции значений метрик с субъективной оценкой качества видео

Использование коэффициента корреляции Пирсона является одним из способов количественной оценки взаимосвязи между двумя переменными. Коэффициент корреляции Пирсона всегда принимает значение от -1 до 1, где:

- -1 указывает на отрицательную линейную корреляцию между двумя переменными;
- 0 указывает на отсутствие линейной корреляции между двумя переменными;
- 1 указывает на положительную линейную корреляцию между двумя переменными.

Коэффициент корреляции рассчитывается по формуле:

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \cdot \sigma_y} \tag{5}$$

где σ – среднеквадратическое отклонение.

Коэффициенты корреляции MOS с метриками BRISQUE, MetaIQA, NIQE, PIQE и RankIQA:

- $r_{\text{brisque}} = 0,38;$
- $r_{\text{niqe}} = -0,68;$
- $r_{\text{piqe}} = -0,45;$
- $r_{\text{rankiqa}} = -0,32;$
- $r_{\text{metaiqa}} = 0,11.$

Между переменными метрик NIQE и PIQE с MOS наблюдается наибольшая корреляция, а у MetaIQA она практически отсутствует.

2. Оценка времени обработки видео метриками.

В режиме реального времени является недопустимым использование метрик с большим временем работы, так как она сильно замедляют процесс анализа видеопоследовательности. Ниже на рисунке 11 представлен график зависимости анализируемого разрешения метриками от времени.

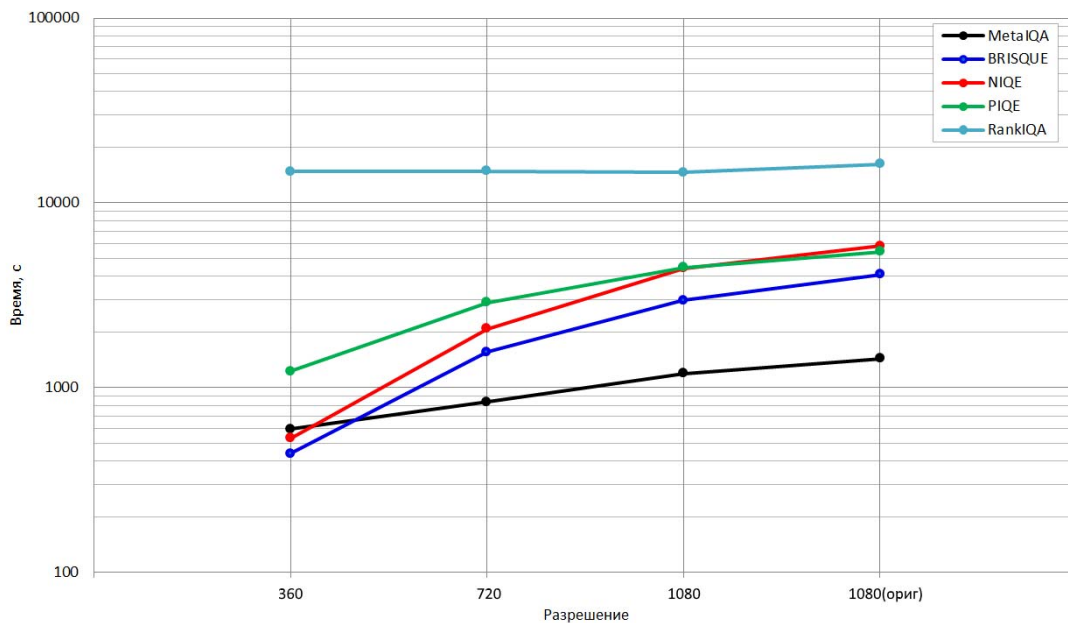


Рис. 11. График зависимости анализируемого разрешения метриками от времени

На рисунке 11 видно, что MetaIQA имеет самую высокую скорость обработки видео, и время работы значительно меньше зависит от качества видеопоследовательности в сравнении с другими метриками.

На оценку качества видеопоследовательности больше всего времени требуется у RankIQa, что недопустимо для использования в режиме реального времени. У RankIQa файл весов предварительно обученной нейронной сети использует большой объем дискового пространства, и время работы алгоритма не зависит от количества кадров обрабатываемого видео и их разрешения. Таким образом, алгоритм эффективен при использовании современных видеоускорителей (GPU), ориентированных на распределенные вычисления и выбора для обработки длинных видеопоследовательностей в высоком качестве.

Размер файла предварительно обученной модели алгоритма MetaIQa составляет десятки килобайт данных, что незначительно по сравнению с объемами современных накопителей, однако объем весов нейронной сети для алгоритма RankIQa составляет несколько гигабайтов [11], что накладывает ограничения на выбор используемой аппаратуры при проведении анализа.

Метрики BRISQUE, NIQE и PIQE имеют схожую скорость обработки, пропорционально зависящую от качества используемой видеопоследовательности.

3. Зависимость средней оценки и среднеквадратического отклонения от разрешения видеопоследовательности

Следующим критерием сравнения метрик является мера изменения их значений в зависимости от разрешения видеопоследовательности. На рисунке 12 представлен график зависимости среднего значения метрик от разрешения видеопоследовательности, а в таблице 1 их среднеквадратическое отклонение.

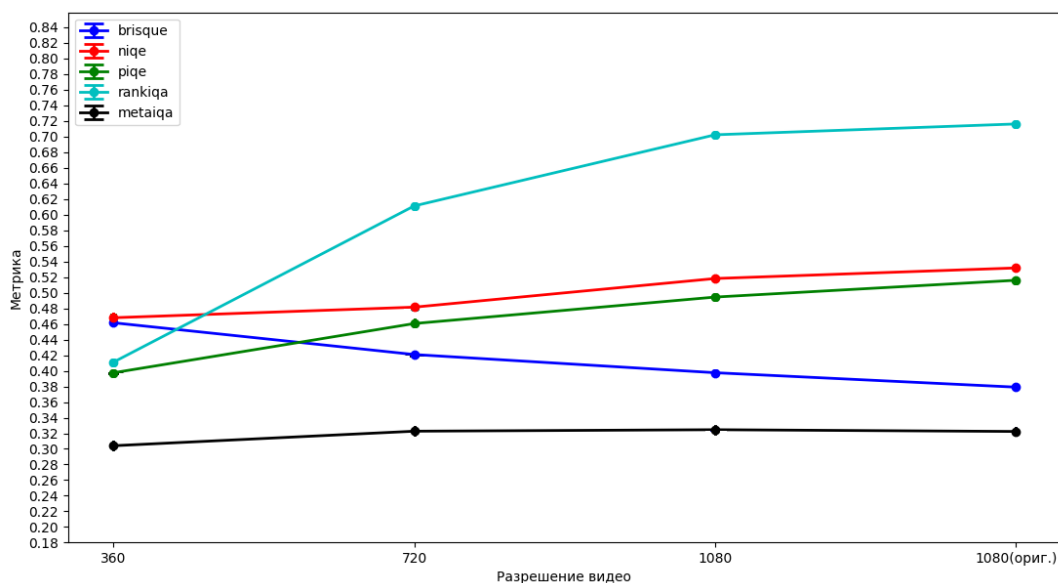


Рис. 12. График зависимости среднего значения метрик от разрешения видеопоследовательности

По рисунку 12 и таблице 1 можно сделать вывод, что наибольшая зависимость от качества видеопоследовательности наблюдается у метрик BRISQUE, PIQE

Таблица 1

Среднеквадратическое отклонение значений метрик по разрешениям видеопоследовательности

Разрешение видео	Среднеквадратическое отклонение для метрик				
	BRISQUE	NIQE	PIQE	MetaIQa	RankIQa
640x360	0,078	0,086	0,077	0,093	0,155
1280x720	0,073	0,088	0,06	0,096	0,133
1920x1080	0,072	0,093	0,09	0,095	0,118
1920x1080 (оригинал)	0,068	0,091	0,077	0,095	0,112

Заключение

В данной работе рассмотрены такие безреференсные метрики, как BRISQUE, MetaIQA, NIQE, PIQE и RankIQA. Был произведен сравнительный анализ указанных алгоритмов, по результатам которого можно сделать вывод, что наиболее подходящей метрикой для оценки качества видео в режиме реального времени является PIQE, так как она показала наилучшие результаты по времени работы, по корреляции значений с MOS и по зависимости среднего значения и среднеквадратического отклонения от разрешения. Метрики BRISQUE и NIQE также показали хорошие результаты, однако у BRISQUE низкая корреляция с данными MOS, а у NIQE практически не наблюдается зависимость от изменений разрешения видеопоследовательности. Метрики MetaIQA и RankIQA меньше всего подходят для использования, так как их корреляция с MOS близка к 0, время работы алгоритма RankIQA немного превышает остальные метрики, а у MetaIQA с повышением разрешения видео значения практически не меняются.

Литература

1. Keelan B. Handbook of Image Quality: Characterization and Prediction, CRC Press: Boca Raton, FL, USA, 2002, pp.143-147.
2. Chubarau, A., Akhavan, T., Yoo H., Mantiuk R.K., Clark, J. Perceptual image quality assessment for various viewing conditions and display systems, Electron. Imaging 2020, 2020.
3. Saupe D., Hahn F., Hosu V., Zingman I., Rana M. Crowd workers proven useful: A comparative study of subjective video quality assessment, In Proceedings of the QoMEX 2016: 8th International Conference on Quality of Multimedia Experience, Lisbon, Portugal, 2016, pp. 1-2.
4. Bampis C.G., Li Z., Katsavounidis I., Huang T., Ekanadham C., Bovik A.C. Towards Perceptually Optimized End-to-end Adaptive Video Streaming, submitted to IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2018, pp. 2-5.
5. Rao R. R. R., Goring S., Robitza W., Feiten B., Raake A. AVT-VQDB-UHD-1: A Large Scale Video Quality Database for UHD-1, IEEE ISM 2019, pp. 1-8.
6. Geisler W. S. Visual Perception and the Statistical Properties of Natural Scenes, Annual Review of Psychology 59(1), 2008, pp. 167-192.
7. Moorthy A. K., Bovik A. C. Statistics of natural image distortions, In International Conference on Acoustics Speech and Signal Processing, pp. 962-965.
8. Mittal A., Moorthy A. K., Bovik A. C. Blind/Referenceless Image Spatial Quality Evaluator, 11th Asilomar Conference, 2011, pp. 723-727.
9. Vanschoren J. Meta-learning: A survey, arXiv preprint arXiv:1810.03548, 2018, pp. 1-3.
10. Hancheng Z., Leida L., Jinjian W., Weisheng D., Guangming S. MetaIQA: Deep Meta-Learning for No-Reference Image Quality Assessment, 2020 IEEE/CVF Conference on CVPR.
11. Xialei L., Joost W., Andrew D. B. RankIQA: Learning from Rankings for No-reference Image Quality Assessment, IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1040-1042.
12. Как работает YouTube: функции, принципы и достижения [Электронный ресурс]. URL: <https://www.youtube.com/howyoutubeworks/> (Дата обращения: 29.09.2022).
13. Mittal A., Soundararajan R., Bovik A. C. Making a 'Completely Blind' Image Quality Analyzer, 2013.
14. Mozhaeva A. I., Vlasuyk I. V., Potashnikov A. M., Cree M. J., Streeter L. The Method and Devices for Research the Parameters of the Human Visual System to Video Quality Assessment, 2021 Systems of Signals Generating and Processing in the Field of on Board Communications, 2021, pp. 50.
15. Mozhaeva A. I., Potashnikov A. M., Vlasuyk I. V., Streeter L. Constant Subjective Quality Database: The Research and Device of Generating Video Sequences of Constant Quality, 2021 International Conference on Engineering Management of Communication and Technology, 2021.

ANALYSIS OF THE EFFECTIVENESS OF NO-REFERENCE METRICS FOR ASSESSING THE QUALITY OF VIDEO STREAMING

Igor V. Vlasuyk,

Associate Professor of the Department of T&SB MTUCI, PhD., Moscow, Russia,
ru3dlp@yandex.ru

Kiseleva S. Aleksandra,

Graduate MTUCI, Moscow, Russia,
alexandra.080600@gmail.com

Abstract

At present the no-reference video quality estimation is a difficult task due to the fact that not all no-reference metrics correlate with subjective quality estimation. In this article presents an overview of the currently most well-known no-reference metrics: BRISQUE, METAIQA, NIQE, PIQE u RANKIQA. The review summarizes the algorithms used by metrics. The aim of this work is a comparative analysis of the no-reference metrics. The comparison is based on the correlation of no-reference metric values with subjective assessment, the dependence of these values on video resolution, and the time spent on analysis.

Keywords: *no-reference metrics of video quality estimation, subjective methods of video quality estimation, video quality analysis.*

ВЛИЯНИЕ ВОССТАНОВИТЕЛЬНОЙ КОРРЕКЦИИ НА ПАРАМЕТРЫ РЕЧЕВОГО СИГНАЛА

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

Василевский Кирилл Антонович,
магистрант МТУСИ, Москва, Россия,
alaxtver@yandex.ru

Гаранин Тимофей Дмитриевич,
магистрант МТУСИ, Москва, Россия,
veneriec@gmail.com

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

Цыплаков Дмитрий Иванович,
магистрант МТУСИ, Москва, Россия,
tsyplakov-di@mail.ru

Аннотация

В статье рассматривается объективное влияние дополнительной коррекции на параметры речевого сигнала. Произведена оценка статистических (энергетических и спектральных) параметров речевого сигнала с дополнительной обработкой на основе алгоритма CELP выполненной до и после кодека. Обоснован выбор этого алгоритма и описаны его преимущества в сравнении с другими алгоритмами. Произведена оценка как положительного, так и отрицательного влияния на качество речевого сигнала аудиопроцессорной обработки до и после кодека

Ключевые слова: кодирование с линейным предсказанием, алгоритм CELP, ОСМ, динамическая обработка аудиопроцессором, спектральные характеристики.

Введение

Исследования параметров речевого сигнала (РС), определяющих его возможность передачи смысловой (разборчивость) и эмоциональной информации показали наличие большой избыточности, как статистической (избыточности описания), так и психофизической, определяемой восприятием сигнала с помощью слухового анализатора человека. Известно, например [1], что сохранение в РС только участков нестационарности («атак»), составляющих около 15% его длительности, не влияет на передачу смысловой информации, хотя качество звучания ухудшается. При представлении РС в виде произведения аналитической огибающей на косинус мгновенной фазы, замена огибающей на постоянное число не приводит к уменьшению смысловой информативности [1]. В зависимости от условий эксплуатации (характеристик канала передачи) скорость передачи РС может составлять от 64кБ/с до 600 бит/с. Наиболее часто используются алгоритмы на основе адаптивной дифференциальной импульсно-кодовой модуляции (АДИКМ), линейное прогнозирующее кодирование (LPC) и линейное предсказание с мультикодовым управлением (CELP).

Все перечисленные алгоритмы обеспечивают достаточно высокое качество передачи, но при этом не сохраняют форму сигнала. В тоже время методики контроля канала передачи, как правило, контролируют именно форму гармонического стационарного сигнала. Возникает задача определения параметров РС, которые позволяют прогнозировать субъективную оценку качества передачи.

К наиболее информативным параметрам РС можно отнести относительную среднюю мощность (ОСМ). Различают два вида ОСМ, в зависимости от способа нормировки: канальную и сигнальную. Канальная ОСМк – это отношение мощности реального сигнала к мощности гармонического с амплитудой равной номинальной для канала, а сигнальная ОСМс – к мощности гармонического сигнала с амплитудой равной максимальному значению на каждой выборке.

Канальная ОСМк, практически, оценивает громкость РС, соответственно канальная дифференциальная ОСМк.д – перепады громкости, определяющие эмоциональную информативность сигнала. Сигнальная ОСМс является очень чувствительным индикатором наличия искажений в канале передачи [2]. В работе исследуются объективные изменения РС

Главная идея исследования заключается в оценке изменения энергетических и спектральных параметров речевого сигнала при предварительной его обработке, а также после «прохождения» сигнала через выбранный нами кодек (CELP), с дальнейшей коррекцией РС.

Сравнительный анализ экспериментально полученных статистических характеристик интервальных величин ОСМ

Наибольшие трудности при исследованиях вызывает формирование базы высококачественных РС. Поэтому в качестве эталонного РС были использованы речевые дикторские фрагменты программы радиостанции «Маяк» из библиотеки программ кафедры ТиЗВ МТУСИ. Программы были записаны на выходе тракта формирования программ со студийным качеством, а затем их качество было снижено до «телефонного». Были проведены сравнительные исследования изменения статистических параметров РС при введении коррекции до и после кодирования. В список параметров для сравнения входят: относительная средняя мощность, спектральные характеристики РС без коррекции и при обработке сигнала исследуемым алгоритмом на аудиопроцессоре.

Для расчётов характеристик исходного и обработанного аудиосигналов использовалось программные продукты, обеспечивающие обработку и сжатия аудиоданных, а именно ПО NewEstim и Adobe Audition [3,4].

Предварительно, для имитации прохождения радиосигнала по каналу вещания отрывок аудиофайла подвергался обработке полосовым фильтром с полосой от 300 до 3400 Гц. Далее производилось оценивание изменения параметров РС. Оценке подверглись спектральные параметры сигнала и ОСМ. После прохождения аудиосигналом процедуры сжатия посредством речевых кодеков так же была произведена оценка его параметров и получены характеристики сжатого вещательного сигнала. Затем была проведена предварительная обработка исходного сигнала с использованием исследуемого алгоритма при помощи корректора звукового сигнала (КЗС) [4]. Особенностью алгоритма является обработка аналитической огибающей РС, что резко повышает его эффективность. Произведены сопоставление и сравнение полученных данных, для дальнейшей оценки эффективности работы алгоритма.

В качестве кодеков, применяемых для обработки исходного вещательного РС, были использованы кодеки с алгоритмом CELP (со скоростями 16, 8 и 4.8 кбит/с).

Следует отметить, что при кодировании с линейным предсказанием (CELP) происходит моделирование некоторых параметров речи, и при передаче закодированного сигнала эти параметры передаются вместо некоторых отсчётов уровней РС или их разностей. Это позволяет использовать канал с гораздо меньшей пропускной способностью. Кодек, в свою очередь, работает с блоками отсчётов в отличие от АДКМ кодирования, при котором работа происходит с каждым отдельными отсчётами. Каждый блок, характеризуется частотой основного тона и амплитудой, а также некоторыми второстепенными параметрами. Такой метод сжатия затрудняет разработку алгоритмов эффективной обработки аудиосигналов. В то же время без эффективного сжатия РС потребуются слишком большие полосы пропускания каналов передачи, что в современных условиях является критичным фактором. Задача разработки алгоритма обработки РС, эффективность которого мало зависит, от дальнейшего кодирования сигнала является одной из основных задач многих исследований в этой области.

Для оценки эффективности обработки, надо чётко сопоставить все параметры сигналов разных типов. Так же следует оценить эффективность разных сочетаний алгоритмов обработки.

На рисунках 1-4 приведены полученные экспериментально статистические параметры: интервальных сигнальной ОСМс; сигнальной дифференциальной ОСМс.д; канальной ОСМк и канальной дифференциальной ОСМк.д для РС при разных скоростях его сжатия (кбит/с) без предварительной обработки аудиопроцессором [3]. На рисунках 1-4 отображены гистограммы средних значений, среднего квадратического отклонения (СКО), медианных значений и интегральных абсолютных отклонений статистик (ИАОС) речевого сигнала для соответствующих режимов его сжатия.

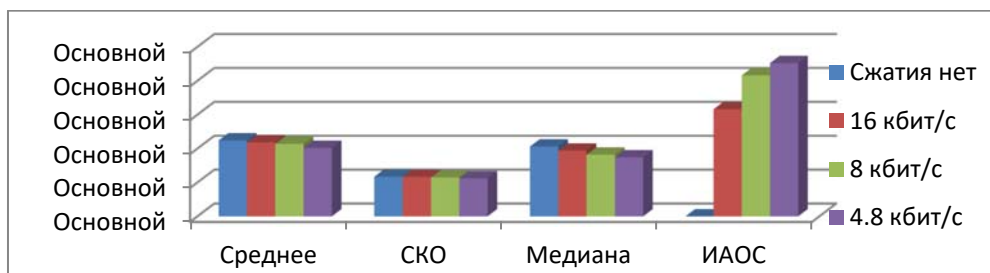


Рис. 1. Интервальные значения ОСМс РС без обработки аудиопроцессором

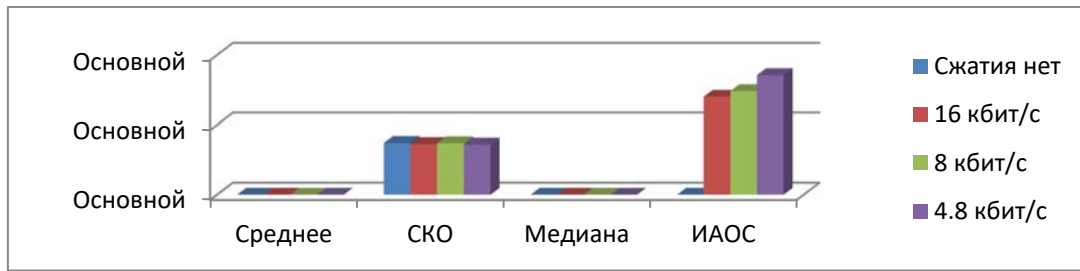


Рис. 2. Интервальные значения ОСМс.д РС без обработки на аудиопроцессоре

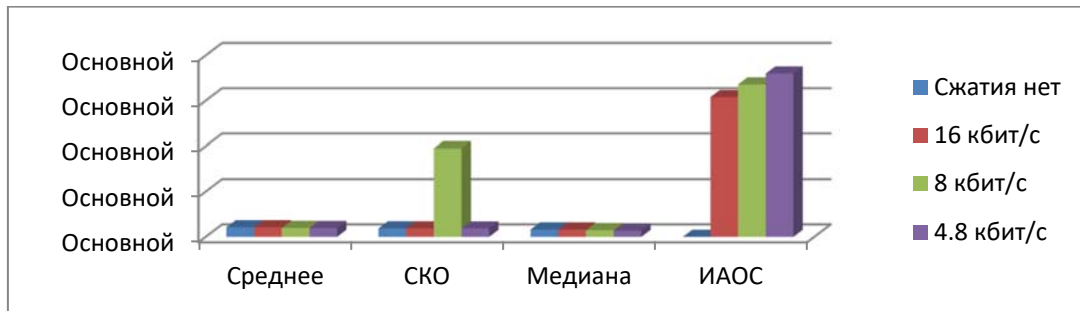


Рис. 3. Интервальные значения ОСМк РС без обработки на аудиопроцессоре

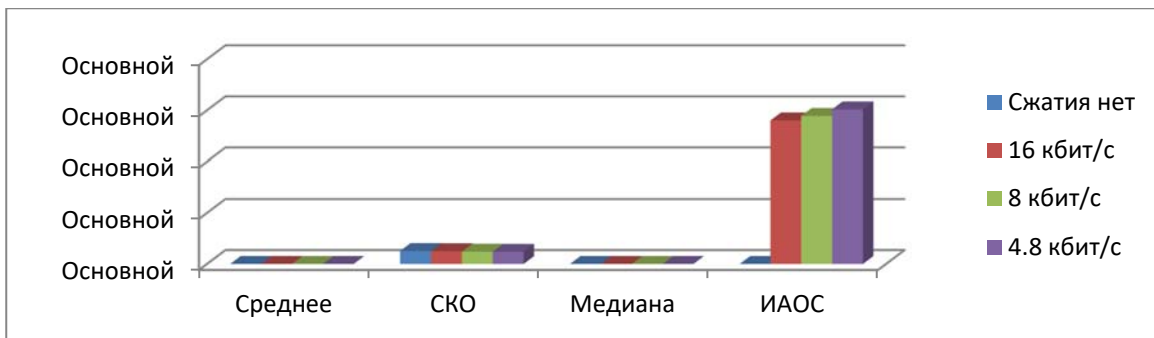


Рис. 4. Интервальные значения ОСМк.д РС без обработки на аудиопроцессоре

Соответственно на рисунках 5-8 приведены гистограммы значений статистических параметров интервальных ОСМ РС при его обработке аудиопроцессором до кодека

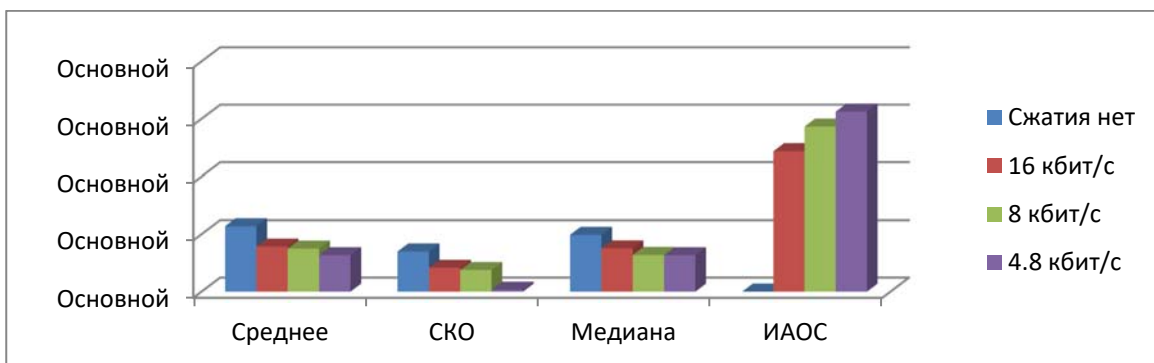


Рис. 5. Интервальные значения ОСМс РС при обработке аудиопроцессором до кодека

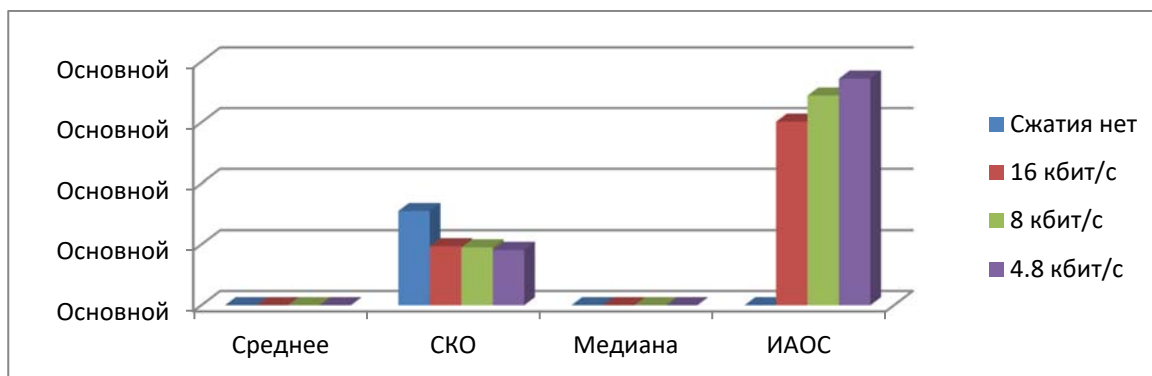


Рис. 6. Интервальные значения ОСМс.д РС при обработке аудиопроцессором до кодека

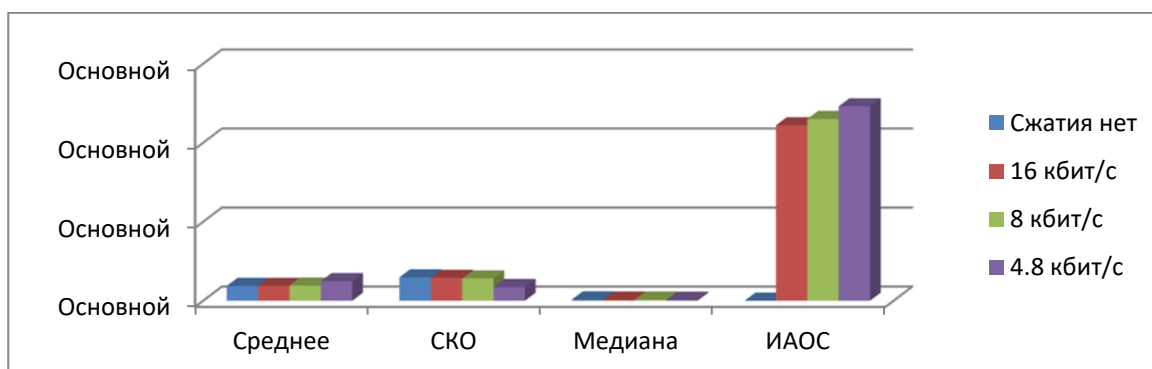


Рис.7. Интервальные значения ОСМк РС при обработке аудиопроцессором до кодека

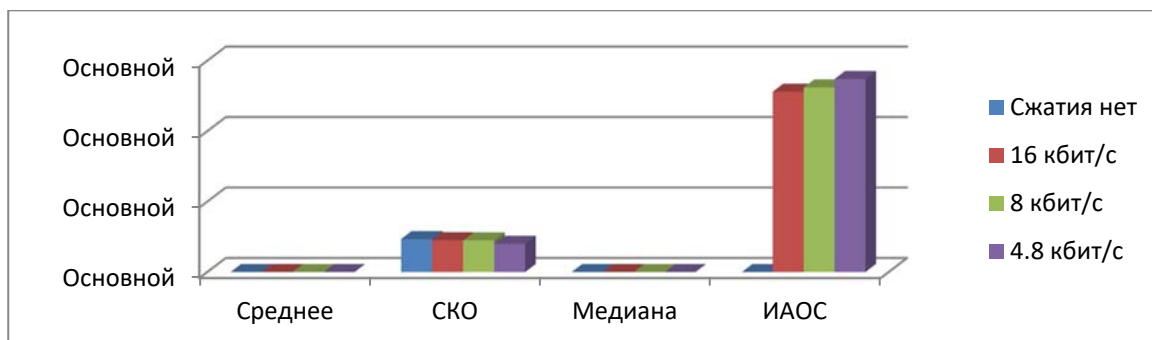


Рис. 8. Интервальные значения ОСМк.д РС при обработке аудиопроцессором до кодека

Гистограммы значений статистических параметров интервальных ОСМ РС при его обработке аудиопроцессором после кодека приведены на рисунках 9-12

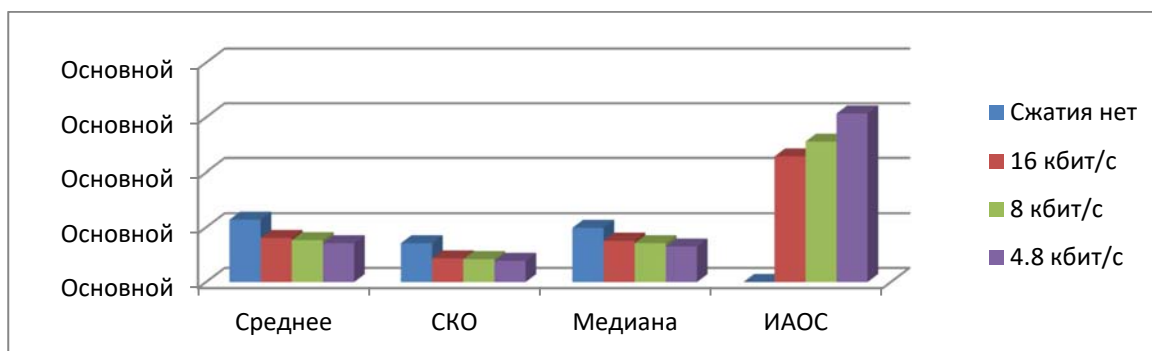


Рис. 9. Интервальные значения ОСМс РС при обработке аудиопроцессором после кодека

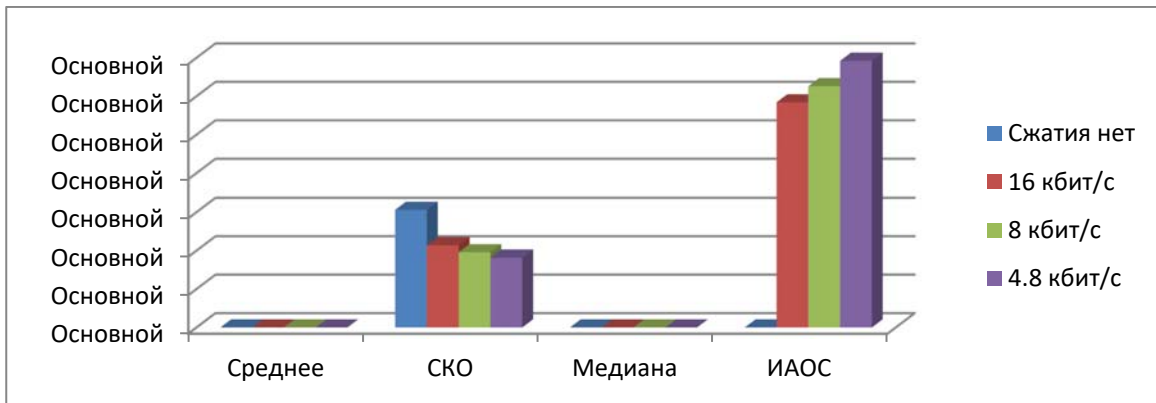


Рис. 10. Интервальные значения ОСМс.д РС при обработке аудиопроцессором после кодека

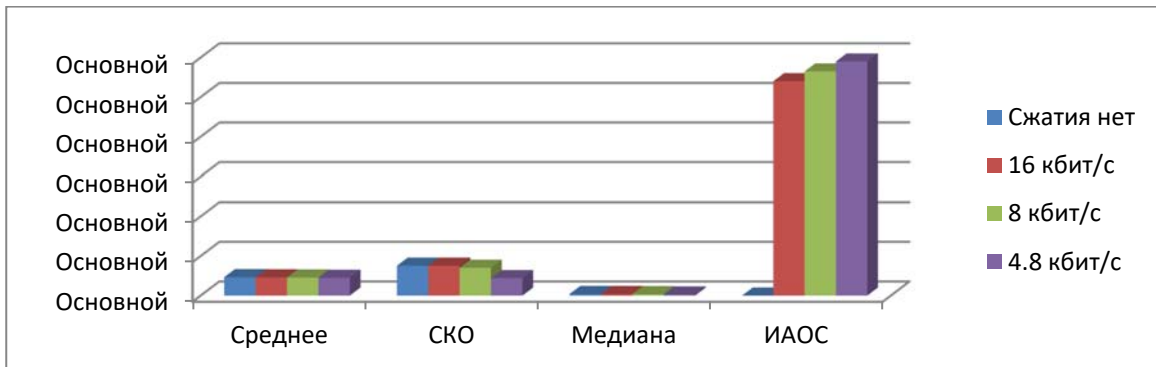


Рис. 11. Интервальные значения ОСМк PC при обработке аудиопроцессором после кодека

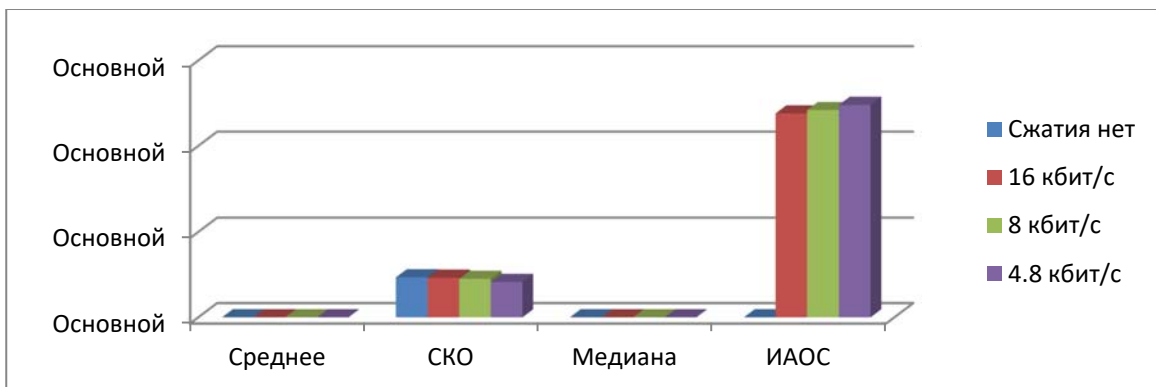


Рис. 12. Интервальные значения ОСМк.д PC при обработке аудиопроцессором после кодека

На рисунках 13-16 приведены гистограммы значений статистических параметров интервальных ОСМ РС при его обработке аудиопроцессором до и после кодека.

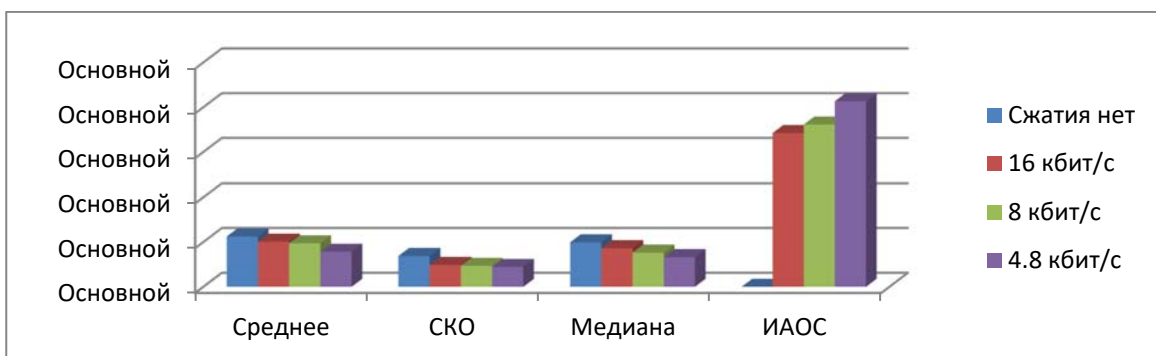


Рис. 13. Интервальные значения ОСМс PC при обработке аудиопроцессором до и после кодека

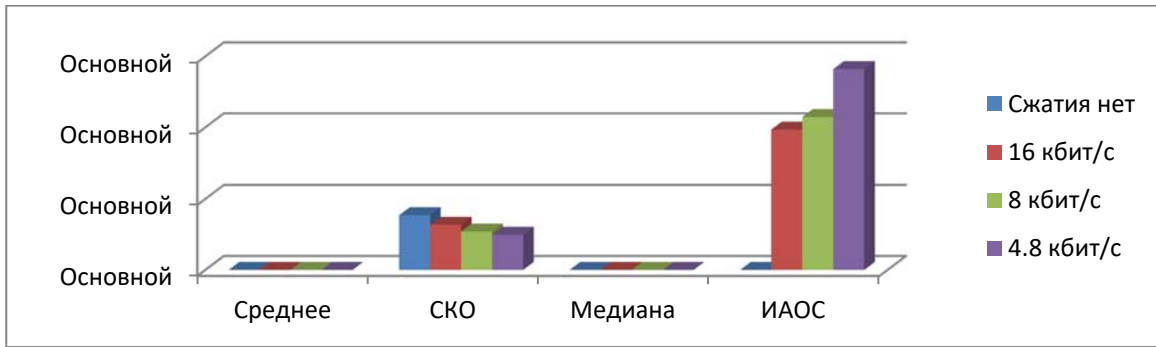


Рис. 14. Интервальные значения ОСМс.д РС при обработке аудиопроцессором до и после кодека

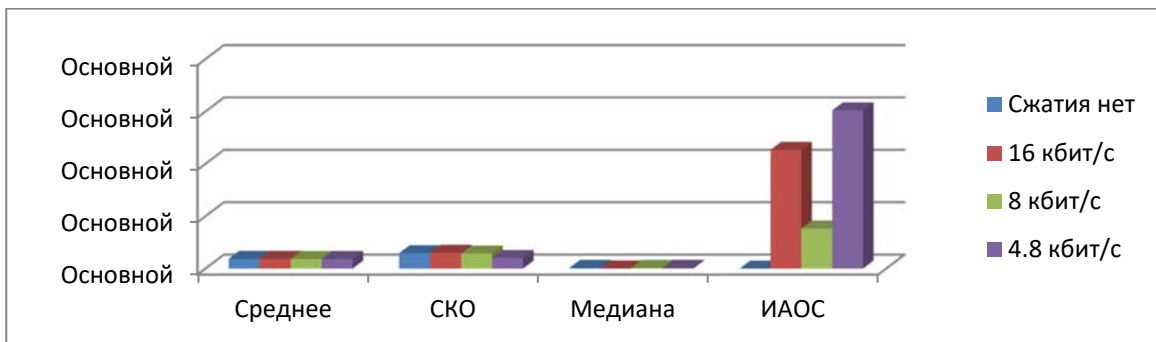


Рис. 15. Интервальные значения ОСМк РС при обработке аудиопроцессором до и после кодека

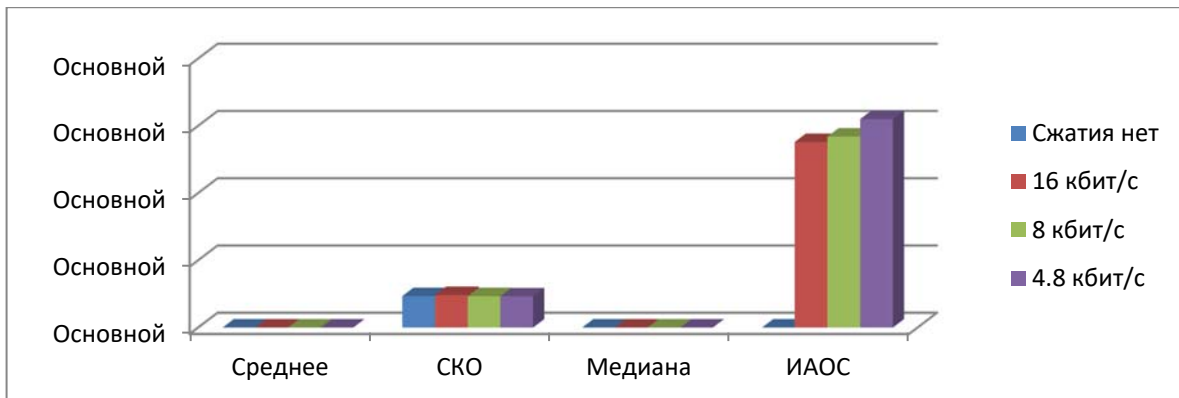


Рис. 16. Интервальные значения ОСМк.д РС при обработке аудиопроцессором до и после кодека

Исходя из полученных статистических данных (рис. 1-16), можно сформулировать некоторые предварительные выводы. В частности, среди характеристик для канальных интервальных ОСМ, видна чёткая тенденция: чем больше мы увеличиваем ОСМк на входе канала, тем больше она снижается в процессе кодирования. Правда, эти потери ОСМк можно компенсировать, проводя восстановительную коррекцию на выходе канала, но в некоторых случаях это невозможно, например, при циркулярной передаче информации с обработкой только на входе канала.

Использование алгоритма коррекции с представлением РС в виде аналитической огибающей и мгновенной фазы, позволяет увеличивать ОСМк в 2-2,2 раза в отличие от увеличений ОСМк всего на 30-40%, характерных для традиционных канальных авторегуляторов [6].

Громкостные соотношения в сигнале оцениваются изменением СКО (среднее квадратическое отклонение) ОСМк, в нашем случае это интервальные дифференциальные канальные ОСМк.д, гистограммы которых (рис. 4,8, 12 и 16) также позволяют сделать некоторые выводы. Из результатов измерений следует, что не обработанные сигналы подвергаются наименьшей деградации, то есть во время сжатия предварительная обработка сигналов влечёт большую деградацию РС [7].

Классическая обработка сигналов, также подвержена негативным влияниям сжатия в прямой зависимости от скорости сжатия РС. Несмотря на то, что классическая обработка меньше «разрушается» во время сжатия, сигнал на выходе всё равно имеет худшие характеристики, по сравнению с исходным. В зависимости от

конкретных характеристик канала, эффективность той или иной обработки будет меняться [5]. В нашем случае рассматриваются объективные характеристики, применимые в большинстве случаев.

Оценка изменений ОСМс и ОСМс.д в канале передачи показывает их незначительность, что подтверждается широким использованием таких кодеков в типовых каналах связи.

Рассмотрим изменение спектральных параметров РС в результате его обработки с использованием аудио-процессора (ЗКС) с целью устранения избыточности сигнала [6].

В процессе устранения психофизической избыточности часть спектра РС «отбрасывается» (не передается), так как не будет воспринята слушателем вследствие частотной и временной маскировки сигнала. В качестве меры оценки искажений спектра используется величина отклонения в процентах мгновенных АЧХ от исходных. При этом среднестатистическая тенденция, определяющая интегральное отклонение (ИО), сохраняется близкой к исходной АЧХ.

Результаты оценки искажений спектра при различных вариантах введении корректоров звуковых сигналов (ЗКС) приведены на рисунках 17-20.

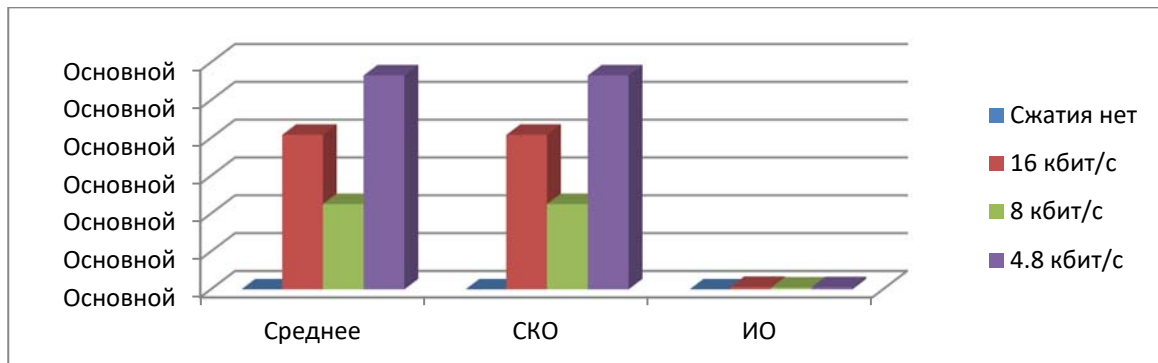


Рис. 17. Сигнал, не подвергнутый дополнительной обработки

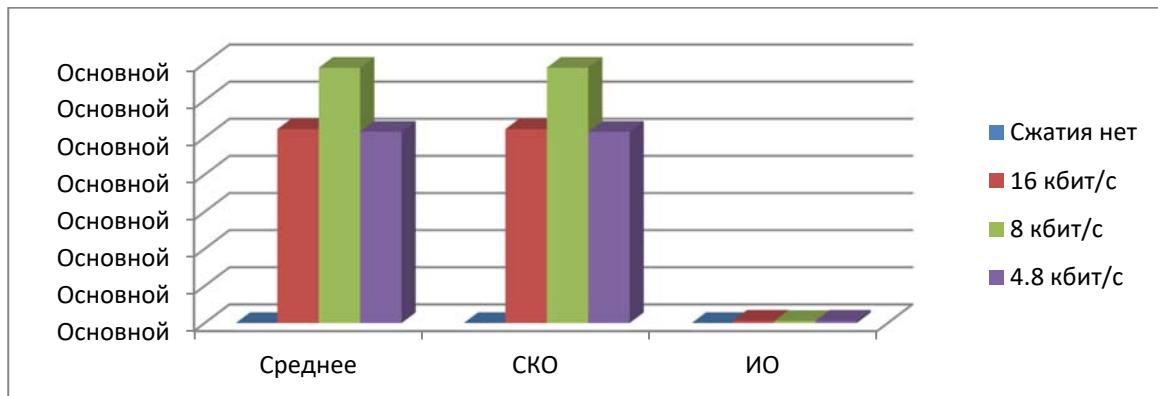


Рис. 18. Сигнал, подвергнутый обработке аудиопроцессором до кодека

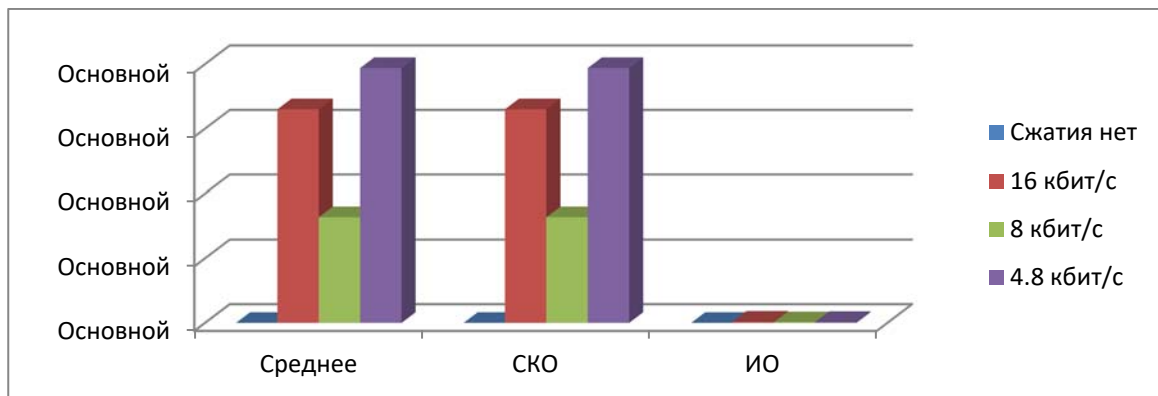


Рис. 19. Сигнал, подвергнутый обработке аудиопроцессором после кодека

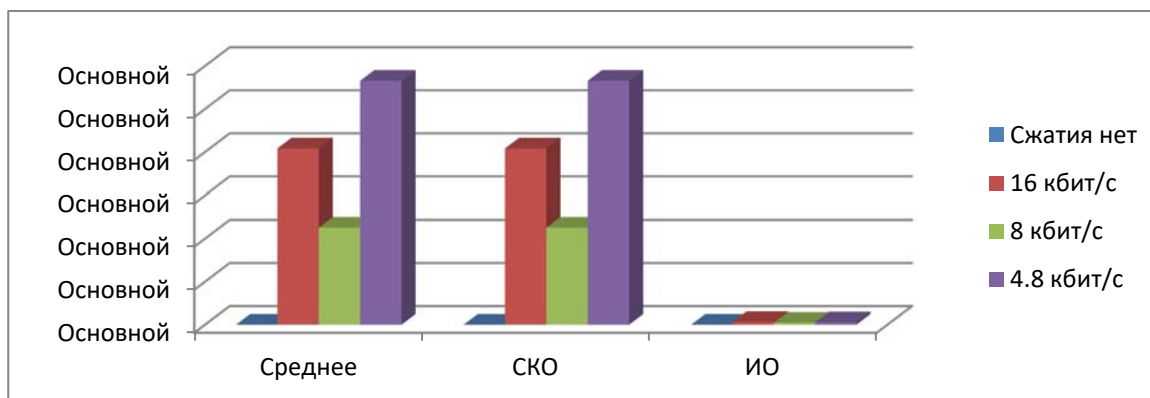


Рис. 20. Сигнал, подвергнутый обработке аудиопроцессором до и после кодека

Выводы

Оценив энергетические и спектральные характеристики сигналов, подвергнутых различной обработке, можно сделать следующие выводы:

- В процессе кодирования, энергетические параметры речевых сигналов заметно изменяются, происходит значительное снижение относительной средней мощности (ОСМ). Эти искажения РС приводят к изменению его динамического диапазона, нарушению формы и к появлению гармонических помех в паузах. Предварительная обработка РС с использованием корректора звукового сигнала (КЗС), в котором используется, специально разработанный алгоритм позволяет снизить негативные изменения структуры речевого сигнала при его сжатии

- Частотные характеристики звукового сигнала также существенно изменяются. Заметно искажается спектра и особенно в области высоких частот. В процессе кодирования появляются мощные узкополосные высокочастотные помехи. Всё это ухудшает основную характеристику речевого сигнала – разборчивость. Если объектом обработки (сжатия) является речевой сигнал, то в целях повышения его разборчивости можно использовать аудиопроцессор как перед, так и после кодера.

В целом, обработка речевого сигнала с использованием алгоритма коррекции (ЗКС) позволяет снизить негативные последствия его сжатия в процессе кодирования, при этом сохраняются объективные параметры качества сигнал: разборчивость, тембр и др.

Литература

1. Попов О.Б., Рихтер С.Г. Цифровая обработка сигналов в трактах звукового вещания, Учебное пособие для вузов. М.: Горячая линия - Телеком, 2015. 342 с.
2. Малов А.В., Ождихин Г.А., Попов О.Б., Черников К.А., Чернышева Т.В. «Методические указания по использованию программы исследования статистических свойств сигнала звукового вещания «ESTIM». М.: МТУСИ, 2013. 19 с.
3. Свидетельство о государственной регистрации программы для ЭВМ № 2013616645 от 15.07.2013м Программа «Анализ параметров сигналов звукового вещания «ESTIM»» Абрамов В.А., Ождихин Г.М., Черников К.В., Малов А.В.
4. Мишенков С.Л., Попов О.Б. Электроакустика и звуковое вещание конспект лекций. М.: Горячая линия - Телеком, 2014. 240 с.
5. Петелин Р.Ю., Петелин Ю.В. «Cool Edit Pro. Секреты мастерства». СПб: БХВ – Санкт-Петербург, Арлит, 2002. 432 с.
6. Свидетельство о регистрации программы для ЭВМ RUS 2014614408 27.11.2013. «Корректор звукового сигнала «КЗС»» Попов О.Б., Абрамов В.А., Литвин С.А., Касьянов А.А., Ождихин Г.М.
7. Гавриков О. В., Мальгин А. Д., Орлов В. Г., Попов О. Б., Самарин Е. В. Снижение искажений звукового сигнала при его передискретизации // Телекоммуникации и информационные технологии. 2020. Т. 9. № 1. С. 72-78.

INFLUENCE OF RECOVERY CORRECTION ON PARAMETERS SPEECH SIGNAL

Oleg B. Popov,

*Professor of the Department of T&ZV MTUCI,
PhD., Moscow, Russia,
o.b.popov@mtuci.ru*

Kirill A. Vasilevsky,

*Graduate of MTUCI, Москва, Россия,
alaxtver@yandex.ru*

Timofey D. Garanin,

*Graduate MTUCI, Москва, Россия,
veneriec@gmail.com*

Vladimir G. Orlov,

*Chief specialist of the Department
of the OofSRWofS of MTUCI, PhD., Moscow, Russia,
v.g.orlov@mtuci*

Dmitry I. Tsyplov,

*Graduate MTUCI, Moscow, Russia,
tsyplov-di@mail.ru*

Abstract

The article considers the objective impact of additional correction on the parameters of the speech signal. The statistical (energy and spectral) parameters of the speech signal were estimated with additional processing based on the CELP algorithm performed before and after the codec. The choice of this algorithm is substantiated and its advantages are described in comparison with other algorithms. An assessment was made of both positive and negative effects on the quality of the speech signal of audio processor processing before and after the codec.

Keywords: *linear prediction coding, CELP algorithm, relative average signal strength, dynamic processing by an audio processor, spectral characteristics.*

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ МОДЕЛЕЙ УДАЛЕНИЯ ФОНА

Давыдова Анастасия Анатольевна,
студент МТУСИ, Москва, Россия,
davydovaanastasia@gmail.ru

Егоров Дмитрий Аркадьевич,
старший преподаватель кафедры ТуЗВ МТУСИ, Москва, Россия,
d.a.egorov@mtuci.ru

Аннотация

В статье исследуются различные модели для «удаления» фона, сравнивается эффективность работы. Актуальность данной проблемы вызвана, прежде всего, необходимостью защиты конфиденциальности в связи с увеличением использования сервисов видеосвязи, а также для создания контента, визуального отслеживания и оперативного решения задач. В материале сравнивается работа наиболее популярных моделей, создается собственная метрика для оценки выходных данных, выявляется лучшая модель удаления заднего фона.

Ключевые слова: удаление фона, изображение, нейросеть, обработка, датасет, энкодер, декодер, модель, метрика, объект.

Введение

Потребность удалить фон с фото возникает довольно часто. Современные технологии тесно связаны с нашей жизнью и, следовательно, с ее изменением их нужно адаптировать под новые реалии. Увеличение количества видеоконференций во время эпидемии COVID-19 выдвинуло концепцию виртуальных фонов на передний план видеоизображения. Любой сервис видеосвязи, будь то Google Meet, Zoom или MS Teams и другие, имеет эту функцию. Виртуальные фоны могут быть действительно полезными для защиты своей конфиденциальности или скрытия своего хаотичного видео окружения.

Развить технологию удаления заднего фона необходимо в связи с тем, что, во-первых, люди все больше хотят оставить свою частную жизнь приватной и поэтому все чаще стараются скрывать место пребывания. Во-вторых, применение данной технологии используется при визуальном отслеживании и сегментации изображения. Но чаще всего удаление фона применяют при создании контента, так как к вырезанному фрагменту можно с легкостью подставить любой фон и тем самым получить совершенно новое изображение.

Существуют множество программ, с помощью которых можно убрать задний план, при этом самым проверенным методом является его ручная коррекция. Однако это трудоемкий процесс, требующий длительного времени. Зачастую людям требуется оперативное решение редактирования изображений, а сервисы, предоставляющие данную возможность либо платные, либо справляются с этой задачей не так качественно, как это необходимо. В данной работе сравниваются три самые популярные модели удаления заднего фона на базе нейросетей, выявляются их плюсы и минусы, а также определяется лучшая из них.

Нейронная сеть

Нейронная сеть представляет собой серию алгоритмов, которые пытаются распознать основные отношения в наборе данных с помощью процесса, имитирующего работу мозга. [1-5]. В этом смысле нейронные сети относятся к системам нейронов органического или искусственного характера. Нейронные сети могут адаптироваться к изменению входных данных, и поэтому сеть может генерировать наилучший возможный результат без необходимости адаптировать критерии вывода данных.

Нейросети состоят из трех основных компонентов (слоев): входной слой, слой обработки и выходной слой, (рис. 1).

Во входном слое данные, поступившие в нейросеть, взвешиваются на основе различных критериев. В слое обработки, скрытом от глаз, есть узлы и связи между этими узлами, аналогичные нейронам и синапсам в мозге живых организмов.

На данный момент существует большое количество различных видов нейросетей, но в основном выделяют три основных типа: искусственные нейронные сети (ANN), сверточные нейронные сети (CNN) и рекуррентные нейронные сети (RNN).

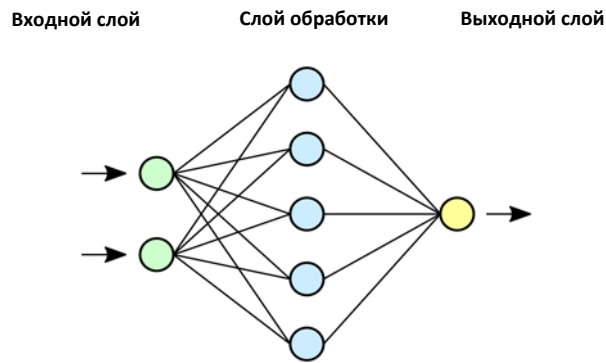


Рис. 1. Простая нейросеть

В данной работе будут рассматриваться только сверточные нейронные сети, так как данная сеть адаптирована для анализа и идентификации визуальных данных, таких как цифровые изображения или фотографии.

Особенности рассматриваемых моделей

Вначале рассмотрим модель Background Removal for Video Conferences by BigVision.

В отличие от других разработчиков, BigVision не предоставляет сведений о работе и устройстве своих моделей. Имеются лишь сведения о структуре нейросети, на которой основана данная модель – это DeepLab v3 [6]. DeepLab v3 является одной из последних нейросетей, на базе которых были разработаны все последующие модели.

DeepLab v3 – это архитектура семантической сегментации. Для сегментации объектов в нескольких масштабах разработаны модули, которые используют каскадную или параллельную свертку для захвата многомасштабного контекста за счет применения нескольких жестких оценочных блоков [7-11].

Архитектура модели DeepLab v3 приведена на рисунке 2.

Функции извлекаются из магистральной сети (VGG, DenseNet, ResNet). Чтобы контролировать размер вероятностных карт, в последних нескольких блоках магистрали используется жесткая свертка. Помимо извлеченных функций из магистрали добавляется сеть ASPP [12], для классификации каждого пикселя. Выходные данные сети ASPP проходят через свертку 1 x 1 для получения фактического размера изображения, которое будет окончательной сегментированной маской для изображения.

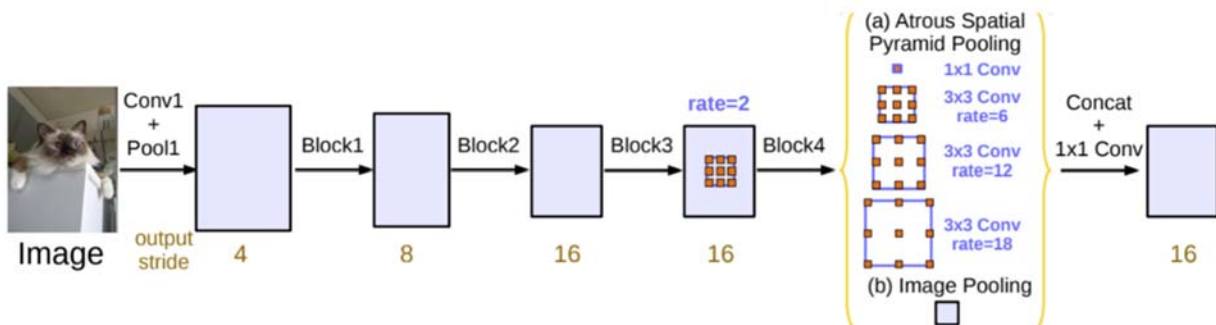


Рис. 2. Архитектура модели DeepLabV3

На рисунке 3 приведена модель DeepLabv3+ by Pinto Model Zoo

Данная модель реализуется в нейросети с энкодером-декодером с разделяемой сверткой Atrous: модуль кодирует многомасштабную контекстную информацию, применяя параллельную свертку в нескольких масштабах [6], в то время как простой, но эффективный модуль декодера уточняет результаты сегментации вдоль границ объекта.

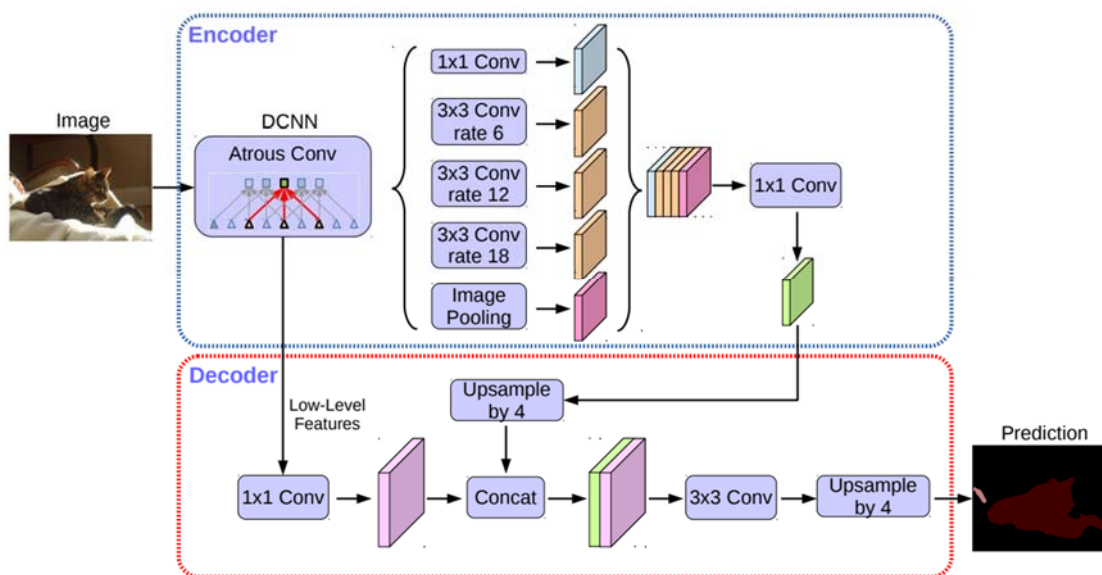


Рис. 3. Архитектура модели DeepLabv3+ by Pinto Model Zoo

Тренировочный датасет. DeepLabv3+ предварительно обучается на наборе данных

MS-COCO, который включает в себя более 330 тысяч изображений. Предварительно обучение также осуществляется на JFT, который состоит из одного миллиарда изображений. Аналогично используется модель Xception, которая была предварительно обучена на наборах данных ImageNet-1k содержащий 1,2 млн изображений и JFT-300M содержащий 300 млн изображений.

Оценочный датасет. Данная модель оценивается на тесте семантической сегментации PASCAL VOC 2012 [13], который содержит 20 классов объектов переднего плана и один класс фоновых объектов. Исходный набор данных содержит 1456 изображений. Этот набор данных был дополнительно расширен, в результате чего было получено 10 582 обучающих изображений [14].

Оценочная метрика. Чтобы оценить эффект свертки 1×1 с 48 слоями в модуле декодера, используется свертка 3×3 с 256 фильтрами и функции Conv2 из сетевой магистрали ResNet-101, т. е. последняя карта вероятности в остаточном блоке res2x. После сопоставлений карт признаков Conv2 с картами признаков DeepLabv3 выяснилось, что более эффективно использовать две свертки 3×3 с 256 фильтрами, чем использовать одну или три свертки. Изменение количества фильтров с 256 до 128 и размера ядра с 3×3 до 1×1 ухудшает производительность [15, 16].

Обучение нейросети. В DeepLabv3+ применяется нейросеть ResNet-101 [17] тренированная на ImageNet-1k [18]. Модель обучается напрямую без предварительного обучения каждого компонента.

Вывод по модели: модель «DeepLabv3+» использует структуру кодер-декодер, в которой DeepLabv3 используется для кодирования богатой контекстной информации, а простой, но эффективный модуль декодера применяется для восстановления границ объекта.

Модель Background Removal by Xuebin Qin. Данная модель использует U-Net-подобную структуру, а именно двухуровневую вложенную структуру (U2 -Net), приведенную на рисунке 4.

U2 -Net состоит из трех частей:

1. 6 ступеней энкодера
2. 5 ступеней декодера
3. Модуль слияния карт значимости со ступенями энкодера и декодера.

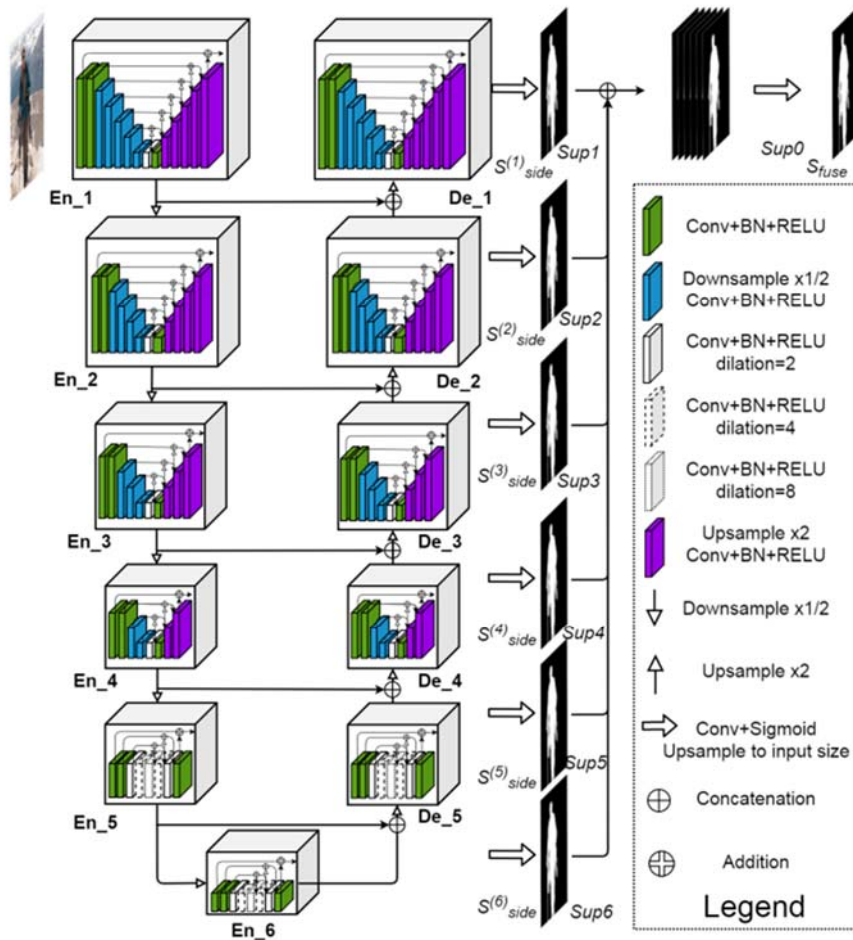


Рис. 4. Архитектура модели Background Removal

Описание работы нейросети

В нейросеть подается изображение. Как известно, в нейронных сетях каждый слой передает данные следующему слою. В нашем случае в сети с остаточными блоками каждый уровень передает информацию следующему слою и непосредственно уровням, находящимся на расстоянии примерно 2–3 перехода. Таким образом, дизайн такой нейросети имеет более глубокую архитектуру с богатыми многомасштабными функциями и относительно низкими затратами на вычисления и память. Кроме того, поскольку архитектура U2-Net построена только на Residual U-блоках (RSU) остаточных блоках без использования каких-либо предварительно обученных магистралей, адаптированных из классификации изображений, она отличается гибкостью и легко адаптируется к различным рабочим средам с незначительной потерей производительности.

Тренировочный датасет. Тренировка нейросети осуществлялась на DUTS-TR, которая является частью DUTS dataset [19]. DUTS-TR состоит из 10553 изображений. На данный момент, это самый большой и наиболее часто используемый датасет для тренировки нейросети выделения объектов. Данный датасет был увеличен путем горизонтального поворота, чтобы собрать 21106 тренировочных изображений оффлайн.

Оценочный датасет. Оценка работы нейросети осуществлялась на шести самых популярных анализирующих датасетах: DUTOMRON [20], DUTS-TE [19], HKU-IS [21], ECSSD [22], PASCAL-S [23], SOD [24]. DUTOMRON включает в себя 5168 изображений, большинство из которых содержат один или два структурно сложных объекта переднего плана. DUTS датасет состоит из двух: DUTS-TR и DUTS-TE. Как упоминалось ранее, DUTS-TR был использован для обучения. Следовательно, DUTS-TE, который содержит 5019 изображений, выбран в качестве одного из оценочных наборов данных. HKU-IS содержит 4447 изображений с несколькими объектами переднего плана. ECSSD содержит 1000 структурно сложных изображений, многие из которых содержат крупные объекты переднего плана. PASCAL-S содержит 850 изображений со сложными объектами переднего плана и загроможденным фоном. SOD содержит всего лишь 300 изображений, но так как изначально он был разработан для сегментации изображений, они зачастую имеют низкую контрастность или содержат сложные объекты переднего плана, перекрывающиеся с границей изображения.

Оценочная метрика. Выходные данные метода выделения объектов представляют собой вероятностные карты, которые имеют то же пространственное разрешение, что и входные изображения. Каждый пиксель предсказанных карт выделенного объекта имеет значение в диапазоне от 0 до 1 (или (0, 255)). Маской являются карты бинарных значений, в которых каждый пиксель равен 0 или 1 (или 0 или 255), где 0 указывает фоновые пиксели, а 1 указывает пиксели выступающего объекта переднего плана. Чтобы всесторонне оценить качество этих вероятностных карт по сравнению с маской, используется шесть способов:

- Кривые точности-отзыва,
- Максимальные F-измерения [7],
- Средняя абсолютная ошибка [25, 28, 26],
- Взвешенные F-измерения [25],
- Структурированные измерения [29],
- Ослабленные F-измерения границы [28].

Обучение нейросети. В процессе обучения нейросети размер каждого изображения сначала изменяется до 320×320 , случайным образом переворачивается по вертикали и обрезается до 288×288 . Обучение сети осуществляется с нуля. Веса потерь w (m) side и w_{fuse} устанавливаются равными 1. Оптимизатор Adam (адаптивной скорости обучения) [30] используется для обучения сети. Сеть обучается до тех пор пока потери не сойдутся без использования способов проверки, которые описаны ранее [25, 26, 31]. После 600 тыс. итераций потери при обучении сходятся, и весь процесс обучения занимает около 120 часов. Во время тестирования входные изображения ($B \times H$) изменяются до 320×320 и передаются в сеть для получения карт значимости. Прогнозируемые карты значимости размером 320×320 изменяются до исходного размера входного изображения ($B \times H$). В обоих процессах изменения размера используется билинейная интерполяция.

Вывод по модели: Преимущество данного дизайна заключается в том, что он способен захватывать больше контекстной информации из разных масштабов благодаря смешению рецептивных полей разного размера в ReSidual U-блоках (RSU).

Эксперимент

Чтобы выявить лучший способ удаления заднего фона необходимо сравнить выходные данные методов в одинаковых условиях, и поэтому мы создаем свою метрику для оценки. Для этого был сделан собственный тестовый набор данных - фото с человеком на фоне и фото этого фона при тех же условиях, (рисунок 5).

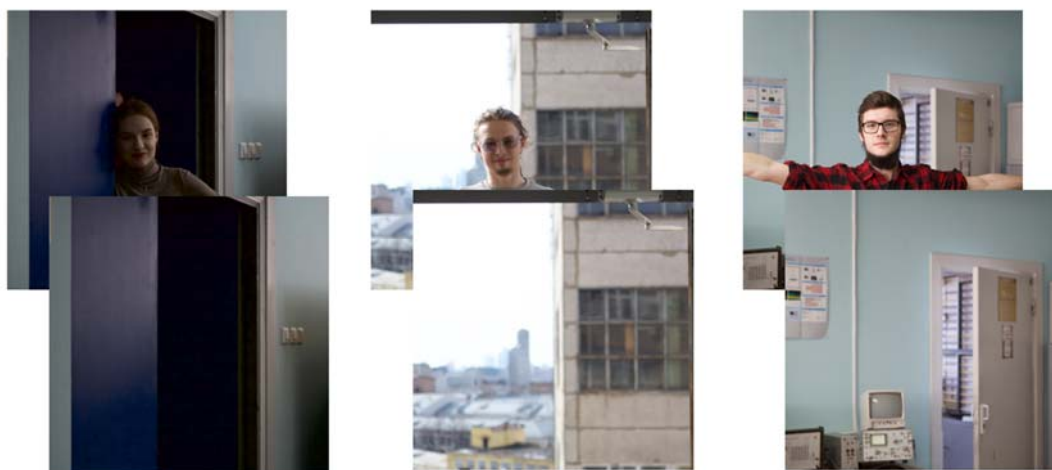


Рис. 5. Тестовые изображения

Таким образом, мы получаем путем вычитания двух изображений маску – «человек без фона» (рис. 6).



Рис. 6. Эталонные маски

Затем мы обрабатываем наши тестовые изображения через выбранные нами нейронные сети и, тем самым, получаем изображения, из которых для оценки качества мы вновь делаем маски, но уже с использованием изображений с удаленным фоном, по которым, в свою очередь, формируем маски способом аналогичным созданию тестовых изображений.

Далее мы сравниваем маски, полученные с фотографий и обработанных с помощью нейросетей с эталонной маской, и выбираем лучшую модель удаления фона путем получения разности эталонной и полученной маски. Модель, в которой ошибка минимальна, принимается за лучшую.

Согласно закону Вебера-Фехнера, интенсивность ощущения линейно зависит от логарифма интенсивности возбуждения, и поэтому для большей репрезентативности ошибку будем считать в децибелах,

Ввиду того, что изображения, полученные с помощью модели DeepLabv3+ by Pinto Model Zoo, визуально уже отличались большой ошибкой, было принято решение о нецелесообразности расчетной части для этой нейронной сети.

Результаты сравнения моделей приведены на рисунке 7.

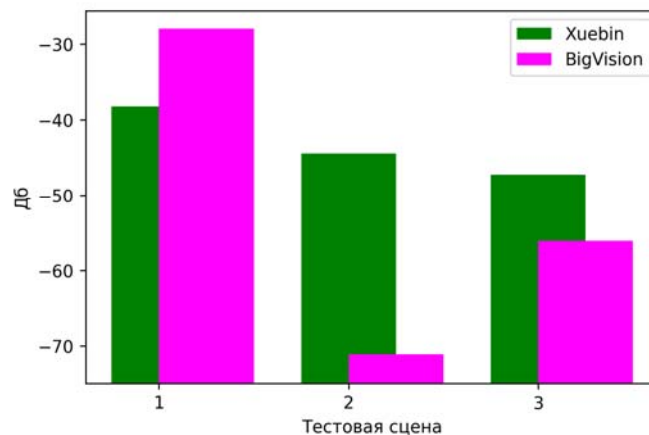


Рис. 7. Сравнение ошибок моделей BigVision и Xuebin Qin

Заключение

По результатам эксперимента нельзя однозначно сделать вывод о том, какая из моделей является лучшей. На рисунке 7 видно, что хотя и модель Background Removal for Video Conferences by **BigVision** имеет минимальную ошибку на двух тестовых изображениях, модель Background Removal by **Xuebin** Qin при этом демонстрирует более стабильную работу. Стоит также учесть, что в данной работе были проведены исследования лишь с технической точки зрения и не учитывались аспекты зрительного восприятия изображения человеком. С учётом этого было принято решение о продолжении исследований с использованием тестовой базы большего объема и более эффективной метрики оценки качества удаления фона, а также о необходимости проведения субъективно-оценочной экспертизы.

Литература

1. *Rosenblatt F.* The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain // *Psychological Review*, № 65 (6), pp. 386-408, 1958.
2. *Weng J., Ahuja N., Huang T. S.* Learning recognition and segmentation using the Cresceptron // *International Journal of Computer Vision*, vol. 25, no. 2, pp. 105-139, Nov. 1997.
3. *Siegelmann H.T., Sontag E.D.* Turing computability with neural nets (PDF) // *Appl. Math. Lett.* №4 (6), pp. 77-80, 1991.
4. *Nicola Secomandi.* Comparing neuro-dynamic programming algorithms for the vehicle routing problem with stochastic demands // *Computers & Operations Research*. 27 (11-12), pp. 1201-1225, 2000.
5. *Goodfellow Ian, Pouget-Abadie Jean, Mirza Mehdi, Xu Bing, Warde-Farley David, Ozair Sherjil, Courville Aaron, Bengio Yoshua.* Generative Adversarial Networks // *Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014)*, pp. 2672-2680.
6. *Chen L.C., Papandreou G., Schroff F., Adam H.* Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587, 2017.
7. *Achanta R., Hemami S., Estrada F., Susstrunk S.* Frequency-tuned salient region detection // *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597-1604, 2009.
8. *Brandt A.* Multi-level adaptive solutions to boundary-value problems // *Mathematics of computation*, №31(138), pp. 333-390, 1977.
9. *Terzopoulos D.* Image analysis using multigrid relaxation methods // *TPAMI*, №(2), pp. 129-139, 1986.
10. *Briggs W. L., Henson V. E., McCormick S. F.* A multigrid tutorial. SIAM, 2000.
11. *Papandreou G., Maragos P.* Multigrid geometric active contour models. *TIP*, № 16(1), pp. 229-240, 2007.
12. *Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A. L.* Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv:1606.00915, 2016.
13. *Everingham M., Eslami S.M.A., Gool L.V., Williams C.K.I., Winn J., Zisserman A.* The pascal visual object classes challenge a retrospective // *IJCV*, 2014.
14. *Hariharan B., Arbel'aez P., Bourdev L., Maji S., Malik J.* Semantic contours from inverse detectors // *ICCV*. 2011.
15. *Ronneberger O., Fischer P., Brox T.* U-net: Convolutional networks for biomedical image segmentation // *MICCAI*. 2015.
16. *Badrinarayanan V., Kendall A., Cipolla R.* Segnet: A deep convolutional encoder-decoder architecture for image segmentation // *PAMI*. 2017.
17. *He K., Zhang X., Ren S., Sun J.* Deep residual learning for image recognition // *CVPR*. 2016.
18. *Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A.C., Fei-Fei L.* ImageNet Large Scale Visual Recognition Challenge // *IJCV*. 2015.
19. *Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan.* Learning to detect salient objects with image-level supervision // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 136-145, 2017.
20. *Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang.* Saliency detection via graph-based manifold ranking // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3166-3173, 2013.
21. *Guanbin Li and Yizhou Yu.* Visual saliency detection based on multiscale deep cnn features // *IEEE Transactions on Image Processing*, №25(11), pp. 5012-5024, 2016.
22. *Yu Zeng, Yunzhi Zhuge, Huchuan Lu, Lihe Zhang, Mingyang Qian, and Yizhou Yu.* Multi-source weak supervision for saliency detection // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6074-6083, 2019.
23. *Yin Li, Xiaodi Hou, Christof Koch, James M Rehg, and Alan L Yuille.* The secrets of salient object segmentation // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 280-287, 2014.
24. *Vida Movahedi and James H Elder.* Design and perceptual validation of performance measures for salient object segmentation // *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pp. 49-56. IEEE, 2010.
25. *Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang.* A simple pooling-based design for realtime salient object detection // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3917-3926, 2019.
26. *Nian Liu, Junwei Han, and Ming-Hsuan Yang.* Picanet: Learning pixel-wise contextual attention for saliency detection // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3089-3098, 2018.
27. *Егоров Д.А.* Выявление аномального поведения группы объектов в видеопоследовательности. Сборник трудов XIII международной научно-технической конференции // *Технологии информационного общества. Материалы XIII Международной отраслевой научно-технической конференции*. 2019. С. 167-169.
28. *Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand.* Basnet: Boundary-aware salient object detection // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7479-7489, 2019.
29. *Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji.* Structure-measure: A new way to evaluate foreground maps // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4548-4557, 2017.
30. *Diederik P Kingma and Jimmy Ba.* Adam: A method for stochastic optimization. arXiv preprint, 2014.
31. *Lu Zhang, Ju Dai, Huchuan Lu, You He, and Gang Wang.* A bi-directional message passing model for salient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1741-1750, 2018.
32. *Власюк И.В.*, метод оценки эффективности алгоритмов обработки видеосигналов в телевизионных камерах при передаче движущихся объектов // *Технологии информационного общества. Материалы XIII международной отраслевой научно-технической конференции*. 2019. С. 158-160.

EFFICIENCY EVALUATION OF BACKGROUND REMOVAL ALGORITHMS

Anastasia A. Davydova,
Student MTUCI, Moscow, Russia,
davydovaanastasiaan@gmail.com

Dmitry A. Egorov,
Senior Lecturer of the Department of T&SB MTUCI,
Moscow, Russia,
d.a.egorov@mtuci.ru

Abstract

The article explores different modules of background removal, evaluates the efficiency of their work. The actuality of this problem is caused with the need of confidentiality protection with the increasing usage of services for videoconferences and also for making content, visual detection and operative problem solution. In this material we take a look at the methods used in the models, evaluate the work of the most popular models for background removal with the help of proposed new metrics for output results. All of this helps to define pluses and minuses of studied algorithms.

Keywords: *background removal, image, neural network, processing, dataset, encoder, decoder, model, metrics, object.*

МОДЕЛИРОВАНИЕ КОМПЕНСАЦИИ НЕЛИНЕЙНЫХ ИСКАЖЕНИЙ В УСИЛИТЕЛЕ МОЩНОСТИ НА ОСНОВЕ ЕГО ЦИФРОВОЙ ПОВЕДЕНЧЕСКОЙ МОДЕЛИ

Максимов Антон Алексеевич,
аспирант МТУСИ, Москва, Россия,
mash2525@yandex.ru

Комаров Игорь Валерьевич,
магистрант МТУСИ, Москва, Россия
ivk1406@mail.ru

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

Аннотация

Разработана поведенческая модель усилителя мощности на основе уравнения регрессии, позволяющая разложить искаженный сигнал на выходе усилителя на неискаженную составляющую и аддитивную совокупность помех нелинейных искажений. Установлено, что такая помеха состоит из безынерционных нелинейных составляющих и включений, соответствующих проявлению эффектов памяти. Показано, что коррекция только безынерционных составляющих помехи нелинейных искажений позволяет эффективно бороться с внеполосными излучениями, а компенсация эффектов памяти – с внутриполосной помехой. Установлено, что при максимально возможной компенсации безынерционной составляющей помехи лишь компенсация также и эффектов памяти позволяет обеспечить необходимое подавление помехи нелинейных искажений.

***Ключевые слова:** усилитель мощности, поведенческая модель, уравнение регрессии, модель Кима и Константину, компенсация нелинейных искажений, предварительные цифровые искажения, внеполосная помеха, внутриполосная помеха.*

Введение

Основные требования, предъявляемые к современным радиопередающим устройствам – это линейность, широкая полоса пропускания и высокий КПД при усилении широкополосных сигналов с значительным пик-фактором [1,2]. Основным источником нелинейных искажений в радиопередатчике является усилитель мощности (УМ), и выбор режимов его работы всегда приводит к неизбежным противоречиям между его линейностью и энергетической эффективностью (КПД) [1,4-8]. При этом обычно добиваются максимально возможной энергетической эффективности УМ при сохранении нелинейных искажений усиленного сигнала на уровне, который может быть скомпенсирован при использовании методов линеаризации его работы. В настоящее время используются цифровые методы линеаризации работы УМ в сочетании с его поведенческими моделями. Наиболее общими и сложными являются модели с нелинейными эффектами памяти, причины, возникновения которых в УМ весьма многочисленны [3,9-11].

Необходимость рассмотрения нелинейных эффектов памяти усилителя УМ при решении проблемы компенсации искажений усиленного сигнала зависит, как правило, от вида амплитудной огибающей и ширины спектра, а также от характеристик рассеиваемой мощности и работы системы охлаждения УМ, то есть от тепловых эффектов применяемого активного элемента (транзистора). По-видимому, если ширина спектра сигнала не превышает 10 МГц, а флуктуации температуры в области переходов транзистора нескольких градусов, то методами схемотехнического проектирования усилителя мощности можно добиться пренебрежимо малого влияния динамических эффектов памяти на искажения усиленного сигнала. В этом случае нелинейные эффекты объясняются естественными условиями (режимами) функционирования высокоэффективного УМ, который в этом случае можно рассматривать как нелинейный статический прибор, описываемый с помощью модели Салеха, ряда Тейлора и т.д. [5,6]. Но в современных системах связи, с шириной полосы усиленного сигнала несколько десятков МГц и более, практически всегда возникают нелинейные эффекты памяти в передающей аппаратуре, компенсация которых позволяет повысить информационную эффективность канала связи.

Целью данной работы является моделирование компенсации как безынерционных искажений, так и нелинейных эффектов памяти в УМ при использовании его цифровых регрессионных моделей на основе ряда Вольтерры.

Инерционные регрессионные модели на основе ряда Вольтерры

Ряд Вольтерры можно записать в дискретной форме [1-4,6]:

$$x_{\text{ВЫХ}}(k) = \sum_{n=1}^N \sum_{m_1=0}^M \sum_{m_2=0}^M \cdots \sum_{m_n=0}^M h_n(m_1, m_2, \dots, m_n) \prod_{r=1}^n z_{\text{ВХ}}(k - m_r), \quad (1)$$

где: $x_{\text{ВЫХ}}(k)$ и $z_{\text{ВХ}}(k)$ – отсчеты выходного и входного сигналов в момент времени с номером k , далее, $h_n(m_1, m_2, \dots, m_n)$ – ядра Вольтерры порядка n , представляющие собой многомерные импульсные отклики нелинейной инерционной системы, M – глубина памяти модели, N – ее порядок.

Как видно, модель (1) весьма сложна, и содержит большое число коэффициентов, подлежащих идентификации. Кроме того, она относится к высокочастотному сигналу, и можно попробовать получить выражения, связывающие отсчеты комплексных огибающих входного и выходного сигналов.

Известны различные варианты упрощенных моделей для комплексных огибающих [1], самая простая из которых – модель Кима и Константину [2]. Ее можно получить, удалив из ряда Вольтерры все кросс-компоненты. В результате, для отсчетов комплексной огибающей сигнала на выходе нелинейного объекта с памятью можно записать:

$$\dot{x}_{\text{ВЫХ}}(k) = \sum_{m=0}^M \sum_{n=1}^N \dot{a}_{mn} \dot{z}_{\text{ВХ}}(k-m) \left| \dot{z}_{\text{ВХ}}(k-m) \right|^{n-1}, \quad (2)$$

где \dot{a}_{mk} – коэффициенты модели.

Последнее выражение можно записать в матричной форме:

$$\dot{x}_{\text{ВЫХ}}(k) = \dot{Z}_{\text{ВХ}M}(k)A, \quad (3)$$

где коэффициенты модели рассматриваются в виде матрицы-строки

$$A = \left[\dot{a}_{01} \quad \dots \quad \dot{a}_{0N} \quad \dot{a}_{11} \quad \dots \quad \dot{a}_{1N} \quad \dots \quad \dot{a}_{M1} \quad \dots \quad \dot{a}_{MN} \right]^T,$$

а вектор-столбец $\dot{Z}_{\text{ВХ}M}(k)$ из произведений сдвинутых отсчетов комплексной огибающей входного сигнала имеет вид:

$$\dot{Z}_{\text{ВХ}M}(k) = \begin{bmatrix} \dot{z}_{\text{ВХ}}(k) \\ \vdots \\ \dot{z}_{\text{ВХ}}(k) \left| \dot{z}_{\text{ВХ}}(k) \right|^{N-1} \\ \dot{z}_{\text{ВХ}}(k-1) \\ \vdots \\ \dot{z}_{\text{ВХ}}(k-1) \left| \dot{z}_{\text{ВХ}}(k-1) \right|^{N-1} \\ \dot{z}_{\text{ВХ}}(k-M) \\ \vdots \\ \dot{z}_{\text{ВХ}}(k-M) \left| \dot{z}_{\text{ВХ}}(k-M) \right|^{N-1} \end{bmatrix}^T. \quad (4)$$

На рисунке 1 изображена блок-схема данной модели. Из его анализа следует, что модель можно рассматривать как комбинацию $(M+1)$ полиномиальных функций, каждая из которых применяется к задержанному отсчету входного комплексного сигнала $\dot{z}_{\text{ВХ}}(k)$.

Таким образом, общее число коэффициентов, подлежащих идентификации, будет равно $N(M+1)$.

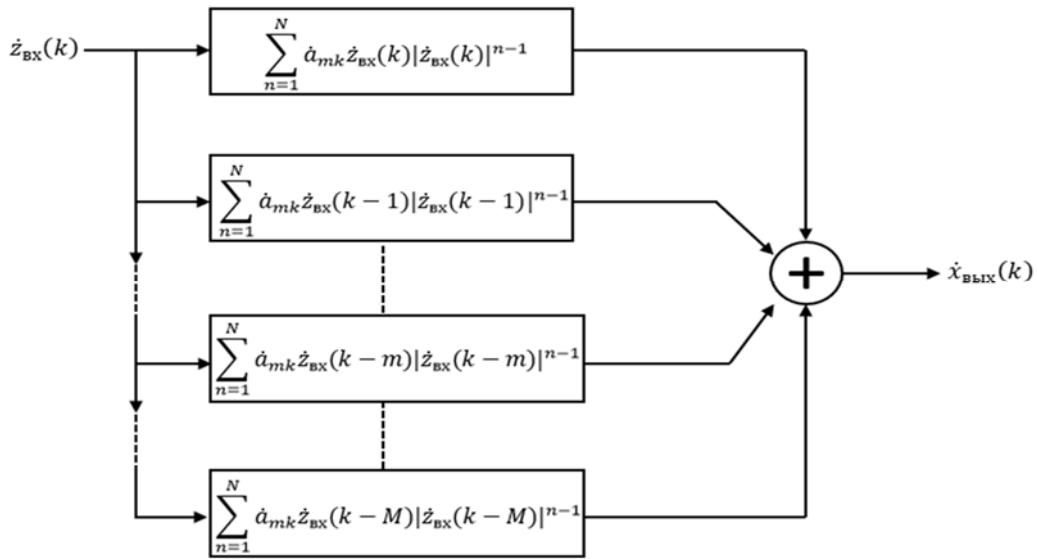


Рис. 1. Иллюстрация модели Кима и Константину

Предварительные цифровые искажения для полиномиальных моделей нелинейных эффектов памяти

Блок-схема устройства компенсации нелинейных искажений УМ с устройством предварительного цифрового искажения (ПЦИ) сигнала для обеспечения его линейного усиления приводится на рисунке 2.

На первом этапе исследования необходимо выбрать адекватную полиномиальную модель искажений сигнала в УМ с нелинейными эффектами памяти и идентифицировать ее коэффициенты. С этой целью в данном разделе будем рассматривать подход, основанный на решении полных систем линейных уравнений (СЛУ). На самом первом этапе следует использовать максимально сложную полиномиальную модель при учете имеющихся вычислительных ресурсов. В результате можно гарантированно выбрать наиболее точную, с точки зрения используемых критериев, модель искажений в УМ с нелинейными эффектами памяти. В этом случае, вероятно, придется решать СЛУ большой размерности из нескольких сотен или даже тысяч уравнений.

С целью разъяснения сути данного подхода рассмотрим утрированно простой пример, когда в результате первого этапа исследований выяснилось, что с высокой точностью нелинейный УМ описывается моделью Кима и Константину при $M = 1$ и $N = 2$, тогда:

$$\dot{x}_{\text{вых.УМ}}(k) = \sum_{m=0}^1 \sum_{n=1}^2 \dot{a}_{mn} \dot{z}_{\text{вх.УМ}}(k-m) \left| \dot{z}_{\text{вх.УМ}}(k-m) \right|^{n-1}, \tag{4}$$

и для идентификации коэффициентов модели необходимо было для любого выбранного значения k решить СЛУ лишь из 4-ех уравнений относительно комплексных коэффициентов $\dot{a}_{01}, \dot{a}_{02}, \dot{a}_{11}, \dot{a}_{12}$.



Рис. 2. Блок-схема УМ с устройством предварительного цифрового искажения

Если в (4) поменять местами $\dot{x}_{\text{вых.УМ}}(k)$ и $\dot{z}_{\text{вх.УМ}}(k)$, то получим уравнение, описывающее искажения сигнала $\dot{x}_{\text{вых.УМ}}(k)$ в устройстве ПЦИ - такие, что после его искажения в УМ получится опять $\dot{x}_{\text{вых.УМ}}(k)$. То есть в этом случае коэффициенты модели (4) относятся к устройству ПЦИ.

Таким образом, для идентификации коэффициентов модели устройства ПЦИ достаточно решить СЛУ той же размерности, что и размерность СЛУ исходного нелинейного устройства.

Отметим, что при этом мы всегда будем оценивать коэффициенты модели устройства ПЦИ по искаженному сигналу, имеющему меньший по значению пик-фактор по сравнению с пик-фактором исходного неискаженного сигнала. При больших амплитудных искажениях сигнала в УМ это может привести к несоответствию модели УМ, работающему вместе с устройством ПЦИ, полученному ранее уравнению регрессии для него (без устройства ПЦИ).

Имитационное моделирование компенсации искажений с нелинейными эффектами памяти

Моделирование УМ с нелинейными эффектами памяти и устройства его предварительной цифровой коррекции производилось при использовании программной среды MATLAB/Simulinc с целью создания модели, достаточно точно имитирующей особенности их совместной работы. Производилась также оценка влияния параметров модели на качественные показатели эффективности работы УМ.

Модель УМ

Рассматривалась полиномиальная модель Кима и Константину с параметрами $M = 5$ и $N = 5$ Входной сигнал имитировался с помощью генератора белого гауссовского шума, пропущенного через низкочастотный эллиптический фильтр.

После запуска модели входные и выходные сигналы передавались в рабочую область MATLAB для последующего извлечения коэффициентов модели предварительного цифрового корректора. Полученные передаточные характеристики УМ показаны на рисунке 3. Спектральные плотности мощности входных и выходных сигналов показаны соответственно на рисунках 4 и 5.

В таблице 1 приведена оценка АСРР [1] для двух разработанных моделей УМ (буквой В обозначена верхняя полоса, а буквой Н – нижняя).

Коррекция нелинейности

Поскольку работа УМ уже была промоделирована, то его полиномиальная модель и ее коэффициенты считаем известными. Их количество составляет $MN = 25$. Но если бы необходимо было их идентифицировать, то при известных отсчетах входного и выходного сигналов УМ можно было бы воспользоваться оператором MATLAB'a \ (обратный слэш или деление слева), что соответствует решению переопределенной системы методом наименьших квадратов.

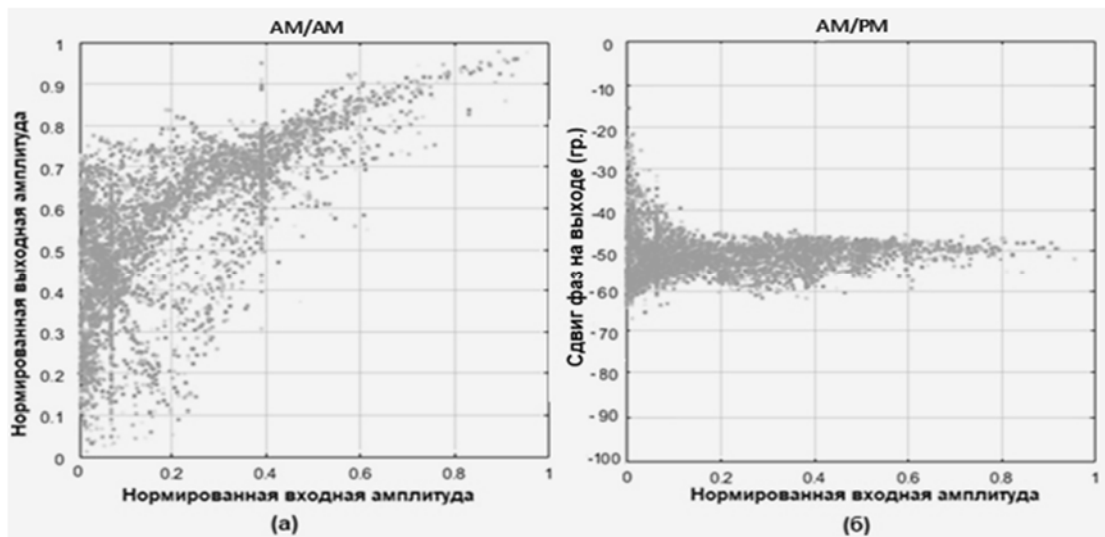


Рис. 3. Экспериментальные передаточные характеристики для полиномиальной модели с памятью

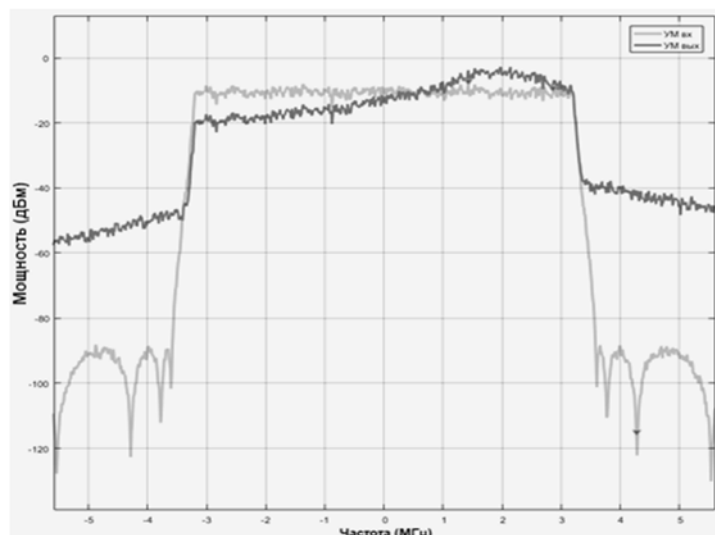


Рис. 4. Спектральная плотность мощности сигнала на входе и выходе модели Винера

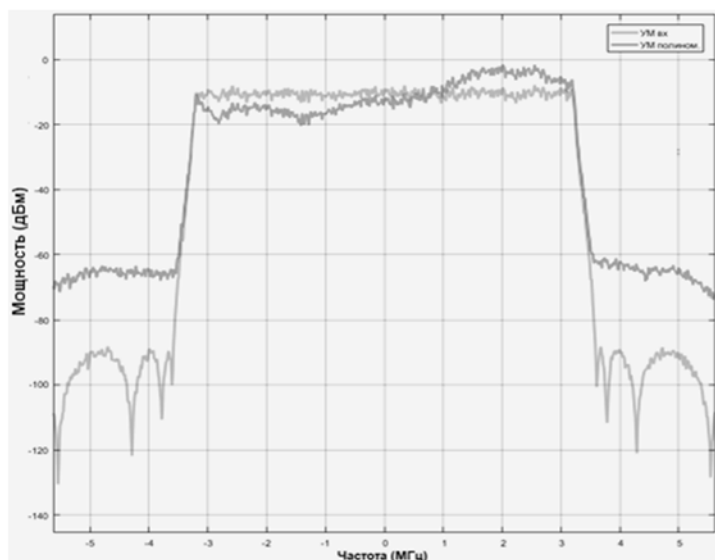


Рис. 5. Спектральная плотность мощности сигнала на входе и выходе полиномиальной модели при $M = 5$ и $N = 5$

Таблица 1

ACPR для двух разработанных моделей УМ (В - верхняя полоса, Н – нижняя полоса)

Модель	$ACPR_{Н_{ВХ}}$, дБ	$ACPR_{В_{ВХ}}$, дБ	$ACPR_{Н_{ВЫХ}}$, дБ	$ACPR_{В_{ВЫХ}}$, дБ
Полином	-82.4	-82.4	-55	-56

Далее в программе MATLAB вычислялись 50 отсчетов искаженного сигнала на выходе устройства предварительной цифровой коррекции в соответствии с алгоритмом, описанным выше, последние 25 из которых использовались для расчета коэффициентов модели корректора.

Результаты вычислений отображены на рисунке 6.

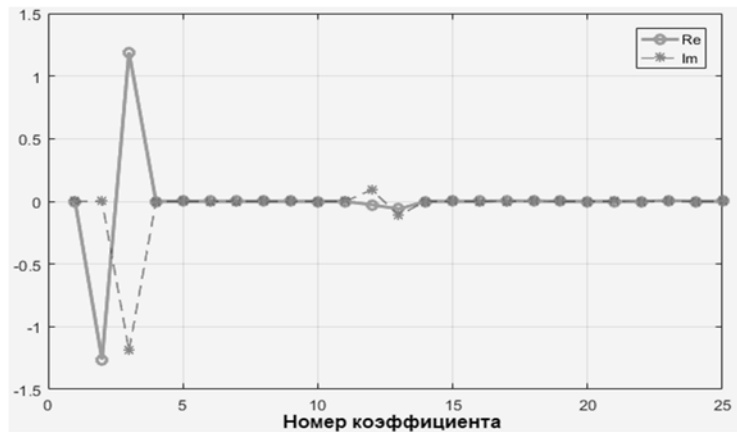


Рис. 6. Действительная и мнимая компоненты полученных коэффициентов коррекции

На рисунках 7 и 8 приведены соответственно структурная схема модели статической коррекции нелинейности УМ и подсистема цифрового корректора.

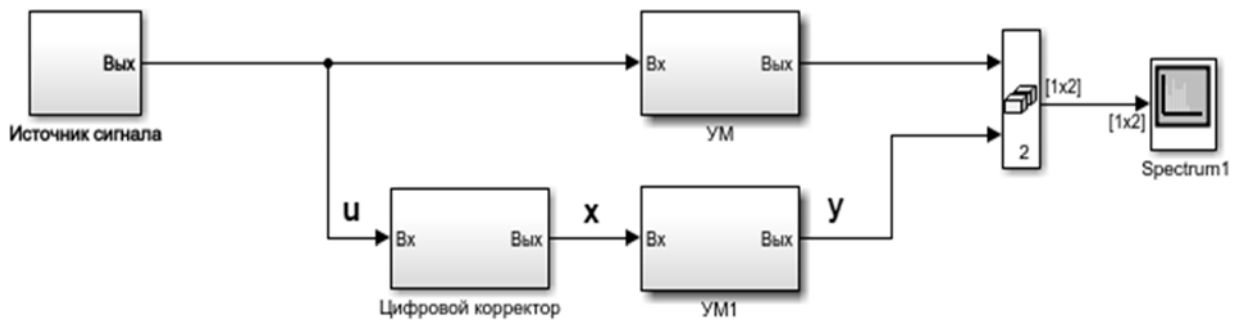


Рис. 7. Модель статической коррекции нелинейности УМ

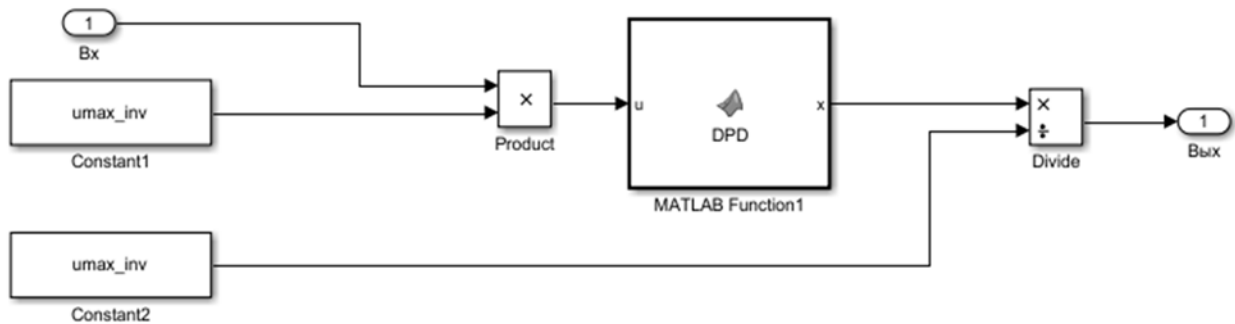


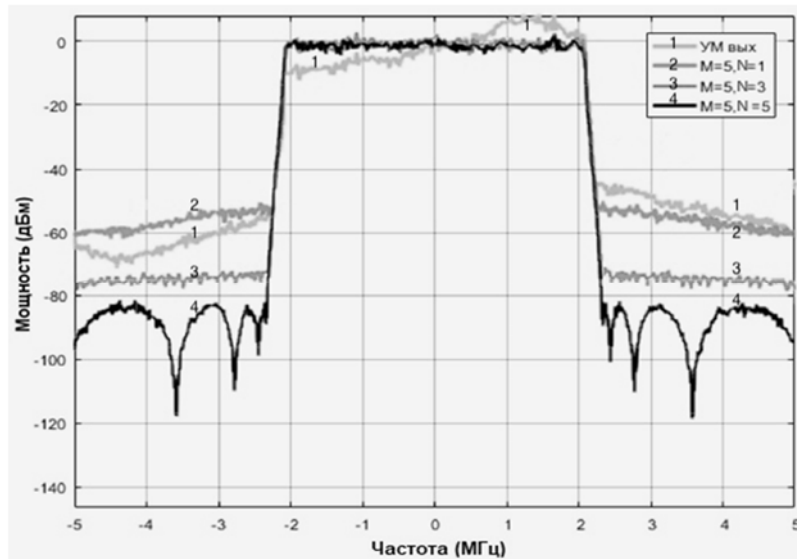
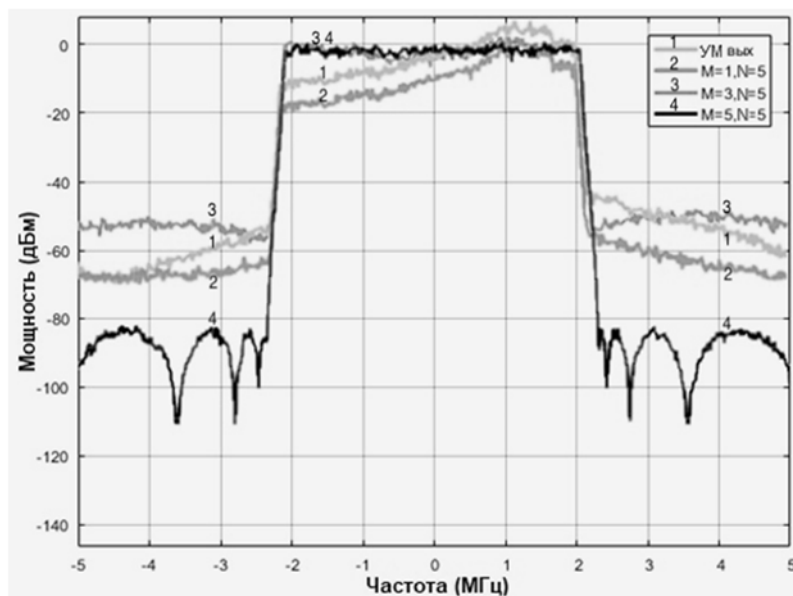
Рис. 8. Подсистема цифрового корректора

Результаты симуляции модели при различных значениях параметров M и N отражены в таблице 2 и на рисунках 9, 10.

Таблица 2

Результаты симуляции модели при различных значениях параметров M и N

Параметры DPD	$ACPR_{Н\text{Вых}}$, дБ	$ACPR_{В\text{Вых}}$, дБ
УМ выход	- 60.2	- 56.3
$M = 5, N = 1$	- 55.5	- 58.4
$M = 5, N = 3$	- 75	- 75.1
$M = 1, N = 5$	- 67.8	- 67.1
$M = 3, N = 5$	- 53	- 52.8
$M = 5, N = 5$	- 94.9	- 95.2

Рис. 9. Коррекция нелинейности при изменении параметра N Рис. 10. Коррекция нелинейности при изменении параметра M

Заключение

Из анализа двух последних рисунков следует, что все рассмотренные варианты предварительных цифровых искажений позволяют эффективно бороться с внутрисполосной помехой, но корректор, учитывающий в большей степени эффекты памяти, чем безынерционные нелинейные искажения, не способен эффективно бороться с подросом спектра в соседних каналах. Вместе с тем, из рисунка 10 следует, что при максимально возможной компенсации безынерционных нелинейных искажений только компенсация ещё и эффектов памяти позволяет обеспечить необходимое подавление помехи нелинейных искажений.

Литература

1. Горгадзе С.Ф. СВЧ усилители мощности для мобильной связи и радиодоступа. М.: Горячая линия – Телеком, 2022, 456 с.
2. Ghannouchi F., Hammi O., Helaoui M. Behavioral modelling and predistortion of wideband wireless transmitters. John Wiley & Sons Ltd, 2015, 250 p.
3. Schetzen M. The Volterra and Wiener Theories of Nonlinear Systems. M. Schetzen - Malabar: Krieger Pub Co, 2006. 595 p.
4. Горгадзе С.Ф., Максимов А.А. Модели регрессии на основе ряда Вольтерры и исследование нелинейных искажения сигналов с большим пик-фактором // Системы синхронизации, формирования и обработки сигналов. 2021. Т.12. №1. С. 13-21.
5. Горгадзе С.Ф., Пилюгин И.С. Самонастраивающийся алгоритм линеаризации характеристики усилителя мощности на основе LUT // REDS: Телекоммуникационные устройства и системы. 2020. Т.10. №2. С. 34-39.

6. Горгадзе С.Ф. Исследование поведенческих моделей усилителей мощности на основе ограничения динамического диапазона девиации ряда Вольтерры // В сборнике: Технологии информационного общества. Материалы XIII отраслевой международной научно-технической конференции. 2019. С. 264-267.
7. Горгадзе С.Ф. Выбор поведенческой модели нелинейного твердотельного усилителя мощности // Системы синхронизации, формирования и обработки сигналов. 2018. Т.9. №1. С. 54-58.
8. Горгадзе С.Ф., Максимов А.А. Теория гармонического баланса для схемотехнического проектирования // Т-Comm: Телекоммуникации и транспорт. 2020. Т.14. №11. С. 21-32.
9. Haber R., Keviczky L. Nonlinear Structures for System Identification (Volterra, Wiener and Hammerstein Models for Dynamic Systems) // Periodica Polytechnica, Electronical Engineering. 1974. Vol. 18. № 4, pp. 393-404.
10. Gilbert P. L., Montoro G., Bertran E. On the Wiener and Hammerstein Models for Power Amplifier Predistortion // Asia-Pacific Microwave Conference Proceedings. 2005. Vol. 2, pp. 4-7.
11. Liu T., Boumaiza S. and Ghannouchi F.M. Deembedding static nonlinearities and accurately identifying and modeling memory effects in wide-band RF transmitters // IEEE Transactions on Microwave Theory and Techniques. 2005. 53 (11), pp. 3578-3587.

**NONLINEAR DISTORTION COMPENSATION MODELING IN A POWER AMPLIFIER
BASED ON ITS DIGITAL BEHAVIORAL MODEL**

Anton A. Maksimov,
Postgraduate MTUCI, Moscow, Russia,
mash2525@yandex.ru

Igor V. Komarov,
Graduate MTUCI, Moscow, Russia,
ivk1406@mail.ru

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

Abstract

A behavioral model of a power amplifier based on a regression equation has been developed, which makes it possible to decompose a distorted signal at the amplifier output into an undistorted component and an additive set of non-linear distortion noise. It has been established that such interference consists of inertial non-linear components and inclusions corresponding to the manifestation of memory effects. It is shown that the correction of only non-inertia components of the interference of nonlinear distortions makes it possible to effectively deal with out-of-band emissions, and the compensation of memory effects - with in-band interference. It has been established that with the maximum possible compensation of the inertialess component of the interference, only the compensation of the memory effects also makes it possible to provide the necessary suppression of the interference of non-linear distortions.

Keywords: *power amplifier, behavioral model, regression equation, Kim and Konstantinu model, non-linear distortion compensation, digital pre-distortion, out-of-band interference, in-band interference.*

ОБЗОР ПОСТРОЕНИЯ СОЛНЕЧНЫХ ПАНЕЛЕЙ ДЛЯ СПУТНИКОВ ТИПА CUBESAT

Пронина Евгения Дмитриевна,
старший преподаватель кафедры СиСРТ
МТУСИ, Москва, Россия,
e.d.pronina@mtuci.ru

Свирская Анастасия Владимировна,
студент МТУСИ, Москва, Россия,
a.v.svirskaya@mtuci.ru

Аннотация

Спутники типа CubeSat становятся важной частью освоения космоса, как в области научных исследований, так и в области промышленного использования околоземного пространства. Спутники такого типа становятся всё более востребованными благодаря дешевизне и простоте в изготовлении в сравнении со спутниками других типов. Важной частью любого спутника является энергосистема. Оптимальным вариантом энергообеспечения спутника является использование солнечных панелей. В данной статье рассматриваются вопросы, связанные с выбором, установкой и использованием солнечных панелей на спутнике типа CubeSat.

Ключевые слова: наноспутники, CubeSat, солнечные панели, системы электропитания, хранение энергии, автономность.

Спутники типа CubeSat

Спутник типа CubeSat – искусственный спутник земли, имеющие малый вес и размер, в сравнении со спутниками других типов. Стандартный CubeSat имеет размеры всего 10x10x10 см, и называется CubeSat с конфигурацией 1U (рис. 1). Наноспутник можно масштабировать до 2U (20x10x10 см), 3U (30x10x10 см), и максимально возможная конфигурация на данный момент состоит из 24U. Вес единицы измерения 1U ограничен 1,33 кг., а, например, 3U – 4 кг. [1-7].

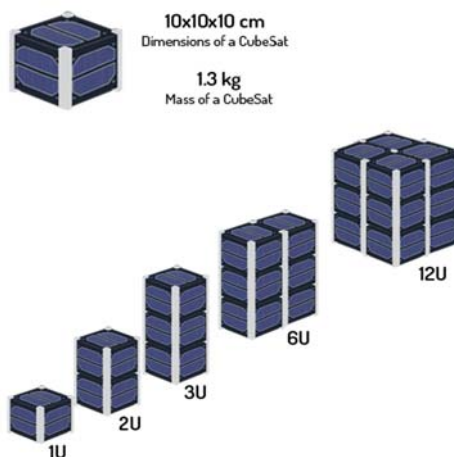


Рис. 2. Конфигурации спутников типа CubeSat

Малые космические аппараты такого типа могут применяться для:

1. Наладживания систем связи;
2. Изучения космического пространства;
3. Дистанционного Зондирования Земли (ДЗЗ);
4. Всепогодного наблюдения;
5. Образовательных и других целей.

В качестве полезной нагрузки на наноспутнике могут быть расположены исследовательские приборы, заменяемые в различных сферах науки, например для изучения температуры, плотности и состава окружающего

вещества, электрического и магнитного полей гравитации, а также для излучения в различных диапазонах радиации, солнечного и космического ветра, различных свечений в верхних слоях атмосферы Земли.

Как правило, каждый спутник типа CubeSat состоит из каркаса, антенн, датчиков температуры, акселерометров, гироскопа, дозиметра, модуля ГНСС, камер и т.д.

Автономность спутника напрямую связана с самообеспечением энергией и поэтому вопрос энергоснабжения весьма актуален для современного спутникостроения.

Солнечные панели для CubeSat

Требования полезной нагрузки к потребляемой средней и пиковой электрической мощности вместе с профилем орбиты спутника являются важными факторами, влияющими на размеры солнечной батареи и аккумуляторного блока. Чтобы система электроснабжения могла выполнять все свои функции [8] также необходимо учитывать потребляемую спутником мощность как в начале срока его службы, так и в конце.

Система электроснабжения должна обладать следующими основными функциями:

1. Регулирование источника питания,
2. Накопление энергии,
3. Распределение электроэнергии.

Вторичные функции системы:

1. Телеметрические измерения и кондиционирование,
2. Управление системой развертывания антенны,
3. Защита от перегрузки по току [9].

Учитывая ограничения по размерам и весу спутников типа CubeSat, единственным долговечным и стабильным вариантом для получения первичной энергии является солнечная фотоэлектрическая система (рисунок 2), так как размеры наноспутника не подразумевают включение в полезную нагрузку топлива или объемных батарей.

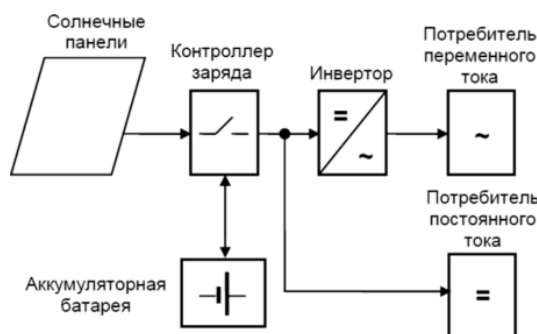


Рис. 3. Солнечная фотоэлектрическая система спутника

Спутники типа CubeSat используют солнечные элементы для преобразования солнечного света в электроэнергию, которая затем сохраняется в перезаряжаемых литий-ионных батареях (рис. 3), обеспечивающих питание во время затмения, а также в периоды пиковой нагрузки [10].

В отличие от более крупных спутников, у которых площадь поверхности, доступная на внешних конструкциях намного больше, наноспутники CubeSat имеют гораздо меньшую площадь, которую можно использовать под солнечные батареи. Кроме того, солнечные панели не должны располагаться на местах, отведенных для других компонентов, таких как планарные антенны, оптические датчики, объективы камер и др. В настоящее время на рынке источников питания представлен широкий спектр солнечных панелей и разнообразных солнечных элементов, которые совместимы со спутниками типа CubeSat [11].



Рис. 4. Батареи для спутника типа CubeSat

Фотогальванический элемент

Все солнечные панели состоят из фотогальванических элементов. Каждый фотогальванический элемент (рис. 4) состоит из двух пластин полупроводникового материала. Фотоэлектрические элементы обычно изготавливаются из кремния — материала, который чаще всего используется в микроэлектронике, так как он может проводить ток, сохраняя при этом электрический дисбаланс, необходимый для создания электрического поля.

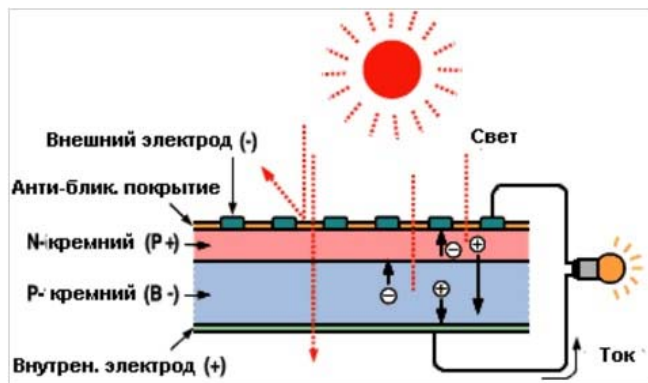


Рис. 5. Фотогальванический элемент

Когда солнечный свет попадает на полупроводник в фотоэлектрическом элементе, энергия света в виде фотонов поглощается, выбивая некоторое количество электронов, которые затем свободно дрейфуют в элементе. Солнечная батарея состоит из положительно и отрицательно заряженных полупроводниковых пластин, соединенных вместе для создания электрического поля. Это электрическое поле заставляет дрейфующие электроны двигаться в определенном направлении - к проводящим металлическим пластинам, выстилающим ячейку. Этот энергетический поток и сила тока определяют количество электроэнергии производимой каждой ячейкой. Как только свободные электроны попадают на металлические пластины, ток направляется в соединительные провода подобно любым источникам генерации электроэнергии [12].

Солнечные панели

Большинство солнечных элементов изготавливаются из полупроводниковых мультикристаллических кремниевых элементов, эффективность которых в настоящее время составляет от 10 до 15%. Самая большая проблема, с которой сталкиваются системы, работающая на солнечных панелях это их низкая эффективность и высокая стоимость.

В настоящее время применяются четыре типа солнечных панелей (рис. 5):

1. Монокристалл;
2. Поликристалл;
3. Тонкая пленка (имеет низкую эффективность);
4. Прозрачная солнечная панель (пока не распространена).

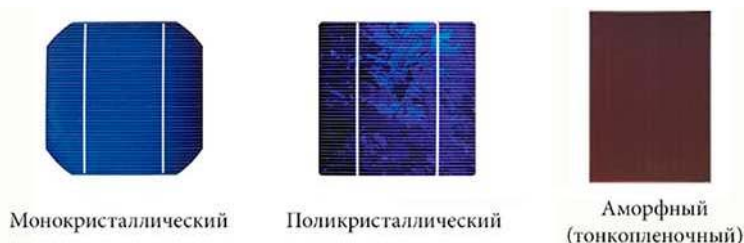


Рис. 6. Типы солнечных панелей

Самые распространенные группы панелей – это монокристаллические и поликристаллические панели.

В монокристаллических солнечных панелях солнечные элементы представляют собой единый кристалл чистого кремния. Для формирования из них модуля используют метод постепенного выращивания кристалла кремния из расплава. В качестве «затравки» используют маленький кристалл кремния, который в последствие обрастает чистым кремнием. Модули, созданные таким образом, имеют однородный цвет и текстуру.

Поликристаллические солнечные панели изготавливаются на основе солнечных элементов из множества кристаллов. В данном случае кристаллическая «затравка» опускается в расплавленный кремний. Когда вся конструкция остывает, внутри её формируется множество небольших разнонаправленных кристаллов. Затем такие

слитки кристаллов нарезаются на тонкие пластины, которые в дальнейшем используются в изготовлении солнечных панелей.

В производстве наноспутников, как правило, применяются солнечные панели на основе монокристаллов, так как они обладают следующими преимуществами:

1. Имеют более высокий КПД по сравнению с поликристаллическими (22% против 12%);
2. Монокристаллические модули более компактны;
3. Такие модули служат гораздо дольше, так как монокристалл обладает стабильными характеристиками, которые не изменяются с течением времени;
4. Монокристаллические модули при низкой освещенности работают лучше, чем поликристаллические.

Критерии выбора солнечных панелей

Помимо размера самих панелей, есть несколько важных критериев, которые следует учитывать при выборе спутниковой солнечной панели или массива панелей:

1. Генерация энергии в соответствии с выбором орбиты и полезной нагрузки;
2. Соблюдение допустимых уровней напряжения и тока системой кондиционирования и распределения электроэнергии внутри спутника;
3. Соответствие в габаритах выпускающего модуля ракет-носителей и самого спутника и нагрузкам, заявленных в техническом описании;
4. Совместимость с температурным диапазоном, с которым может столкнуться спутник при запуске и на орбите;
5. Совместимость со всеми материалами, используемыми при построении самих солнечных панелей и материалов спутника, и характером их поведения из-за теплового расширения и сжатия при запуске и на орбите;
6. Надежное срабатывание при разворачивании панелей (доступно несколько вариантов разворачивания, таких как гибкие соединения, сплавы с памятью формы, торсионная пружина в петле, винтовые пружины, двигатели и т. д.);
7. Способность выдерживать физические нагрузки при запуске.

Наряду с местоположением запуска, бюджетом и аналогичными коммерческими факторами, влияющими на стоимость спутника в целом и солнечных панелей в частом, эти критерии следует использовать для оценки того, какая система лучше всего подойдет для космической миссии и временных рамок работы спутника.

Архитектуры системы электропитания с использованием солнечных панелей

Электроэнергетическая система отвечает за непрерывную выработку и поставку электроэнергии на одном нерегулируемом уровне напряжения.

Солнечные элементы либо батарея, должны обеспечивать электропитание спутника, если соответствующий другой источник электропитания в определённое время не может поддерживать энергообеспечение спутника. Это подразумевает четыре режима работы системы электропитания спутника (табл. 1).

Таблица 1

Режимы работы системы электропитания спутника

		Батарея	
		<i>работает</i>	<i>не работает</i>
Солнечные панели	Генерация энергии	Высокая мощность	Только солнечная энергия
	Нет генерации	Питание только от батареи	Нет энергии

Используются две архитектуры системы электропитания, наиболее часто применяемые при проектировании наноспутников.

1. Система с использованием отслеживания максимальной пиковой мощности – Maximum Power Point Tracking (MPPT) (рис. 6);

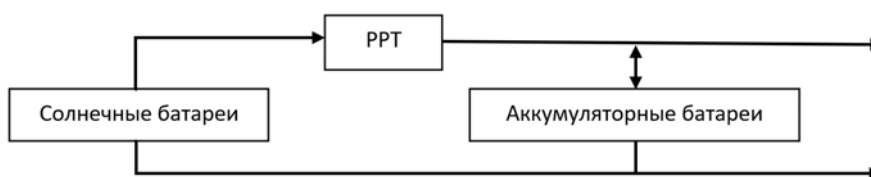


Рис. 7. Система MPPT

2. Система прямой передачи энергии от солнечных батарей - Direct Energy Transfer (DET) (рис. 7).

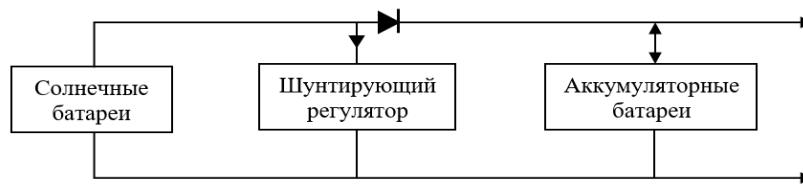


Рис. 8. Система DET

Система МРРТ [13] использует коммутационный преобразователь для управления рабочей точкой солнечной батареи таким образом, чтобы солнечная батарея всегда фиксировала идеальную нагрузку, что делает передачу энергии очень эффективной. МРРТ реализуется путем подачи нерегулируемой шины напряжения по всей энергосистеме, к которой подключены различные подсистемы спутника. К шине также подсоединены солнечные элементы и вторичные батареи.

В отличие МРРТ, DET использует диод в качестве интерфейса между солнечной батареей и ее нагрузкой. Затем нагрузка непосредственно определяет рабочую точку солнечной батареи, которая находится в оптимальной точке максимальной передачи энергии. Интерфейсный диод позволяет подключать несколько солнечных батарей к одной шине аналогично МРРТ. Диод предотвращает обратную подачу напряжения на любую цепь (solar string) при нахождении спутника в тени [14].

Энергетическая система спутника (EPS – Electrical Power System) [15] состоит из трёх типов модулей: солнечный модуль, аккумуляторный модуль и модуль полезной нагрузки. Солнечный модуль отвечает за выработку энергии и передачу ее на общую шину. Аккумуляторный модуль ответственен за накопление и подачу энергии. Модуль полезной нагрузки обеспечивает регулирование и использование энергии. Модульность подсистем обеспечивает масштабируемость, при правильном проектировании системы. Три типа модулей могут быть подключены к общей шине в любом количестве в соответствии с требованиями функционирования спутника. Аккумуляторные модули можно разместить последовательно друг за другом, чтобы увеличить напряжение на общей шине без потери функциональности.

Закрепленные на корпусе и управляемые разворачиваемые солнечные панели

Одним из основных ограничений общей шины спутников типа CubeSat является доступная бортовая мощность. Максимальная мощность, полученная с использованием солнечных панелей, установленных на спутнике, как правило, составляет менее 10 Вт. Производительность спутника типа CubeSat и возможности его использования могут быть значительно улучшены за счет увеличения доступной мощности [16].

На данный момент продолжается поиск технологий и методов эффективного производства и хранения энергии на борту нанокосмических аппаратов за счет эффективного взаимодействия нагрузки и батарей.

Одним из наилучших способов оптимизировать выработку энергии на наноспутнике считается использование разворачиваемых солнечных панелей. Они были разработаны для нанокосмических аппаратов для повышения возможности бортовой выработки электроэнергии. Простейшие системы основаны на одиночных разворачиваемых солнечных батареях, соединенных с основным корпусом спутника одним единственным шарниром. В более сложных конфигурациях система основана на ряде соединенных между собой солнечных панелей. Эти системы основаны на нескольких различных механизмах разворачивания, и некоторые из них были разработаны с учетом конкретных сценариев спутниковой орбиты и стабилизации ориентации.

Увеличение производительности панелей происходит за счет правильного расположения панелей относительно солнца (рис. 8). Максимальная мощность приходится на интервал времени, когда солнечные лучи падают на панель строго ортогонально. В таком случае для 3U CubeSat можно добиться общей мощности от 22 Вт до 56 Вт, повышая общую производительность солнечных панелей с 160% до 400% [17].

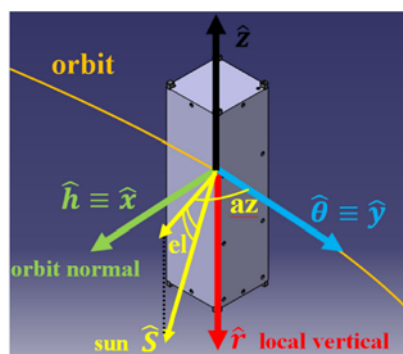


Рис. 9. Система развёртывания солнечных панелей

Размер солнечной панели, подключенной с применением технологии разворачивания, в основном определяется геометрическими ограничениями и размерами спутника типа Cubesat, не предоставляя множества конструктивных вариантов. Следовательно, основной компромисс при проектировании заключается в общей геометрической конфигурации системы солнечных панелей и архитектуры разворачивания. Две возможные конфигурации для компромиссного дизайна показаны на рисунке 9, где представлены варианты продольного и бокового разворачиваний.

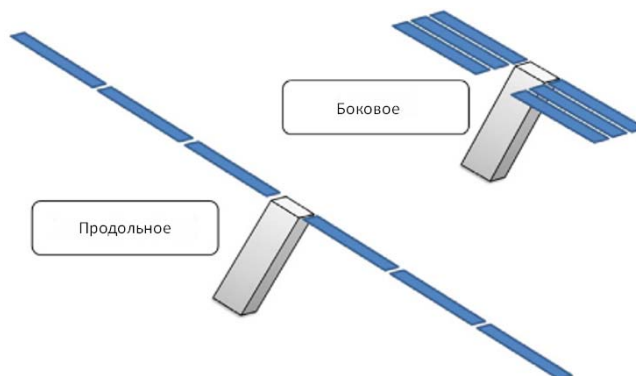


Рис. 10. Варианты разворачиваний солнечных панелей

В конфигурации продольного разворачивания солнечные панели соединяются на стороне, параллельной основному шарниру (к тому, который соединяет систему солнечных панелей с основной конструкцией спутника). В конфигурации бокового разворачивания солнечные панели подключаются на стороне, перпендикулярной основному шарниру.

Продольная конфигурация выбирается как наиболее удачная на основе сравнения с поперечной конфигурацией:

1. Меньший крутящий момент во время разворачивания из-за симметрии;
2. Меньший момент инерции развернутой солнечной панели относительно оси управления;
3. Меньше влияние вращательного момента на ориентацию спутника;
4. Меньшее воздействие солнечной панели на конструкцию в случае неисправности на этапе разворачивания;
5. Больше места для установки вторичных петель;
6. Масштабируемость до кубов разных конфигураций.

Проблемы при использовании солнечных панелей

Использование солнечных панелей для энергообеспечения спутника – один из лучших вариантов питания, но такая система имеет ряд существенных недостатков, над устранением которых работают учёные и инженеры [18]:

1. Расположение на орбите наноспутники, размещенные на низкой околоземной орбите, могут заряжаться и накапливать заряд только 38% времени, что не позволяет реализовать их потенциал на все сто процентов. Вывод спутников на среднюю околоземную орбиту или геостационарную увеличивает эффективность заряда до 50%, но солнечные панели по-прежнему не реализуют свой потенциал полностью.

2. Чувствительность полупроводниковых материалов к радиации. Заряженные частицы и электромагнитное излучение постоянно воздействуют на солнечные элементы, снижая их производительность. Космические лучи также представляют собой проблему из-за высокой ионизационной способности переносить гигаэлектрон-вольты энергии. Когда заряженные частицы ударяются о полупроводниковый материал солнечной батареи, переносимая ими энергия передается атомам материала, смещая их с исходного положения в кристаллической решетке. Смещенные атомы не могут эффективно проводить ток в электрических компонентах системы энергообеспечения из-за снижения проводимости материалов [19].

3. Стоимость установки солнечных панелей является одной из самых дорогих составляющих проектирования спутника.

4. Ограниченные габаритные размеры. У спутников типа CubeSat солнечные панели не могут занимать больше пространства на вынесенной поверхности спутника, чем это позволяет его полезная нагрузка. Из-за этого солнечные панели могут занимать только ограниченную площадь.

5. Вероятность физических повреждений элементов. При разворачивании солнечных панелей существует большая вероятность повреждений. Для ослабления влияния стартовых нагрузок большинство конструктивных решений направлены на повышении жесткости панели за счет применения дополнительных ребер в основании крепления панелей, изготовленных из таких материалов как алюминий и стеклопластик. Также используются панели из армированного углеродным волокном пластика [20].

Заключение

Внедрение и оптимизация энергообеспечения спутников типа CubeSat является одной из важнейших задач на любом из этапов проектирования спутника. Из-за своих малых размеров наноспутники не могут использовать резервуары с топливом или объемных батарей. Поэтому единственным вариантом является использование солнечных панелей, и самыми распространенными являются монокристаллические и поликристаллические панели. Как правило, выбор делается в пользу солнечных панелей на основе монокристаллов, отличающихся рядом технологических преимуществ. Одним из эффективных решений оптимизации энергообеспечения наноспутника является использование разворачиваемых солнечных панелей. Возможны две конфигурации разворачивания панелей: продольное и боковое разворачивания. На основе их сравнения обычно выбирается продольная как наиболее удачная.

Система энергообеспечения спутника с использованием солнечных панелей имеет ряд существенных недостатков, с которыми приходится мириться, в частности: вероятность физических повреждений элементов панели при разворачивании, чувствительность полупроводниковых материалов к радиации и т.д. Тем не менее, солнечные панели остаются самым наилучшим вариантом энергообеспечения для спутников типа CubeSat.

Литература

1. Журнал "Все о Космосе" [Электронный ресурс]. Режим доступа: <https://aboutspacejournal.net/космические-аппараты/искусственный-спутник-земли/малые-спутники/> (Дата обращения: 20.10.2022).
2. Cubesat Space Protocol [Электронный ресурс]. Режим доступа: <https://m.books.ru/books/cubesat-space-protocol-3169452/> (Дата обращения: 20.10.2022).
3. *Satsearch B.V.* [Электронный ресурс] Режим доступа: <https://blog.satsearch.co/2020-04-03-smallsat-and-cubesat-solar-panels-on-the-global-market> (Дата обращения: 20.10.2022).
4. Solar Power Satellites [Электронный ресурс]. Режим доступа: <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/solar-power-satellites> (Дата обращения: 20.10.2022).
5. Журнал "Все о Космосе" [Электронный ресурс]. Режим доступа: <https://aboutspacejournal.net/2016/11/03/наноспутники-cubesat-получат-универсальную/> (Дата обращения: 20.10.2022).
6. Harvard University, The Graduate School of Arts and Sciences [Электронный ресурс]. – Режим доступа: <https://sitn.hms.harvard.edu/flash/2019/future-solar-bright/> 21.03.2019 свободный (Дата обращения: 20.10.2022).
7. Solect Energy [Электронный ресурс]. Режим доступа: <https://solect.com/the-science-of-solar-how-solar-panels-work/> (Дата обращения: 20.10.2022).
8. *Jean Bester, Ben Groenewald, Richardt Wilkinson.* Cape Peninsula University of Technology Electrical power system for a 3U CubeSat nanosatellite incorporating peak power tracking with dual redundant control // *Przegląd Elektrotechniczny (Electrical Review)*, ISSN 0033-2097, R. 88 NR 4a/2012.
9. *Kamel Djamel Eddine Kerrouche, Abderrahmane Seddjar, Nassima Khorchef.* CubeSat project: experience gained and design methodology adopted for a low-cost Electrical Power System DOI:10.1080/00051144.2022.2065420.
10. *Sergio Sanchez-Sanjuan, Jesus Gonzalez-Llorente, Ronald Hurtado-Velasco.* Comparison of the Incident Solar Energy and Battery Storage in a 3U CubeSat Satellite for Different Orientation Scenarios // *Journal of Aerospace Technology and Management*, Jan-Mar 2016.
11. Smallsat and CubeSat solar panels on the global market [Электронный ресурс]. Режим доступа: <https://blog.satsearch.co/2020-04-03-smallsat-and-cubesat-solar-panels-on-the-global-market> (Дата обращения: 20.10.2022).
12. Ваш Солнечный Дом [Электронный ресурс]. Режим доступа: <https://www.solarhome.ru/basics/solar/pv/mono-or-poly-solar-panels.htm> (Дата обращения: 20.10.2022).
13. *Daniel F. Butay, Michael T. Miller.* Maximum Peak Power Tracker: A Solar Application. A Major Qualifying Project Report: April 24th, 2008. С 6-7.
14. *Timothy M. Lim.* A modular electrical power system architecture for small spacecraft University of Kentucky, 2016. С. 12-15.
15. Смердова И.О. Система электропитания малого космического аппарата. УДК 621.311.6:629.78-022.51. Томск, 2019.
16. *FabioSantonia, FabrizioPiergentilib, SerenaDonati.* An innovative deployable solar panel system for Cubesats, *Acta Astronautica*, February-March 2014, pp. 210-217.
17. *FabioSantonia, FabrizioPiergentilib, Gian PaoloCandini.* An orientable solar panel system for nanospacecraft, *Acta Astronautica* Volume 101, August-September 2014, pp. 120-128.
18. *Shankar Bhattarai, IHongrae Kim, 2and Hyun-Ung Oh.* CubeSat's Deployable Solar Panel with Viscoelastic Multilayered Stiffener for Launch Vibration Attenuation, *International Journal of Aerospace Engineering*, 11 Aug 2020.
19. Greener Energy Group [Электронный ресурс]. Режим доступа: <https://greenerenergygroup.co.uk/solar-advantages-and-disadvantages/> (Дата обращения: 20.10.2022).
20. Wikipedia, the free encyclopedia [Электронный ресурс]. Режим доступа: https://en.wikipedia.org/wiki/Solar_panels_on_spacescraft (Дата обращения: 20.10.2022).

AN OVERVIEW OF BUILDING SOLAR PANELS FOR CUBESAT SATELLITES

Evgenia D. Pronina,

*Senior Lecturer of the Department of S&SRT
MTUCI, Moscow, Russia,
e.d.pronina@mtuci.ru*

Anastasia V. Svirskaya,

*Student MTUCI, Moscow, Russia,
a.v.svirskaya@mtuci.ru*

Annotation

Satellites such as CubeSat are becoming an important part of space exploration in both scientific industries and global industry. Satellites of this type are gaining more and more popularity due to their low cost and ease of manufacture relative to other types of satellites. An important part of any satellite is its power system. The best way to power the satellite is to use solar panels. This article discusses issues related to the selection, installation and use of solar panels on a CubeSat satellite.

Keywords: *nanosatellites, CubeSat, solar panels, power supply systems, energy storage, autonomy.*

МЕТОД РАСПРЕДЕЛЕНИЯ ВЫБОРКИ ЦВЕТОВЫХ ОТТЕНКОВ В ЦВЕТОВОМ ПРОСТРАНСТВЕ CIELCH

Иванчев Василий Витальевич,
аспирант МТУСИ, Москва, Россия,
demidra18@gmail.com

Власюк Игорь Викторович,
доцент кафедры ТуЗВ МТУСИ,
к.т.н., доцент, Москва, Россия,
ru3dlp@yandex.ru

Аннотация

В статье представлена разработка метода распределения выборки цветowych оттенков в цветовом пространстве CIELCH. Метод учитывает специфические требования, обусловленные дальнейшей работой с выборкой: соблюдение ограничения на количество цветowych оттенков в выборке (в данном случае представлены точками в пространстве CIELCH) при максимально возможной плотности и равномерности распределения точек. Для увеличения количества точек предлагается использовать двумерную интерполяцию промежуточных результатов для точек, находящихся на одной оси.

Ключевые слова: колориметрия, цветowe пространства, CIELCH, интерполяции, теплoхолодность.

Введение

В настоящее время в разных отраслях промышленности широко используются промышленные системы телевидения, решающие задачу контроля качества выпускаемой продукции. Данная задача вынуждает строить системы телевидения, параметры которой слабо согласованы с параметрами зрительной системы человека. Например, в системах рентгенографического телевидения, динамический диапазон порядка 10000 к уровню шума [1,2]. В болометрических системах телевидения – 30500-61000 к уровню шума [3,4]. При этом, человек видит только порядка 200 градаций яркости, что не дает воспринимать полноценно изображение, полученное с систем промышленного телевидения без обработки [5,6].

Один из методов адаптации изображений с большим динамическим диапазоном для восприятия человеком – это цветowe контрастирование. Метод заключается в замене уровней яркости в изображении на определенные цвета, которые можно задать либо с помощью определенной функции (метод обхода цветowego пространства), либо с помощью тепловой карты, самостоятельно соотнеся определенные яркости к определенным оттенкам. Цветowe контрастирование в настоящее время широко используется в болометрических и рентгенографических системах [7-10].

Контекст данной работы заключается в создании метода цветowego контрастирования, основанного на восприятии человеком теплоты цвета. Для создания данного метода необходимо сделать исследование восприятия теплоты цвета для цветовой модели CIELCH. Исследование предлагается проводить следующим методом. Сначала формируется выборка цветов и проводится опрос восприятия их теплоты. Далее необходимо интерполировать результаты для остальных цветов тела цветовой модели. Таким образом, точки на цветовом пространстве, являющиеся выборкой цветов, будут узлами последующей интерполяции.

Целью данной статьи является решение задачи оптимального распределения точек на цветовом пространстве CIELCH с учётом дальнейшей работы с цветовым пространством CIELCH.

Разработка метода распределения точек в цветовом пространстве CIELCH

Для интерполяции необходимо задать множество узлов интерполяции в пространстве. Выбор точек играет ключевую роль для точности интерполяции. Для максимально возможной точности интерполяции необходимо, чтобы узлы интерполяции находились, как можно ближе друг к другу и были распределены равномерно в пространстве. Однако, при выборе способа распределения точек в пространстве стоит учесть несколько специфических для поставленной задачи условий.

Исходя из того, что узлы интерполяции, определяемые в данной научной работе, являются выборкой цветowych оттенков для проведения тестового испытания, количество узлов интерполяции сильно ограничено. При составлении тестового испытания необходимо учитывать требования, предъявляемые к времени, которое это испытание занимает. Согласно ITU BT.500-14 [11], время тестового испытания не должно превышать 30 минут для того, чтобы уменьшить ошибки, вызванные усталостью и прочими человеческими факторами.

С учетом того, что тестовое испытание представляет собой сравнение каждого оттенка из выборки с остальными, то при 36 оттенках количество элементов тестового испытания составит:

$$K = \sum_1^{35} n = 630, \tag{1}$$

где K – количество элементов тестового испытания.

При этом предполагаемое время прохождения одного элемента тестового испытания составляет 2-3 секунды. то есть общее время тестового испытания при таких условиях равно:

$$T = 630 * 3 = 1890 \text{сек} = 31,5 \text{мин}$$

Что немногим больше времени, требуемого ITU BT.500-14. Количество оттенков (36) было подобрано эмпирически, исходя из представленных выше вычислений.

Из [12,13] известно, что существует зависимость теплохолодности цвета от его яркости: чем цвет ярче, тем он холоднее.

Для более подробного исследования этой зависимости можно задать шаблон точек, который будет повторяться на разных секущих плоскостях. При этом есть возможность не повторять шаблон в каждой секущей плоскости, а в некоторых плоскостях устанавливать значение теплохолодности точки в шаблоне благодаря двумерной интерполяции точек по оси L, имеющих одинаковые значения по осям С и Н (рис. 1):

$$L(z) = \sum_{i=0}^n y_i * l_i(z), \quad l_i(z) = \prod_{j=0, i \neq j}^n \frac{z - z_j}{z_i - z_j} \tag{2}$$

Предположим, что значения восприятия человеком теплоты цвета для точек 1, 2, 4, 5 из рисунка 1 будут следующими (табл. 1):

Таблица 1

Условные значения восприятия теплохолодности цветowych оттенков для ряда точек

Номер точки	Значения восприятия теплоты цвета
1	0.8
2	0.65
4	0.42
5	0.14

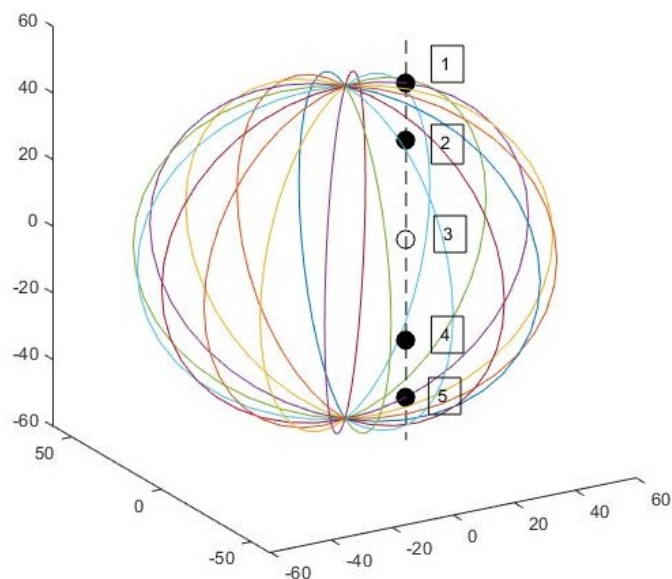


Рис. 1. Цветовая модель CIE LCH с точками, лежащими на одной прямой, параллельной оси Z

Тогда можно восстановить точку 3 с помощью интерполяции (рис. 2).

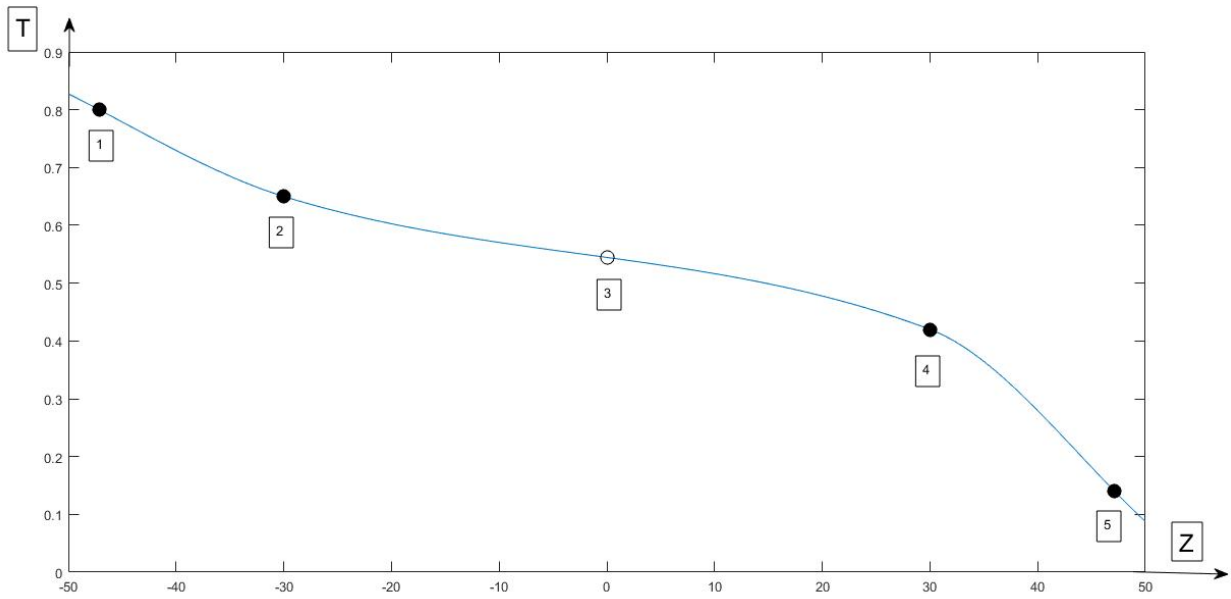


Рис. 2. Интерполяция условных значений восприятия теплоты цвета для точек 1, 2, 4 и 5. (точка 3 восстановлена с помощью интерполяции)

Часть узлов интерполяции необходимо расположить на внешней границе пространства CIELCH. Пространство CIELCH представляет собой эллипсоид вращения [14,15]. Горизонтальные сечения этого эллипсоида будут формировать окружности, а вертикальные – эллипсы (рис. 3 и 4).

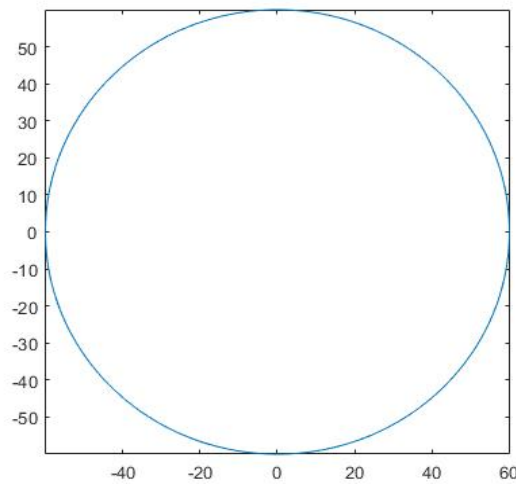


Рис. 3. Горизонтальная секущая CIELCH на уровне яркости 50

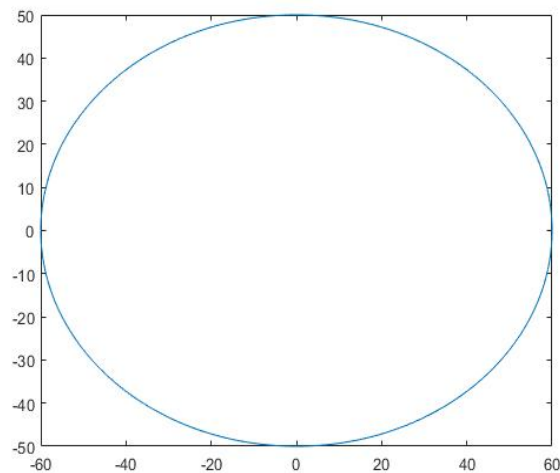


Рис. 4. Вертикальная секущая CIELCH

При расположении точек по шаблону на горизонтальных секущих, получатся точки, расположенные на окружностях, сформированных этими секущими (рис. 5). Точки, лежащие внутри пространства также можно расположить по условным окружностям (рис. 6).

В данной научной работе предлагается формирование трех кругов: первый круг с радиусом 60 (внешняя граница CIELCH), второй – с радиусом 48 и третий круг с радиусом 20 (рис. 7).

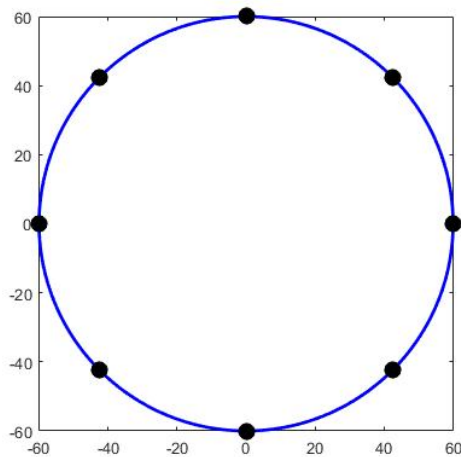


Рис. 5. Горизонтальная секущая CIELCH с 8 равномерно распределенными точками на окружности

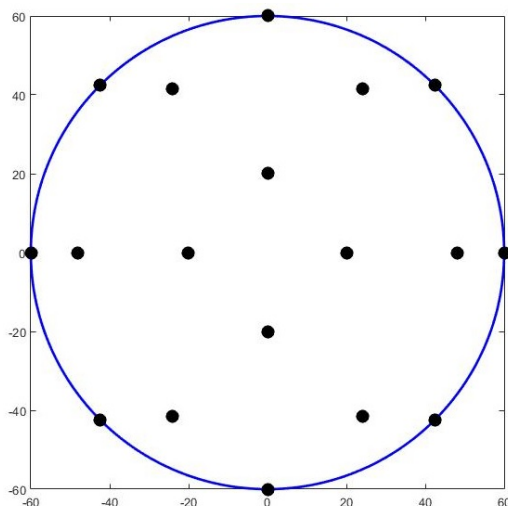


Рис. 6. Горизонтальная секущая CIELCH с точками на окружности и внутри окружности

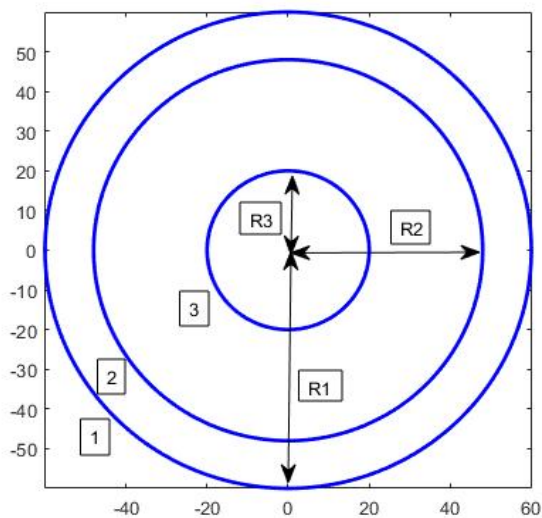


Рис. 7. Круги, на которых располагаются точки: $R_1 = 60$; $R_2 = 48$; $R_3 = 20$

Узлы интерполяции, находящиеся ближе к внешнему краю эллипсоида вращения должно быть больше, так как длина круга на внешней границе, образованная сечением эллипсоида вращения больше, чем длина условного круга, включающего в себя узлы интерполяции, которые лежат внутри эллипсоида вращения.

Данное расположение точек в шаблоне обусловлено тем, что длина внешнего круга будет равна:

$$l_1 = \pi R_1^2 = 11307,6; \tag{3}$$

длина круга №2: $l_2 = 7236,864$

длина круга №3: $l_3 = 1256,4$

Для равномерного расположения точек необходимо их расположить равномерно на окружности так, чтобы расстояние между точками для каждого круга была примерно одинаковой.

В данной научной работе предлагается следующее расположение точек по шаблону: для внешней границы круга – 8 точек, для второго круга – 6 точек, для третьего круга – 4 точки (рис. 8).

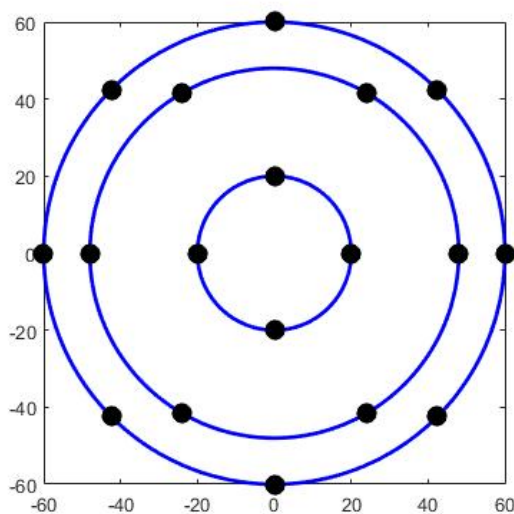


Рис.8. Расположение точек внутри горизонтальной секущей с внутренними кругами

Координаты точек в полярной и декартовой системах координат приведены соответственно в таблицах 2 и 3.

Таблица 2

Координаты точек в полярной системе координат

ρ	ϕ°	Обозначения в системе HEX
60	0	#CE3F7A
60	45	#C05530
60	90	#978319
60	135	#2D8825
60	180	#23AB96
60	225	#1EA8C8
60	270	#157FE0
60	315	#A05AC0
48	0	#BE5079
48	60	#A8662F
48	120	#5D8129
48	180	#249B88
48	240	#1F93C1
48	300	#826ABE
20	0	#976A78
20	90	#807655
20	180	#4D8176
20	270	#687899

Координаты точек в декартовой системе координат

X	Y	Обозначения в системе HEX
-60	0	#CE3F7A
-42,43	42,43	#C05530
0	60	#978319
42,43	42,43	#2D8825
60	0	#23AB96
42,43	-42,43	#1EA8C8
0	-60	#157FE0
-42,43	-42,43	#A05AC0
-48	0	#BE5079
-24	41,57	#A8662F
24	41,57	#5D8129
48	0	#249B88
24	-41,57	#1F93C1
-24	-41,57	#826ABE
-20	0	#976A78
0	20	#807655
20	0	#4D8176
0	-20	#687899

Обозначения в системе HEX требуются для web-программирования, с помощью которого будет составлено тестовое испытание.

При воспроизведении шаблона на секущих, проходящих через уровни яркости, отличающиеся от 50, точки, которые лежали на внешней границе, выйдут за пределы эллипсоида вращения (рис. 9). Целесообразно выбирать такие уровни яркости, чтобы точки, лежащие на внутренних кругах центральной секущей стали точками, которые лежат на внешнем круге.

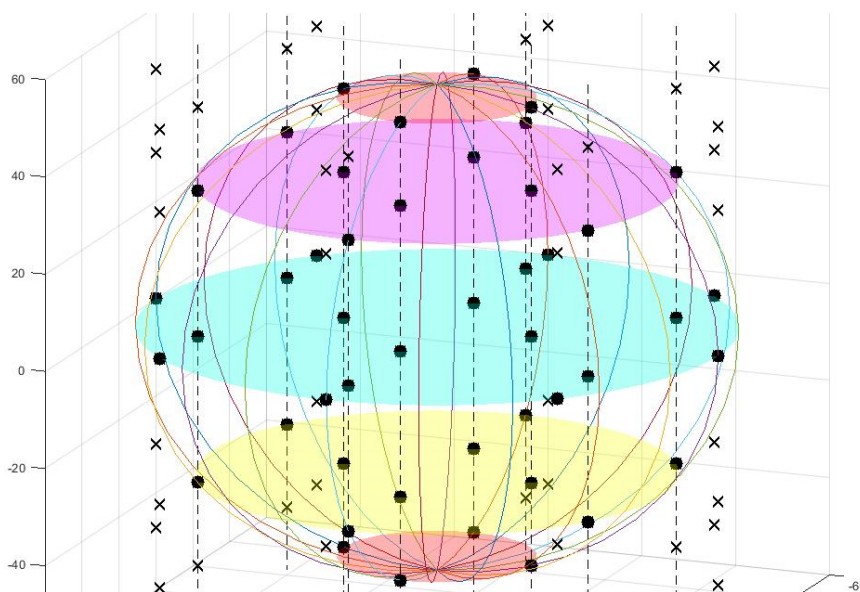


Рис. 9. Иллюстрация метода расположения точек с помощью повторяющегося шаблона

Знаком x обозначены точки, не попадающие в цветное пространство CIELCH.

Так как фигура цветовой модели CIE LCH представляет из себя эллипсоид вращения, то вертикальные секущие будут образовывать эллипсисы. Для расчета необходима вертикальная секущая, проходящая через центр эллипсоида (рис. 4).

При этом, ось X эллипсиса будет соответствовать проекции горизонтальной секущей, а ось Z – проекции перпендикулярной вертикальной секущей. Координаты точек по оси X и Y известны (в данной секущей ось Y не играет роли для расчетов). Необходимо найти координаты для оси Z. Задача сводится к нахождению проекции точек с оси X на эллипс (рис. 10). Для остальных точек, не лежащих в рассматриваемой секущей, значения координат по оси Z будут соответствующими с найденными.

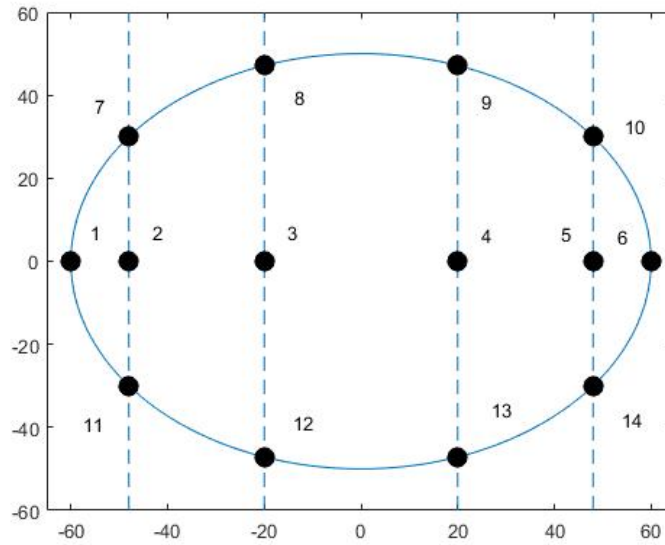


Рис. 10. Поиск точек, лежащих на границе цветового пространства CIELCH

Уравнение эллипса:

$$\frac{x^2}{R_1^2} + \frac{z^2}{R_2^2} = 1 \quad (4)$$

Подставим вместо x значения координаты для точки 5:

$$\frac{48^2}{60^2} + \frac{z^2}{50^2} = 1$$

Получим значения по оси Z для точек 10 и 14.

$$z = \pm 30$$

Подставим вместо x значения координаты для точки 4:

$$\frac{20^2}{60^2} + \frac{z^2}{50^2} = 1$$

Получим значения по оси Z для точек 9 и 13:

$$z = \pm 47,14$$

Для точек 7, 8, 11 и 13 значения по оси Z будут соответствующими.

Можно соединить точки 7 и 10, 8 и 9, 11 и 14, 12 и 13, (рис. 11):

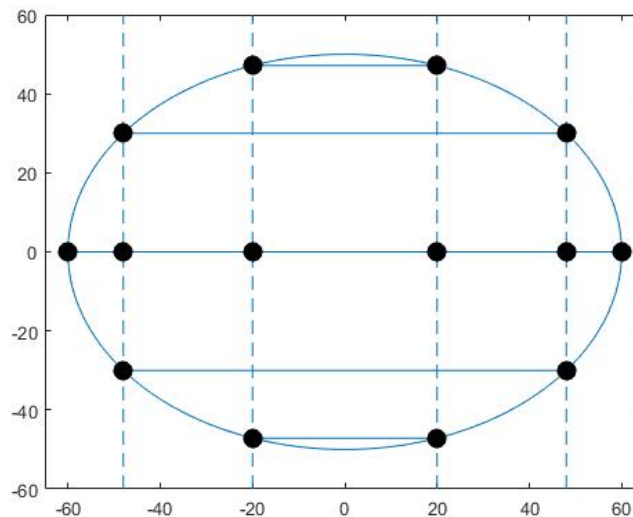


Рис. 11. Вертикальная секущая CIELCH с обозначением ряда горизонтальных секущих

Таким образом, получаем уровни яркости для секущих плоскостей пространства CIELCH. Этими уровнями будут: 47,14, 30, 0, -30, -47,14. Исходя из этого, можно расположить шаблоны на указанных уровнях яркости, убрав точки, выходящие за пределы цветового пространства. Расположение точек будет выглядеть следующим образом (рис. 12):

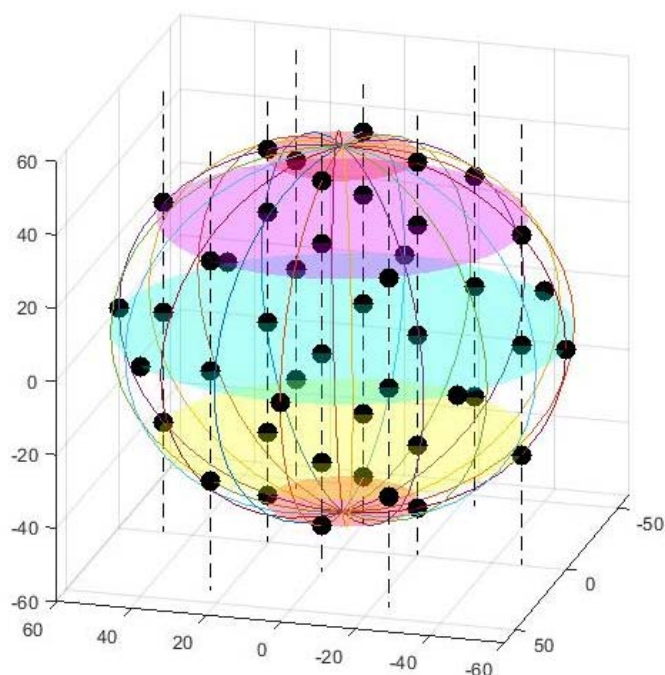


Рис. 12. Цветовое пространство CIELCH с 46 точками, распределенными по предложенному шаблону

В таблице 4 приведены координаты 46 точек из рисунка 12.

Таблица 4

Координаты 46 точек из рисунка 12

z	x	Y	Обозначения в системе HEX
0	-60	0	#CE3F7A
	-42,43	42,43	#C05530
	0	60	#978319
	42,43	42,43	#2D8825
	60	0	#23AB96
	42,43	-42,43	#1EA8C8
	0	-60	#157FE0
	-42,43	-42,43	#A05AC0
	-48	0	#BE5079
	-24	41,57	#A8662F
	24	41,57	#5D8129
	48	0	#249B88
	24	-41,57	#1F93C1
	-24	-41,57	#826ABE
	-20	0	#976A78
	0	20	#807655
20	0	#4D8176	
0	-20	#687899	
30	-48	0	#5F1A36
	-24	41,57	#522E11
	24	41,57	#2D3E15
	48	0	#1A3A34
	24	-41,57	#1B3949
	-24	-41,57	#1D1255
	-20	0	#2B121A
	0	20	#1F1B0C
20	0	#111C19	
0	-20	#851749	

40	-20	0	#2B121A
	0	20	#1F1B0C
	20	0	#111C19
	0	-20	#851749
-30	-48	0	#5F1A36
	-24	41,57	#522E11
	24	41,57	#2D3E15
	48	0	#1A3A34
	24	-41,57	#1B3949
	-24	-41,57	#1D1255
	-20	0	#2B121A
	0	20	#1F1B0C
	20	0	#111C19
	0	-20	#851749
47,14	-20	0	#2B121A
	0	20	#1F1B0C
	20	0	#111C19
	0	-20	#851749

Однако, как видно из рисунка 12, всего точек получилось 46, а в тестовом испытании желательно, чтобы цветовых оттенков было не больше 36. Если учесть зависимость между светлотой и теплохолодностью, можно часть точек убрать из выборки. Как указывалось ранее, значения теплохолодности для этих оттенков можно вычислить с помощью интерполяции.

Обсуждение результатов

Таким образом, окончательный вариант расположения точек на цветовом пространстве CIELCH будет следующим (рис. 13):

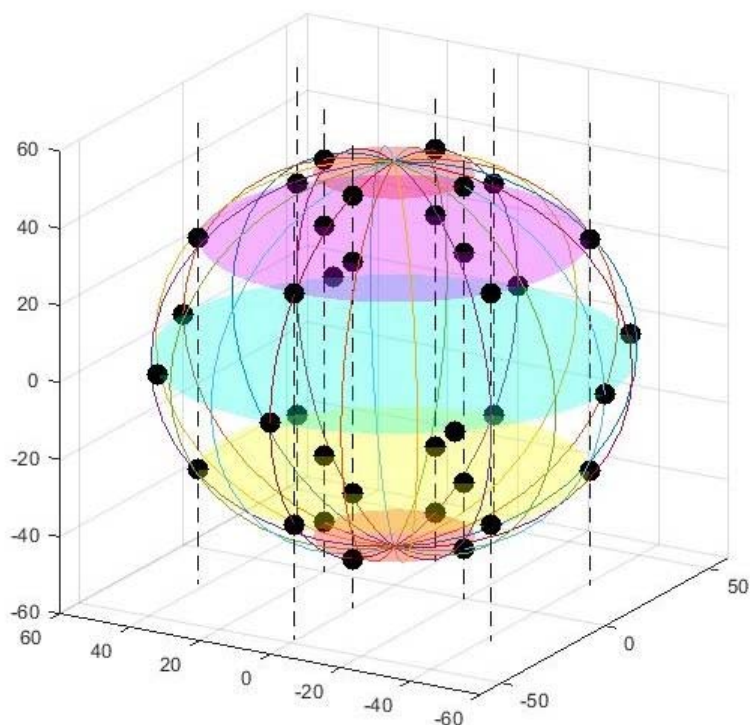


Рис. 13. Цветовое пространство CIELCH с 36 точками, распределенными по предложенному шаблону

В таблице 5 приведены координаты точек (36) к рисунку 13.

Координаты 36 точек из рисунка 13

z	x	Y	Обозначения в системе HEX
0	-60	0	#CE3F7A
	-42,43	42,43	#C05530
	0	60	#978319
	42,43	42,43	#2D8825
	60	0	#23AB96
	42,43	-42,43	#1EA8C8
	0	-60	#157FE0
30	-42,43	-42,43	#A05AC0
	-48	0	#5F1A36
	-24	41,57	#522E11
	24	41,57	#2D3E15
	48	0	#1A3A34
	24	-41,57	#1B3949
	-24	-41,57	#1D1255
	-20	0	#2B121A
47,14	0	20	#1F1B0C
	20	0	#111C19
	0	-20	#851749
	-20	0	#2B121A
-30	0	20	#1F1B0C
	20	0	#111C19
	0	-20	#851749
	-20	0	#2B121A
	-48	0	#5F1A36
	-24	41,57	#522E11
	24	41,57	#2D3E15
	48	0	#1A3A34
-47,14	24	-41,57	#1B3949
	-24	-41,57	#1D1255
	-20	0	#2B121A
	0	20	#1F1B0C

Как видно из рисунка 13 и таблицы 5, удалось соблюсти требования по количеству и равномерности распределения узлов интерполяции. Метод, предлагаемый в данной работе, позволяет увеличивать количество узлов интерполяции (рис. 14), увеличивая их плотность, что также улучшает качество интерполяции.

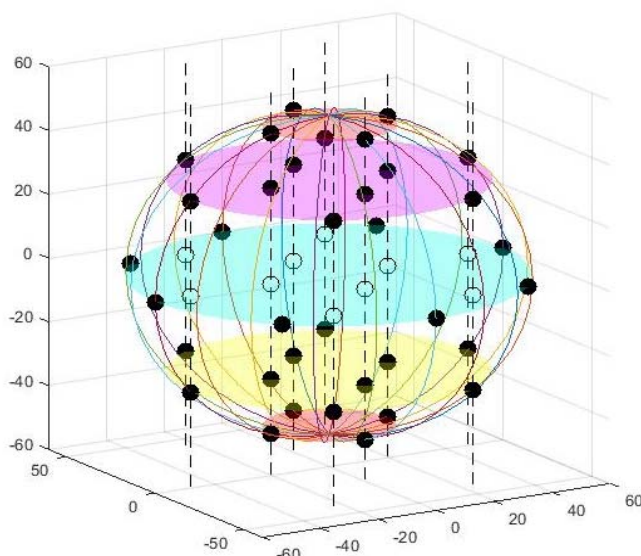


Рис.14. Точки, распределенные в цветовом пространстве CIELCH предложенным методом (восстанавливаемые точки не закрашены)

В частности, для точек со значениями из таблицы 1 получим:

$$L(z) = \sum_{i=0}^n \text{Light}_i * l_i(z) \quad (8)$$

$$L(z) = T_0 * \frac{z-30}{17,14} * \frac{z+30}{77,14} * \frac{z+47,14}{94,28} + T_1 * \frac{z-47,14}{-17,14} * \frac{z+30}{60} * \frac{z+47,14}{77,14}$$

$$+ T_2 * \frac{z-47,14}{-77,14} * \frac{z-30}{-60} * \frac{z+47,14}{17,14} + T_3 * \frac{z-47,14}{-94,28} * \frac{z-30}{-77,14} * \frac{z+30}{-17,14}$$

$$= 0,0000023 * z^3 - 0,00005 * z^2 + 0,0017 * z + 0,579$$

Тогда значения восприятия теплоты цвета для точки 3 ($z=0$), будет:

$$L(0) = 0,579$$

При использовании предложенного метода можно рассчитать значения восприятия теплоты цвета и в других точках пространства CIELCH, что позволяет увеличивать плотность узлов интерполяции.

Заключение

В настоящей работе был предложен метод распределения выборки цветовых оттенков в цветовом пространстве CIELCH, представляющем из себя эллипсоид вращения. Метод отвечает сразу нескольким требованиям, определенным контекстом данной научной работы. Соблюдается требование к ограниченному количеству точек в пространстве CIELCH, однако при этом сохраняется возможность увеличить количество выборки цветовых оттенков путем интерполяции промежуточных результатов. Таким образом, метод пригоден для дальнейшей работы с выборкой цветовых оттенков в цветовом пространстве CIELCH.

Литература

1. Dutton B., Vesga W., Waller J., James S., Seifi M. Additive Manufacturing Defect Formation and NDE Detectability // ASTM Symposium on Structural Integrity of Additive Manufactured Parts, Washington, DC, Nov. 6-8, 2018, ASTM STP 1620, M. Seifi and N. Shamsaei, Eds., ASTM International, West Conshohocken, PA, 2018.
2. Romano S., Brandão A., Gumpinger J., Gschweidl M., Beretta S. Qualification of AM parts: Extreme value statistics applied to tomographic measurements // Mater. Des. 131, pp. 32-48, 2017.
3. Balakrishnan, Ganesh & Yaw, Chong Tak & Koh, Siaw & Abedin, Tarek & Raj, Avinash & Tiong, Sieh Kiong & Chen. Chai. A Review of Infrared Thermography for Condition-Based Monitoring in Electrical Energy: Applications and Recommendations. Energies. 2022. 15. 6000. 10.3390/en15166000.
4. Vinay C., Reddy K., Ramana K.V. A review of methods on Condition Monitoring and Fault Diagnosis using IR Thermography-An Expert System approach // IOP Conf. Ser. Mater. Sci. Eng. 2020, 981, 042027.
5. Le Meur, Olivier & Thoreau, Dominique & Le Callet, Patrick & Barba, Dominique. A spatio-temporal model of the selective human visual attention // Proceedings of the International Conference on Image Processing (ICIP). 2005, pp. 1188-1191. 10.1109/ICIP.2005.1530610.
6. Можаяева А.И., Власюк И.В., Поташиников А.М. Lee Streeter. Эталонная объективная метрика оценки качества видео совместимая с PSNR учитывающая частотные и периферическую характеристики зрения человека // DSPA: вопросы применения цифровой обработки сигналов. 2021. №2.
7. Логинов И. Д. Обработка и сегментация тепловизионных изображений // Молодой ученый. 2017. № 13 (147). С. 62-72.
8. Филиппова Е.В. Общая структура тепловизоров и варианты их классификаций // Известия Тульского государственного университета. Технические науки. 2016. №2. С. 348-351.
9. Cruz-Albarran, Irving A. & Benitez-Rangel, Juan & Osornio-Rios, Roque & Dominguez, Benjamin & Rodriguez Medina, David & Morales-Hernández, Luis. A new approach to obtain a colour palette in thermographic images // Quantitative InfraRed Thermography Journal. 2018. 16. 1-20. 10.1080/17686733.2018.1509199.
10. Olalia Jr, Romulo & Olalia, Jenniea & Carse, Maynard. Evaluating Infrared Thermal Image's Color Palettes in Hot Tropical Area // Journal of Computer and Communications. 2021. 09. 37-49. 10.4236/jcc.2021.911003.
11. МСЭ-R ВТ.500-14. Методики субъективной оценки качества телевизионных изображений. Женева, 2020. 104 с.
12. Ho, Hsin-Ni. Color-temperature correspondence and its impact on object temperature perception. 2015.
13. Bailey, Authors & Grimm, Reynold & Davoli, Cindy & Bailey, Reynold & Grimm, Cindy & Davoli, Christopher. The real effect of warm-cool colors. 2006.
14. Жбанова В.Л. Вопросы применения цифровой колориметрии в современных научных исследованиях // Светотехника. 2021. №2. С. 5-14.
15. Schults, E. & Monogarova, O. & Oskolok, K. Digital Colorimetry: Analytical Possibilities and Prospects of Use // Moscow University Chemistry Bulletin. 2019. №74. С. 55-62. 10.3103/S002713141902007X.

COLOR SAMPLING DISTRIBUTION METHOD IN CIELCH COLOR SPACE

Vasily V. Ivanchev,
Postgraduate MTUCI, Moscow, Russia,
demidra18@gmail.com

Igor V. Vlasuyk,
Associate Professor of the Department of T&SB MTUCI,
PhD., Moscow, Russia,
ru3dlp@yandex.ru

Abstract

The scientific article presents the development of a method for distributing a sample of color shades in the CIELCH color space. The method takes into account specific requirements due to further work with the sample: compliance with the restriction on the number of color shades in the sample (in this case, represented by points in the CIELCH space) with the maximum possible density and uniform distribution of points. To increase the number of points, it is proposed to use two-dimensional interpolation of intermediate results for points located on the same axis.

Keywords: colorimetry, color spaces, CIELCH, interpolations, warm-coldness.

ВЕРОЯТНОСТНЫЕ ХРАКТЕРИСТИКИ УСКОРЕННОГО ПОИСКА ШУМОПОДОБНЫХ СИГНАЛОВ

Ву Ши Дао,
аспирант МТУСИ, Москва, Россия,
vusydaomtusi@gmail.com

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

Аннотация

Рассматриваются причины ограничения длительности шумоподобного сигнала, обрабатываемого в цифровом устройстве его ускоренного поиска, связанные как с нестабильностью частот тактовых генераторов на приемной и передающей стороне при отсутствии предварительной тактовой синхронизации, так и с условной сложностью и быстродействием этого устройства. Разработана методика анализа вероятностных характеристик ускоренного поиска (обнаружения) сигнала, позволяющая связать допустимую длительность времени его обработки и отношение сигнал/шум на входе приемника.

Ключевые слова: шумоподобный сигнал, параметры синхронизации, области неопределенности по частоте и задержке, поиск, ускоренный поиск, вероятностные характеристики обнаружения.

Введение

Бинарные шумоподобные сложные сигналы (СлС), формируемые на основе двоичных псевдослучайных последовательностей (ПСП), широко используются в настоящее время в различных цифровых системах связи, включая спутниковые радионавигационные системы [1,2]. Задачи синхронизации в них решаются, как правило, с помощью специальных периодически повторяющихся пилот-сигналов, излучаемых непосредственно на несущей частоте, либо на одной из ее квадратурных составляющих.

Основной проблемой является точная синхронизация СлС по времени, реже по частоте. Но необходимым условием успешной работы радиосистемы является синхронизация по периоду повторения пилот-сигнала и тактовой частоте псевдослучайного кода с целью формирования опорного на приемной стороне СлС, синхронного по времени с принимаемым. Эта задача может решаться с помощью многоэтапной процедуры поиска сигнала, первый этап которой производится в специальном устройстве ускоренного поиска, где осуществляется грубая синхронизация принимаемого СлС, прежде всего, по времени [3,4].

Устройство поиска сигнала является устройством его совместного обнаружения и оценки параметра задержки по времени, поэтому основным показателем качества его работы является вероятность правильного обнаружения сигнала с некоторой задержкой по времени относительно сигнала с условно нулевой задержкой при заданной вероятности ложной тревоги [7,8].

Целью исследований является разработка методики анализа вероятностных характеристик обнаружения шумоподобного сигнала в устройстве его ускоренного поиска с учетом ограничений на длительность его обработки, связанных с отсутствием синхронизации по тактовой частоте.

Поиск параметров синхронизации

Поиск предполагает обнаружение сигнала и измерение его параметров, как правило, частоты и времени задержки, с точностью, соответствующей размерам сечения основного пика функции неопределенности (ФН) сигнала [5-8]. То есть два соседних значения любого из параметров можно считать неразличимыми, если разность между ними меньше ширины сечения центрального пика ФН.

В результате, можно считать, что любой из рассматриваемых параметров синхронизации является дискретным, причем задержка сигнала по времени τ меняется от 0 до длительности (периода повторения) сигнала T , а Δf – от 0 до F , где F – ширина области неопределенности по частоте.

Если считать, что $\Delta\tau$ и Δf_n – соответственно ширина интервала области неопределенности по времени задержки и частоте, внутри которой значения любого из параметров неразличимы с точки зрения его оценки, то число дискретных значений параметра τ определяется как $n_\tau = T/\Delta\tau$, а число значений параметра Δf – как $n_f = F/\Delta f_n$. Но в действительности при дискретизации обрабатываемого сигнала значения $\Delta\tau$ и Δf_n должны

выбираться в соответствии с теоремой Котельникова, то есть в данном случае, при выборе интервалов дискретизации сигнала по времени и частоте в соответствии с размерами основного пика ФН, они будут в два раза больше интервала дискретизации сигнала, рекомендуемого этой теоремой. Тогда, если ширина спектра сигнала равна ΔF_s , то $\Delta \tau \approx 1/\Delta F_s$, а $\Delta f_n \approx 1/T$.

Учитывая, что в случае СлС $\Delta F_s \approx N/T$, получим, что общее число неизвестных параметров $n_{в,ч}(\text{СлС}) = n_b n_q = FT \Delta F_s T = FTN$, где N – длина (период) псевдослучайного кода [9].

На рисунке 1 приводится частотно-временная плоскость, на которой области неопределенности параметров синхронизации СлС - времени задержки и частоты, ограничены прямоугольником.

Область неопределенности параметров разделена сеткой на ячейки в виде прямоугольников со сторонами $\Delta \tau$ и Δf_n . Общее число ячеек равно $n_{в,ч}(\text{СлС})$. Площадь каждой из них приблизительно равна площади центрального пика ФН, то есть в каждой ячейке может расположиться только один его центральный пик. Поэтому сетка определяет границы между распознаваемыми значениями параметров, а сами параметры соответствуют центрам интервалов неопределенности и могут принимать любые значения из их общего числа n_b и n_q .

Таким образом, дискретные значения параметров синхронизации можно пронумеровать и обозначить их как $\tau_n, n = 1, \dots, n_b$ и $\Delta f_k, k = 1, \dots, n_q$.

На рисунке 1 выделена заштрихованная ячейка, соответствующая принимаемому СлС. Именно ее надо найти в процессе поиска, после чего надо отыскать ее центр. Но последнее уже не относится к задаче поиска.

Построение устройства оценки одновременно τ и Δf может основываться на использовании набора корреляторов (Кор), опорными сигналами которых служат копии сигнала с дискретными значениями параметров синхронизации, то есть при учете рекомендации теоремы Котельникова $\text{Re}[\dot{S}(t - \tau_n) \exp(j(2\pi(f_0 + \Delta f_k)t + \phi))]$, $n = 1, \dots, 2n_b, k = 1, \dots, 2n_q, \dot{S}(t)$ – комплексная огибающая принимаемого сигнала. Общее число корреляторов должно составлять $4n_b n_q$.

Структурная схема устройства, реализующего максимально правдоподобную оценку частоты и времени задержки сигнала, состоящая из набора корреляторов, приведена на рис. 2. Выходным сигналом корреляторов после вычисления модулей их откликов будет модуль периодической автокорреляционной функции СлС (ПАКФ) $|\dot{\chi}(\tau, \Delta f)|$ (с точностью до множителя, соответствующего энергии сигнала и аддитивной шумовой составляющей), значения которой для дискретных τ_n и Δf_k появятся одновременно на выходе корреляторов. Затем используется решающее устройство (РУ), представляющее собой селектор максимума. В последнем сравниваются значения откликов коррелятора и выбирается максимальный из них. В качестве максимально правдоподобной оценки частоты и времени задержки сигнала выдаются параметры опорного сигнала коррелятора с наибольшим выходным откликом.

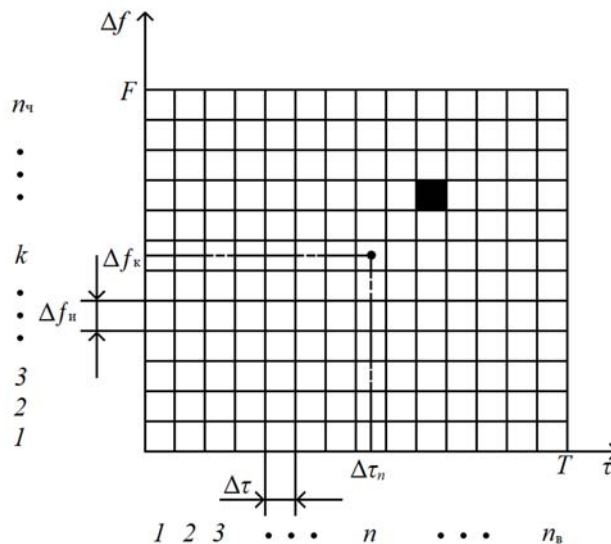


Рис. 1. Область неопределенности параметров синхронизации

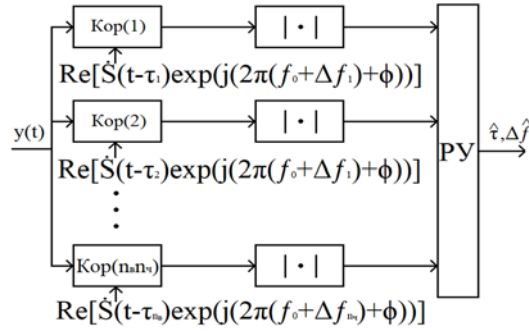


Рис 2. Устройство максимально правдоподобной оценки частоты и времени задержки на основе набора корреляторов

В случае если частота сигнала известна и равна $f_0(\Delta f = 0)$, в схеме на рисунке 2 используется только $2n_k$ корреляторов для оценки задержки сигнала по времени, но все они могут быть заменены одним согласованным фильтром (СФ).

Последний представляет собой линейное устройство с импульсным откликом в виде зеркальной копии полезного сигнала, то есть $h(t) = s(T - t)$, где T – длительность сигнала. СФ максимизирует выходное отношение сигнал/шум при воздействии аддитивного белого гауссовского шума, но в данном контексте важным является его свойство воспроизводить автокорреляционную функцию сигнала в реальном масштабе времени. Действительно, отклик СФ:

$$r(t) = \int_{-\infty}^{\infty} y(x)h(t-x)dx = \int_{-\infty}^{\infty} y(x)s(T-t+x)dx = \frac{1}{2} \operatorname{Re} \left[\int_0^T \dot{y}(x)\dot{s}(T-t+x)dx \right] = \operatorname{Re} \left[\left\{ \frac{1}{2} \int_0^T \dot{Y}(x)\dot{S}(T-t+x)dx \cdot \exp(-j2\pi f_0 T) \right\} \cdot \exp(j2\pi f_0 t) \right]. \quad (1)$$

Как видно, фигурные скобки в последней формуле выделяют комплексную огибающую сигнала на выходе СФ, причем:

$$\left| \frac{1}{2} \int_0^T \dot{Y}(x)\dot{S}(T-t+x)dx \right| = \frac{1}{2} \left| \dot{\chi}(t-T, 0) + \xi_n \right|, \quad (2)$$

где ξ_n – аддитивная помеховая составляющая на выходе СФ.

Но необходимо учитывать, что при записи (1) предполагалось, что СФ согласован с входным сигналом с точностью до смещения начальной фазы его высокочастотной несущей.

Таким образом, отклик СФ в реальном масштабе времени с коэффициентом $\frac{1}{2}$ повторяет действительную часть автокорреляционной функции входного сигнала, сдвинутую по времени на длительность сигнала T .

Данное обстоятельство позволяет использовать для максимально правдоподобной оценки задержки сигнала по времени устройство, схема которого приведена на рисунке 3.

Входной сигнал $y(t)$ вначале обрабатывается в СФ, с выхода которого сигнал поступает на детектор огибающей (ДО). В последнем блоке, представляющем собой решающее устройство (РУ), фиксируется момент времени t_m , когда сигнал на выходе ДО достигнет максимального значения, причем оценка задержки сигнала по времени равна $\hat{t} = t_m - T$. При этом значения $|\operatorname{Re}\{\dot{\chi}(t-T, 0)\}|$ на входе решающего устройства (РУ) появляются последовательно.



Рис. 3. Устройство максимально правдоподобной оценки времени задержки на основе СФ

СФ на практике обычно является цифровым устройством, на входе которого каким-либо способом выделяется комплексная огибающая принимаемого сигнала (для СлС, формируемого на основе двоичной ПСП – это действительная функция), после чего осуществляется ее дискретизация по времени с тактовой частотой $f_T \approx 1 / 2\Delta t$

При этом на длительности элементарного импульса СлС $T_3 = T / N$ берут два его отсчета, сдвинутых друг относительно друга по времени на $T_3 / 2$. Эти две группы отсчетов сигнала, каждая из которых берется через интервалы времени длительностью T_3 должны обрабатываться отдельно – каждая в своем устройстве, а решение об обнаружении сигнала должно приниматься при его обнаружении в любом из этих устройств. Условно будем их называть СФ1 и СФ2.

Необходимо учитывать, что удвоенная частота дискретизации сигнала не совсем точно соответствует тактовой частоте принимаемого сигнала, теоретически равной $2f_T$, из-за нестабильностей задающих генераторов как на передающей, так и на приемной стороне. Кроме того, на этапе поиска тактовая синхронизация принимаемого сигнала еще не произведена. В результате этого спустя некоторый промежуток времени обязательно произойдет проскальзывание, то есть два отсчета СлС при его дискретизации через интервал времени T_3 будут приходиться на один и тот же его элементарный импульс, либо один такой импульс будет пропущен.

Благодаря сдвигу отсчетов СлС на входах СФ1 и СФ2 на $T_3 / 2$ проскальзывание никогда не будет происходить на входах этих устройств одновременно. Тем не менее, длительность времени обработки СлС в этих устройствах не должно превышать длительности времени между двумя соседними проскальзываниями, которое легко оценить, зная нестабильность частот задающих тактовых генераторов на передающей и приемной стороне.

Так, если $\frac{\Delta f_T}{f_T} = 10^{-4}$, то легко посчитать, что длительность псевдослучайного кода, который может обрабатываться в СФ1 и СФ2 не должна превышать приблизительно 5000. Поэтому важно уметь оценивать эффективность поиска СлС в зависимости от длины ПСП, обрабатываемой в соответствующем устройстве.

Эффективность ускоренного поиска СлС

Задача поиска параметров синхронизации СлС, то есть совместного его обнаружения и оценки параметров с точностью, соответствующей размерам сечения основного пика ФН, сводится к задаче распознавания $n_{в,ч} = n_b n_c$ квазиортогональных сигналов.

Действительно, каждой паре параметров Δf_k и τ_n соответствует сигнал $s(t, \tau_n, \Delta f_k)$, $n=1, \dots, n_b$, $k=1, \dots, n_c$, а ФН всех этих сигналов имеют несовпадающие центральные пики. Их взаимные ФН принимают малые значения, соответствующие боковым пикам ФН. Если можно допустить, что форма ФН сигнала имеет приблизительно кнопочную форму и статистические характеристики ее боковых пиков известны, то задача поиска параметров синхронизации сводится к задаче распознавания $n_{в,ч}$ квазиортогональных сигналов [9,10]. В данном случае $n_{в,ч} = n_b = N$.

Кроме того, рассматривался вариант с условно когерентным сигналом накоплением СлС на длительности нескольких периодов его повторения с выхода каждого СФ, когда СлС может обрабатываться в СФ на длительности периода повторения сигнала, а на входе РУ отсчеты ПАКФ параллельно суммируются. Также исследовались характеристики обнаружения сигнала в случае, когда в СФ возможна обработка лишь части периода СлС.

Особенностью разработанной методики исследования эффективности устройства поиска является возможность построения графиков с учетом отношения мощности шума к мощности сигнала на входе приемника $\left[\frac{P_{ш}}{P_c}\right]_{вх}$, а также возможность учета длины обрабатываемого псевдослучайного кода.

На рисунке 4 приводятся графики зависимостей вероятностей правильного обнаружения СлС $P_{обн}$ от $\left[\frac{P_{ш}}{P_c}\right]_{вх}$ при вероятности ложной тревоги $P_{лт} = 10^{-4}$ и 10^{-6} .

Предполагалось, что сигнал сформирован на основе М-последовательности (МП) длиной N , и в СФ обрабатывается один ($k=1$) полный период ее повторения. Аналогичные графики, но для случая, когда в СФ возможна обработка либо части периода СлС ($k=0.1, 0.3, \dots$), где k – его часть, либо одного периода при условно когерентном накоплении нескольких периодов ($k=2, \dots, 10, \dots, 50$) при формировании СлС на основе МП приводятся на рисунках 5-10, а при формировании на основе кода Голда (рис. 11).

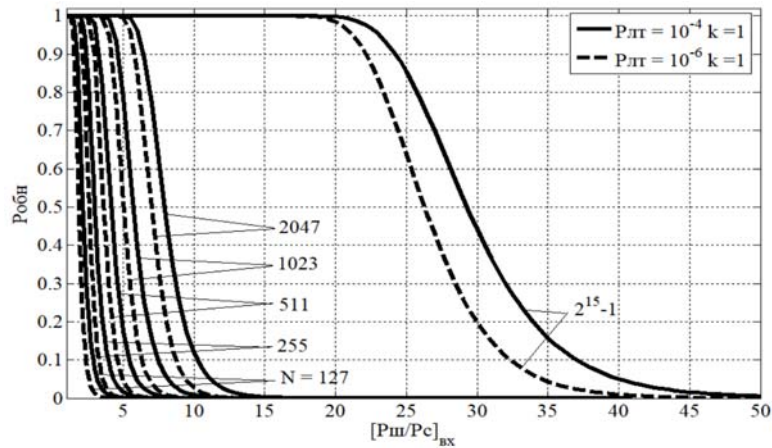


Рис. 4. Зависимости $P_{обн}$ от $\left[\frac{P_{ш}}{P_c}\right]_{вх}$ при формировании СлС на основе МП и заданной $P_{лг} = 10^{-4}$ или 10^{-6} для $N = 127, 255, 511, 1023, 2047, 2^{15} - 1$

Разработанная методика позволяет производить анализ вероятности ошибки от $\left[\frac{P_{ш}}{P_c}\right]_{вх}$ при наличии на входе приемника нескольких сдвинутых по времени копий одного и того же сигнала вследствие его многолучевого распространения.

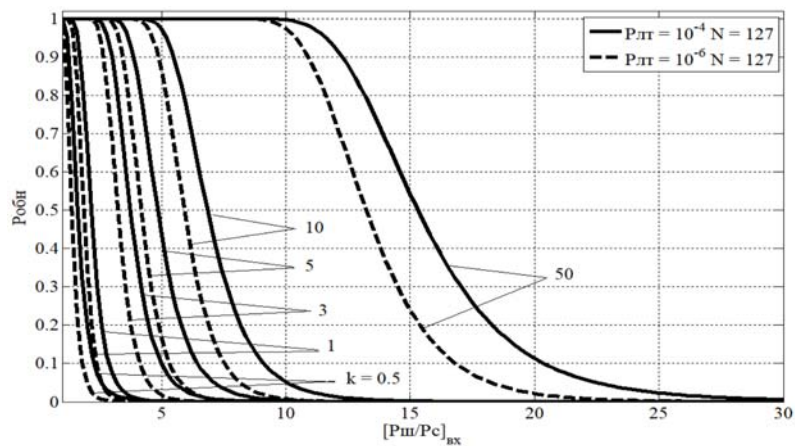


Рис. 5. Зависимости $P_{обн}$ от $\left[\frac{P_{ш}}{P_c}\right]_{вх}$ при формировании СлС на основе МП для $N = 127$ при $k = 0.5, \dots, 50$.

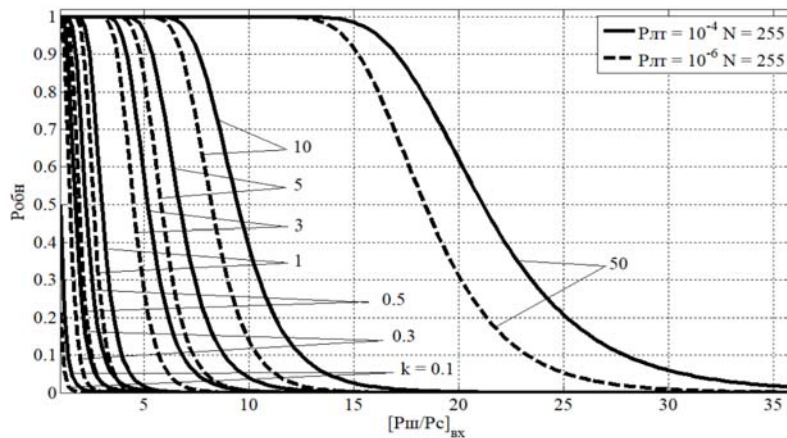


Рис. 6. Зависимости $P_{обн}$ от $\left[\frac{P_{ш}}{P_c}\right]_{вх}$ при формировании СлС на основе МП для $N = 255$

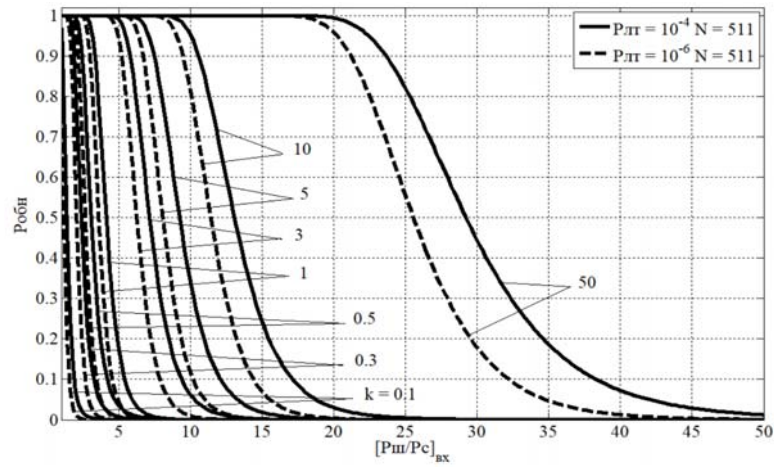


Рис. 7. Зависимости $P_{обн}$ от $\left[\frac{P_{ш}}{P_c}\right]_{вх}$ при формировании СлС на основе МП для $N = 511$

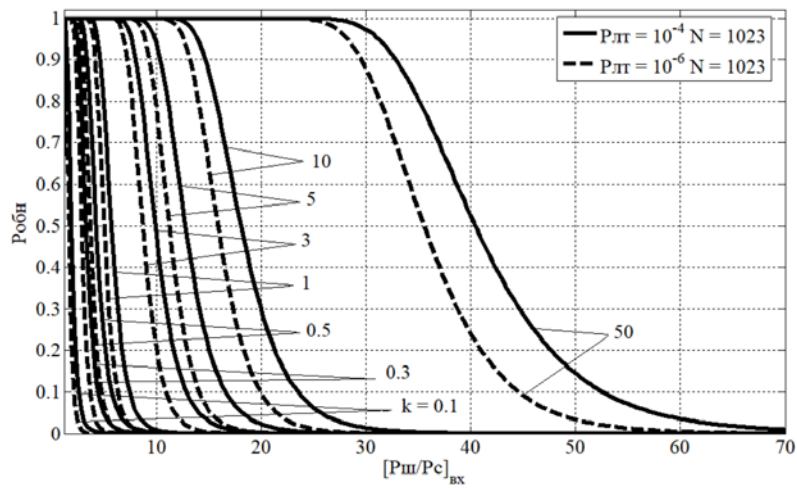


Рис. 8. Зависимости $P_{обн}$ от $\left[\frac{P_{ш}}{P_c}\right]_{вх}$ при формировании СлС на основе МП для $N = 1023$

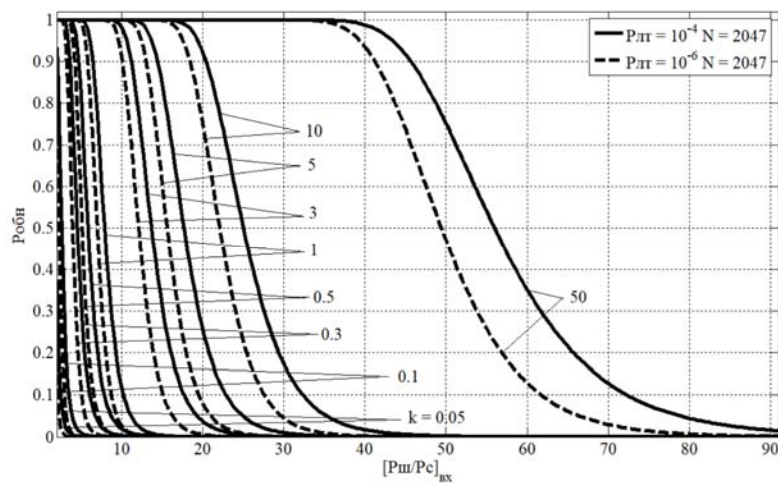


Рис. 9. Зависимости $P_{обн}$ от $\left[\frac{P_{ш}}{P_c}\right]_{вх}$ при формировании СлС на основе МП для $N = 2047$

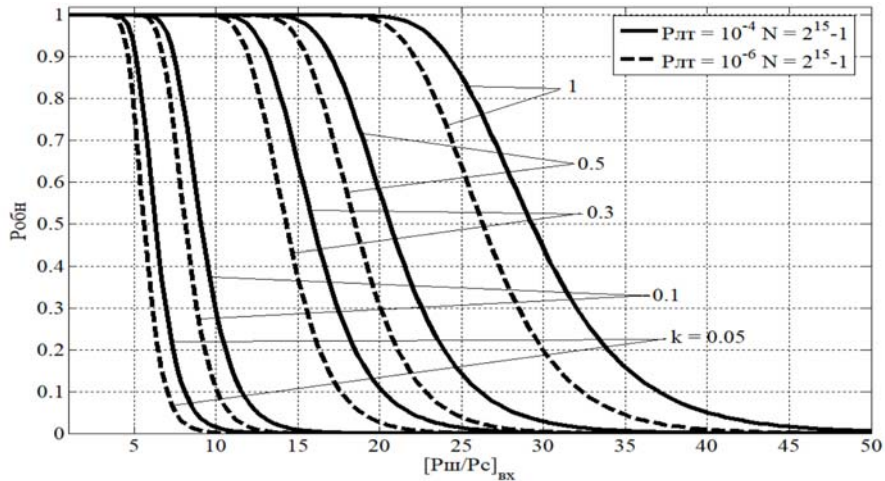


Рис. 10. Зависимости $P_{обн}$ от $\left[\frac{P_{ш}}{P_c}\right]_{вх}$ при формировании СЛС на основе МП для $N = 2^{15} - 1$

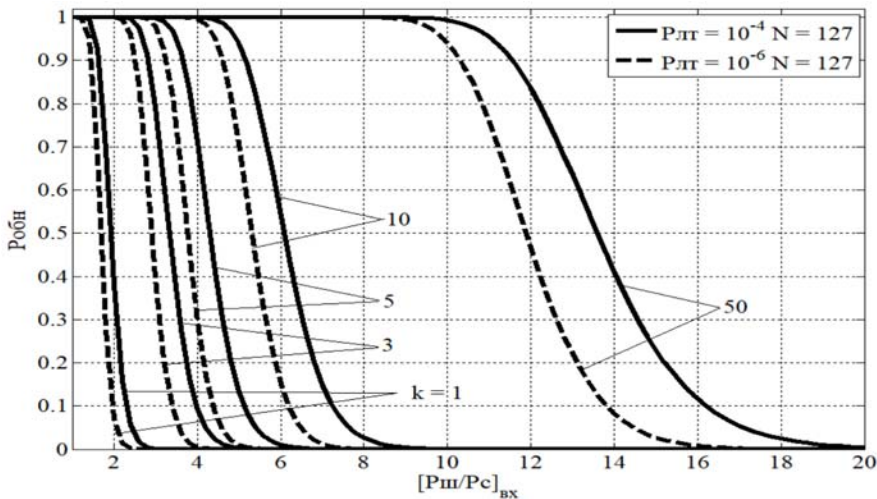


Рис. 11. Зависимости $P_{обн}$ от $\left[\frac{P_{ш}}{P_c}\right]_{вх}$ при формировании СЛС на основе кода Голда для $N = 127$

Заключение

В статье приведена разработанная методика анализа эффективности цифрового устройства ускоренного поиска параметра задержки по времени шумоподобного сложного сигнала с учетом ограничения длительности времени его обработки, связанного с нестабильностью задающих тактовых генераторов на передающей и приемной стороне.

Другим обстоятельством, обстоятельством обуславливающим необходимость обрабатывать относительно коротких отрезков сигнала, является условная относительная сложность устройства поиска и его быстродействие. В следствие этого, базовым параметром при анализе эффективности устройства поиска шумоподобного сигнала является длительность псевдослучайного кода, который может непосредственно обрабатываться в согласованном фильтре или корреляторе, а также допустимое число периодов сигнала, энергия которых может накапливаться условно когерентно с выхода фильтра или коррелятора.

Таким образом, разработанная методика анализа эффективности устройства поиска позволяет судить об его эффективности при наличии априорной информации о нестабильности тактовых задающих генераторов и отношении мощности шума к мощности сигнала на входе приемника.

Литература

1. *Инамов В.П.* Широкополосные системы и кодовое разделение сигналов. М.: Мир связи, 2007.
2. *Beard C., Stallings W.* Wireless Communication Networks and Systems. L.: Pearson, 2016.
3. *Ву Шу.Д., Горгадзе С.Ф.* Устройство ускоренного поиска шумоподобного сигнала // Технологии информационного общества // Сборник трудов XVI Международной отраслевой научно-технической конференции. Москва, 2022. С. 88-90.
4. *Gorgadze S.F., Gut T.M.* Accelerated evaluation of spread spectrum signals synchronization parameters // 2020 Systems of Signals Generating and Processing in the Field of on Board Communications. 2020. С. 9078627.
5. *Гут Т.М., Горгадзе С.Ф.* Характеристики ковариационных функций и оценка параметров шумоподобного сигнала // Телекоммуникации и информационные технологии. 2019. Т. 6. № 2. С. 35-41.
6. *Горгадзе С.Ф.* Ускоренный цифровой алгоритм синхронизация шумоподобных сигналов по времени и частоте // Системы синхронизации, формирования и обработки сигналов. 2016. Т. 7. № 4. С. 16-18.
7. *Горгадзе С.Ф., Бойков В.В.* Измерительные сигналы с многопозиционными поднесущими для спутниковых радионавигационных систем // Радиотехника и электроника. 2014. Т. 59. № 3. С. 264.
8. *Горгадзе С.Ф., Вовк А.С.* Оценка параметров шумоподобного сигнала на негармонической поднесущей // Фундаментальные проблемы радиоэлектронного приборостроения. 2014. Т. 14. № 5. С. 182-185.
9. *Варакин Л.Е.* Системы связи с шумоподобными сигналами. М.: Радио и связи, 1985. 384 с.
10. *Тихонов В.И.* Статистическая радиотехника. М.: Советское радио, 1966. 219 с.

**PROBABILISTIC CHARACTERISTICS OF ACCELERATED SEARCH
SPREAD SPECTRUM SIGNALS**

Vu Sy Dao,

Postgraduate MTUCI, Moscow, Russia,
vsydaomtusi@gmail.com

Svetlana F. Gorgadze,

Professor of the Department of ROS MTUCI,
Doctor of Technical Sciences, Professor, Moscow, Russia,
s.f.gorgadze@mtuci.ru

Abstract

The reasons for limiting the duration of a spread spectrum signal processed in a digital device for its accelerated search are considered, which are associated both with the instability of the frequencies of clock generators on the receiving and transmitting sides in the absence of preliminary clock synchronization, and with the conditional complexity and response speed of this device. A technique has been developed for analyzing the probabilistic characteristics of an accelerated search (detection) of a signal, which makes it possible to relate the allowable duration of its processing time and the signal-to-noise ratio at the receiver input.

Keywords: *spread spectrum signals, synchronization parameters, frequency and delay uncertainty regions, search, accelerated search, probabilistic characteristics of detection.*

ИССЛЕДОВАНИЕ ДОПОЛНИТЕЛЬНЫХ КРИТЕРИЕВ КАЧЕСТВА ЗВУКА В ПОМЕЩЕНИЯХ С КУПОЛЬНОЙ КОНСТРУКЦИЕЙ ПОТОЛКА

Калинина Лилия Владимировна,
магистрант МТУСИ, Москва, Россия,
likoska@mail.ru

Чернышева Татьяна Васильевна,
доцент кафедры ТиЗВ имени С. И. Катаева МТУСИ,
к.т.н., доцент, Москва, Россия,
krba2012@yandex.ru

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

Аннотация

Статья посвящена анализу распределения дополнительных критериев качества помещений с купольной конструкцией потолка разной высоты. Исследования проводились как для купольных зданий, так и для зданий равновеликого объема с плоским потолком. Приводятся результаты расчетов дополнительных параметров качества звука для двух типов помещений и их сравнение для оценки влияния купола на распределение акустических параметров в подкупольном пространстве.

Ключевые слова: критерии качества помещений, прозрачность звучания, разборчивость, четкость, индекс передачи речи, артикуляционная потеря согласных, время реверберации, уровни звукового поля.

Введение

Всё большее количество зданий с купольной конструкцией потолка используется для проведения массовых мероприятий: концертов, организации экскурсий, различного рода конференций, проходящих в помещениях, которые предназначены для воспроизведения как речевых, так и музыкальных программ. В связи с тем, что исследований подобных помещений проводилось недостаточно, изучение влияния купола этих помещений на внутреннюю звуковую картину помещения весьма актуально в наши дни [1].

Целью статьи является поиск мер повышения качества звука в помещении. Для этого было проведено исследование дополнительных критериев оценки качества помещений, таких как прозрачность звука, разборчивость, четкость, индекс передачи речи и артикуляционная потеря согласных. [2,8] С учётом результатов ранее проведенных исследований, направление акустических осей источников звука было выбрано к центру зала [7]. Так как архитектурные особенности помещения – купольная конструкция потолка – вносят изменения в восприятие слушателями акустической картины. Рассмотрены варианты одного и того же помещения с разной высотой купола.

3D-модели помещений

Для исследования акустических характеристик использовался программный продукт, в котором построена 3D-модель помещения со следующими параметрами: высота стен – 55 м; радиус раскрытия пола – 25 м; высоты купола с двумя вариантами: 15 м и 25 м (рис. 1).

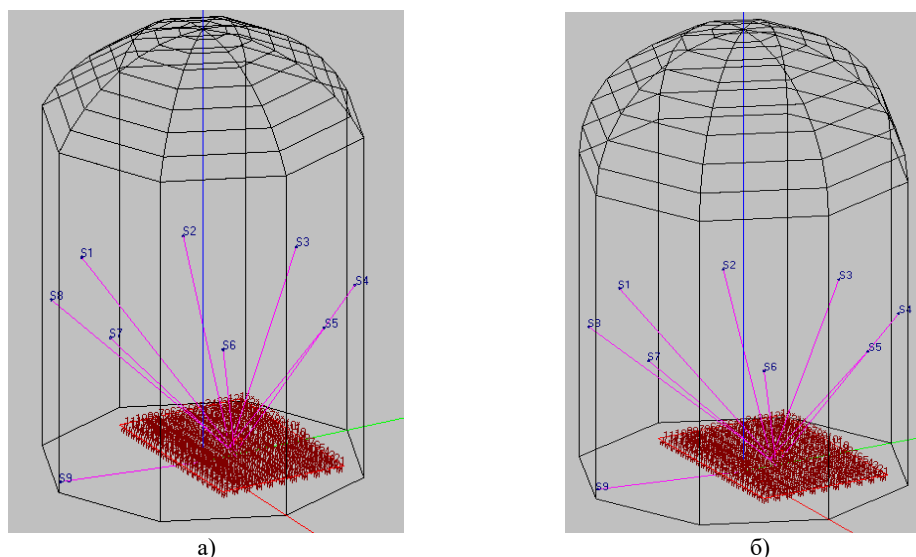


Рис. 1. 3D-модель помещения с высотой купола равной: а) 15 м.; б) 25 м

Для расчетов в выбранном помещении предполагается нахождение 121 слушателя.

Для озвучения помещения были выбраны и установлены 8 симметрично расположенных источников звука, установленный в одном из углов помещения девятый источник звука, является аналогом исполнителя. Размещение восьми источников звука было выбрано на высоте 27,5 м исходя из предварительно заданной неравномерности звучания не более 3 дБ. Для помещения с другими размерами высота подвеса источников звука рассчитывается исходя из существующих рекомендаций. [2] Тип источников звука был выбран с учётом диаграммы направленности звуковой колонки. Расчеты проводились на частоте 1000 Гц. Для оценки влияния купола на дополнительные критерии акустических качеств помещений, были построены две модели помещений с плоской конструкцией потолка, эквивалентных по объему помещениям с купольной конструкцией [5].

Объем помещений с купольной конструкцией потолка составил 111974.9 м³ для помещения с куполом высотой 15 м и 126408.96 м³ для помещения с куполом высотой 25 м. Соответственно были построены модели зданий с плоской конструкцией потолка и равновеликими объемами помещений (рис. 2).

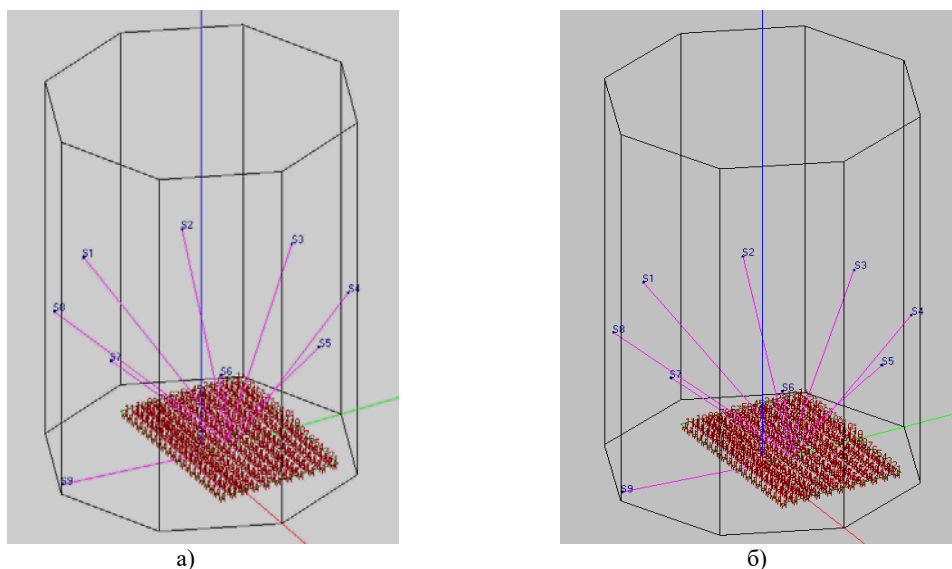


Рис. 2. 3D модели помещений с плоской конструкцией потолка объема, равного: а) 111990,13 м³; б) 126397,7 м³

В помещениях было достигнуто время реверберации, близкое к оптимальному [6].

На рисунках 3 и 4 соответственно приведены частотные характеристики времени реверберации в помещении с куполом высотой 15 м и 25 м и равновеликими помещениями с плоской крышей

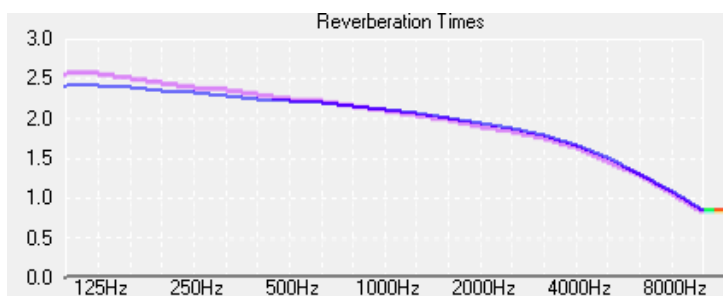


Рис. 3. Частотные характеристика времени реверберации помещений: с куполом высотой 15 м (розовая линия) и с плоским потолком (синяя линия)

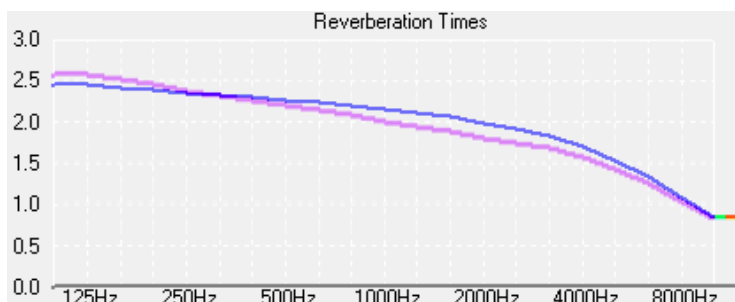


Рис. 4. Частотные характеристика времени реверберации помещений: с куполом высотой 25 м (розовая линия) и с плоским потолком (синяя линия)

На рисунке 5 представлены полученные частотные характеристики времени реверберации для двух купольных помещений с разными высотами куполов.

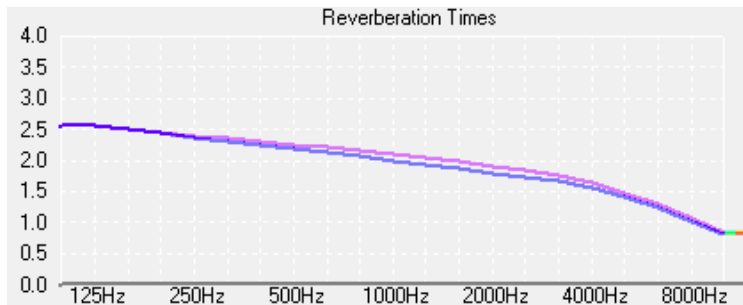


Рис. 5. Частотные характеристика времени реверберации в купольных помещениях при высоте купола: 15 м (розовая линия) и 25 м (синяя линия)

Из рисунков 3 и 4 следует, что купол увеличивает реверберацию на низких частотах по сравнению с помещениями без купола. С увеличением высоты купола увеличивается степень диффузности [8]. При куполе высотой 25 м (рис. 4) заметно снижение времени реверберации на средних и высоких частотах. При сравнении времени реверберации помещений с разной высотой купола видно, что при более высоком куполе время реверберации на средних и высоких частотах снижается (рис. 5).

Результаты расчетов распределения дополнительных коэффициентов качества помещений

Таблица 1

Помещение с купольной конструкцией потолка (высота купола – 15 м)
и помещение с плоской крышей того же объема

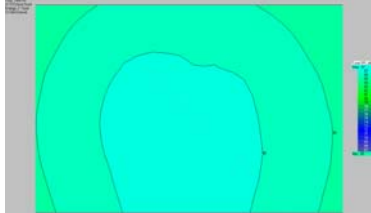
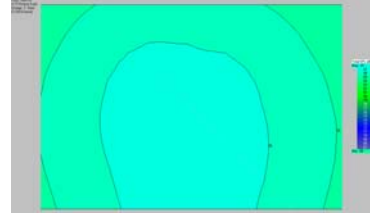
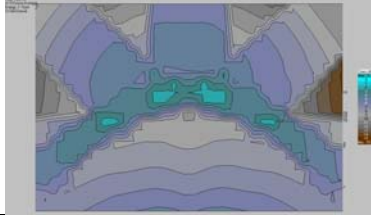
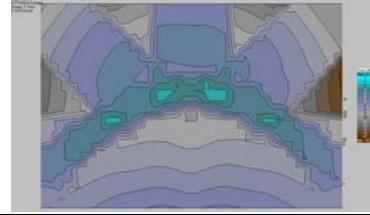
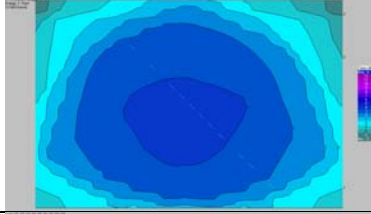
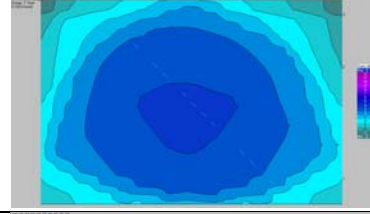




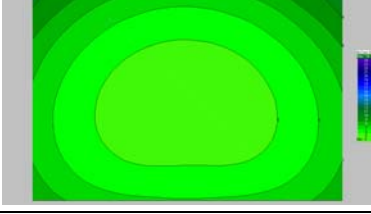
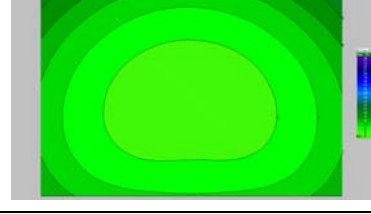
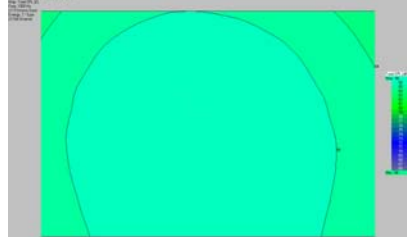
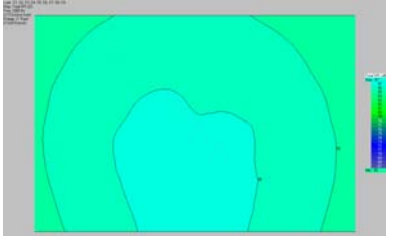
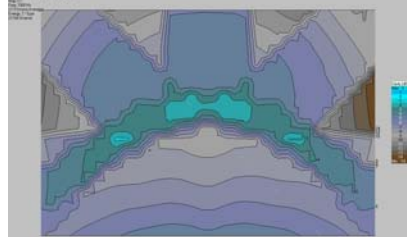
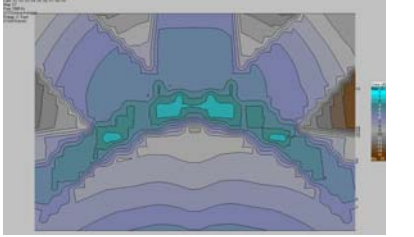
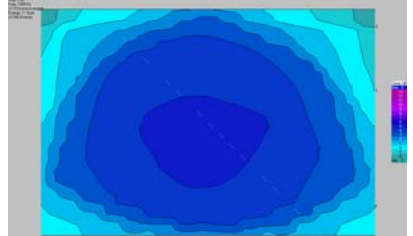
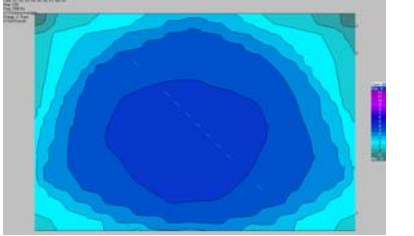
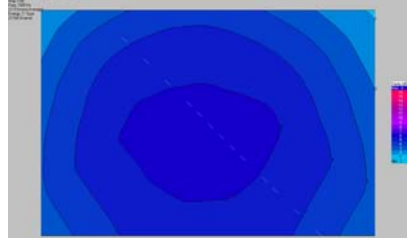
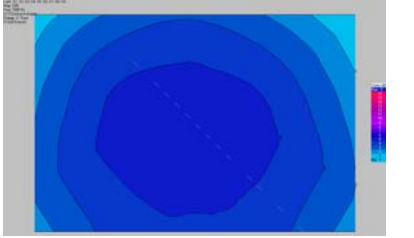
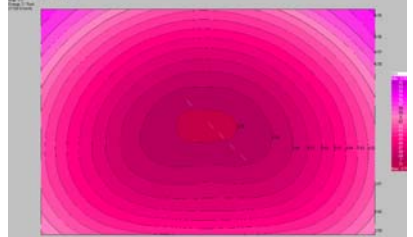
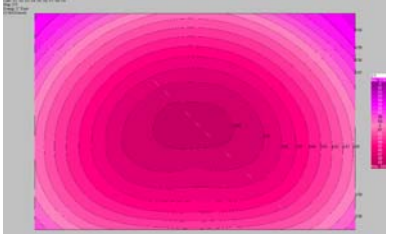
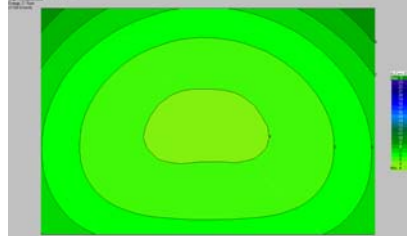
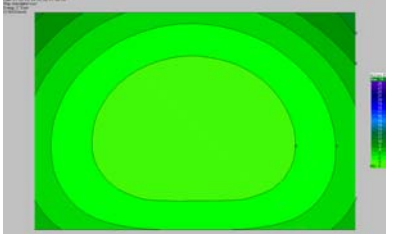
Наименование коэффициента	Помещение с куполом	Помещение с плоской крышей
Распределение уровней суммарного поля		
Распределение коэффициента точности локализации C7		
Распределение коэффициента разборчивости C50		
Распределение коэффициента четкости C80		
Распределение коэффициента индекса передачи речи STI		
Распределение коэффициента артикуляционной потери согласных Articulation Loss		

Таблица 2

Помещение с купольной конструкцией потолка (высота купола – 25 м)
и помещение с плоской крышей того же объема

Наименование коэффициента	Помещение с куполом	Помещение с плоской крышей
Распределение уровней суммарного поля		
Распределение коэффициента точности локализации C7		
Распределение коэффициента разборчивости C50		
Распределение коэффициента четкости C80		
Распределение коэффициента индекса передачи речи STI		
Распределение коэффициента артикуляционной потери согласных Articulation Loss		

Сравнение результатов расчётов дополнительных критериев качества звучания в помещении с купольной конструкцией потолка при высоте купола 15 м и помещении с плоской крышей того же объема, приведённых в таблице 1, позволяет сделать следующие выводы:

1. Хотя в обоих помещениях неравномерность поля составляет 2 дБ, в соответствии с рекомендациями [3] обеспечен требуемый уровень поля величиной 85-87 дБ, в помещении с плоской крышей область со значением поля, равным 86 дБ (близким к верхней границе уровня) несколько ниже, чем в помещении с куполом.

Это показывает, что в помещении с купольной конструкцией потолка уровень поля в большей степени соответствует рекомендациям.

2). Значения коэффициента точности локализации $C7$ колеблются от -15 до -2 дБ. Согласно рекомендациям, наилучшая локализация находится в области центра зала подковообразной формы. Так как локализация улучшается по мере приближения значения параметра к 0 дБ, был сделан вывод о том, что в помещении с купольным потолком локализация лучше, так как в центре области имеет подковообразную форму.

3). Значения коэффициента разборчивости речи $C50$ колеблются от -3 до 4 дБ. Согласно рекомендациям, область с наилучшей разборчивостью [4] находится в центре зала. В помещении с купольной конструкцией потолка эта область больше, чем у помещения с плоской крышей.

3). Значения коэффициента четкости $C80$ колеблются от 0 до 5 дБ. Полученные результаты дают величину четкости, не соответствующую рекомендуемой для речевых программ. Согласно другим рекомендациям, оба помещения подходят для их использования в качестве концертных залов. Лучшая четкость находится в области центра зала. Эта область больше по площади в помещении с купольной конструкцией потолка.

4). Значения индекса передачи речи STI колеблются от 0.53 до 0.68, что, согласно интерпретации показателей STI , соответствует оценкам «удовлетворительно» и «хорошо». Наилучшая область находится в центре зала. В помещении с купольной конструкцией потолка данная область шире, чем у помещения с плоской крышей.

5). Значения артикуляционной потери согласных колеблются от 5% до 10%, что, согласно рекомендациям, соответствует оценкам «хорошо» и «удовлетворительно» [4,8]. Наилучшая область по артикуляции речевой информации находится в центре зала. При детальном сравнении полученных результатов можно заметить, что в помещении с куполом наилучшая область немного больше, чем в помещении с плоским потолком.

6). В целом значения дополнительных критериев качества звучания в помещениях одинакового объема с купольной конструкцией (высота 15 м.) и плоским потолком практически идентичны. Однако в помещении с купольной конструкцией потолка результаты более близки к рекомендуемым.

Приведённые в таблице 2 полученные при моделировании значения дополнительных критериев качества звука в помещении с купольной конструкцией потолка (высота купола 25 м) и в помещениях того же объема с плоской крышей, позволяют сделать следующие выводы.

1). Уровень поля помещения с купольной конструкцией потолка равен 84-86 дБ, что соответствует существующим рекомендациям. Неравномерность составляет 2 дБ. Уровень поля помещения с плоским потолком равен 85-87 дБ, что несколько превышает нормы [3]. Неравномерность поля в обоих помещениях одинакова - 2 дБ.

2). Значения коэффициента точности локализации $C7$ колеблются в помещениях соответственно от -14 до -1 дБ в первом и от -15 до -2 дБ во втором. В обоих помещениях область с наилучшей локализацией находится в центре зала подковообразной формы, а самая лучшая локализация сосредоточена в центре этой области и симметрично по её краям. Значения коэффициента и величина области с наилучшей локализацией больше соответствует рекомендуемым в помещении с купольной конструкцией потолка.

3). Значения коэффициента разборчивости речи $C50$ колеблются соответственно от -2 до 5 дБ и от -3 до 4 дБ для двух сравниваемых помещений. Согласно рекомендациям, область с наилучшей разборчивостью находится в центре зала. В помещении с купольной конструкцией потолка значение коэффициента разборчивости [7] более приближено к рекомендованным значениям, чем в помещении с плоским потолком.

4). Значения коэффициента четкости $C80$ колеблются от 1 до 6 дБ и от 1 до 5 дБ соответственно. Полученные результаты коэффициентов четкости не соответствуют рекомендациям для речевых программ. В тоже время, оба помещения соответствуют рекомендациям по использованию их в качестве концертных залов. Лучшая четкость звучания обеспечивается в обоих помещениях в области центра зала.

5). Значения индекса передачи речи STI в помещении с купольной крышей колеблются от 0.55 до 0.71, что, согласно интерпретации показателей STI , соответствует оценкам «удовлетворительно» и «хорошо». Наилучшая область находится в центре зала. Значения индекса передачи речи STI в помещении с плоской крышей колеблются от 0.54 до 0.69, что, согласно интерпретации показателей STI , также соответствует оценкам «удовлетворительно» и «хорошо». Наилучшая область находится в центре зала. Сравнение значений коэффициентов STI в наилучших областях позволяет сделать вывод о том, что в купольном помещении индекс передачи речи выше.

6). Значения артикуляционной потери согласных колеблются у сравниваемых двух помещений соответственно от 4% до 9% и от 5% до 10%, что, согласно рекомендациям, соответствует оценкам «хорошо» и «удовлетворительно». Наилучшая область по артикуляции находится в центре зала. Согласно рекомендациям, значения данного параметра лучше в помещении с купольной конструкцией потолка [9].

Заключение

Результаты проведённого моделирования и полученные результаты распределения дополнительных коэффициентов оценки качества помещений равных объёмов с различной конструкцией потолка показали, что в помещениях с купольной конструкцией потолка результирующие коэффициенты более близки к рекомендованным. Следовательно, для улучшения качества восприятия звуковой картины слушателями предпочтительнее делать выбор в пользу купольных помещений.

При сравнении значений дополнительных коэффициентов оценки качества помещений с купольной конструкцией потолка, представленных в таблицах 1 и 2 для разных высот купола (15 и 25 м), в большей степени соответствуют рекомендациям помещения с большей высотой купола. Поэтому, при выборе помещения для проведения массовых мероприятий стоит делать выбор в сторону помещения с большим подкупольным пространством.

Литература

1. Калинина Л.В., Чернышева Т.В. Анализ распределения уровней дополнительных критериев качества звука в подкупольном помещении // Сборник трудов XIII молодёжного научного форума «Телекоммуникации и информационные технологии». М.: ОНИРС МТУСИ, 2022. 59 с.
2. Вахитов Ш.Я., Ковалгин Ю.А., Фадеев А.А., Щевьев Ю.П. Акустика. Учебник для вузов. Под ред. профессора Ю.А. Ковалгина. М.: Горячая линия – Телеком, 2016. 660 с.
3. Йордан В.Л. Акустическое проектирование концертных залов и театров / Перевод с английского: Хомутов С.А., редактор: Манриненко Л.И. М.: Стройиздат, 1986. 170 с.
4. Орлов В.Г., Терехов А.Н. Оценка качества речи в аудиоинформационных устройствах для телефонных сетей общего пользования // Фундаментальные проблемы радиоэлектронного приборостроения. 2008. Т. 8. № 1. С. 272-277.
5. Выходец А.В., Гитлиц М.В., Ковалгин Ю.А., Никонов А.В., Однолько В.В. Радиовещание и электроакустика: учебник для ВУЗов. М.: Радио и связь, 1989. 432 с.
6. Ефимов А.П. Три взгляда на акустику помещений // Install-pro. № 3-7, 2000.
7. Алдошина И. А., Приттс Рой. Музыкальная акустика. Учебник для высших учебных заведений. СПб.: Композитор Санкт-Петербург, 2006. 720 с.
8. ГОСТ 25902-2016 Зрительные залы. Метод определения разборчивости речи.
9. Емельянов Е.Д. Звукофикация театров и концертных залов. М.: Искусство, 1989. 342 с.

STUDY OF ADDITIONAL SOUND QUALITY CRITERIA IN ROOMS WITH DOME CEILING STRUCTURE

Liliya V. Kalinina,
Graduate MTUCI, Moscow, Russia
likoska@mail.ru

Tatyana V. Chernysheva,
Associate Professor of the Department of T&SB MTUCI,
PhD., Moscow, Russia,
krba2012@yandex.ru

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

Abstract

The article considers the analysis of the distribution of additional criteria for the quality of premises in buildings with a domed ceiling structure of different heights. The researchers have been conducted for domed buildings and for buildings of equal volume with a flat ceiling. The article presents the results of calculations of additional quality parameters and their comparison for two types of rooms to consider the influence of the dome on the distribution of acoustic parameters in the dome space.

Keywords: criteria for the quality of premises, clarity coefficients, speech transmission index, articulation loss, reverberation time, sound pressure level.

«Сетевые технологии и системы телекоммуникаций»

ПРИМЕНЕНИЕ НАУКОМЕТРИЧЕСКОГО АНАЛИЗА ДЛЯ ВЫЯВЛЕНИЯ ПЕРСПЕКТИВНЫХ РАЗРАБОТОК В ОБЛАСТИ ВОЛС

Рабенандрасана Жослен,
ассистент кафедры НТС МТУСИ, к.т.н., Москва, Россия,
r.jocelin@mtuci.ru

Зайцев Александр Игоревич,
аспирант кафедры НТС МТУСИ, Москва, Россия,
a.zaitcev96@yandex.ru

Аннотация

Рассмотрены технологии спектрального уплотнения каналов связи и пространственного мультиплексирования в волоконно-оптических линиях связи. Проанализирован поток научно-технической литературы в области исследований технологий спектрального уплотнения каналов и технологий пространственного мультиплексирования. На основе известных методик анализа публикационной активности проведена оценка технологической вовлеченности различных стран, выявлены стадии развития технологий спектрального и пространственного уплотнения каналов связи. Представлены данные о динамике публикаций в интервале времени 2017-2022 гг., позволяющие прогнозировать вовлеченность различных коллективов в исследования по данным тематикам. Определены группы стран, в которых работают наиболее компетентные исследовательские группы в области спектрального уплотнения и пространственного мультиплексирования. Выявлен список организаций, которые наиболее активно финансируют исследования по рассмотренным технологиям. Представлено распределение публикационной активности по следующим отраслям знаний: спектральное уплотнение каналов связи и пространственное мультиплексирование каналов связи. На основе сравнительного анализа динамики публикационной активности двух направлений, определены основные тенденции развития перспективных разработок технологии спектрального уплотнения и пространственного мультиплексирования.

Ключевые слова: волоконно-оптические линии связи, спектральное уплотнение, пространственное мультиплексирование, пропускная способность, оптическое волокно, наукометрический анализ.

Введение

Внедрение технологии 5G, увеличение востребованности интернета вещей (IoT – Internet of Things), растущий спрос на мультимедийные приложения [1-3], внедрение потокового видео сверхвысокой четкости [4] и применение технологии прокладки оптического волокна до дома (FTTH – Fiber To The Home), для всех этих технологий необходимо увеличение пропускной способности канала связи. В настоящее время основной поток цифровых данных осуществляется с использованием оптических волоконных линий связи. Одним из способов повышения их пропускной способности является применение технологии спектрального уплотнения (WDM – Wavelength-division multiplexing) и пространственного мультиплексирования (SDM – Space-division multiplexing). Согласно [5,6], к 2037 г. прогнозируется достижение пропускной способности передачи по волоконно-оптическим линиям связи (ВОЛС) при использовании технологии WDM от 32 до 400 Тбит/с, а при дополнительном применении технологии SDM, 5-100 Пбит/с. При прогнозировании перспектив развития различных технических инноваций в области телекоммуникаций, а также жизненного цикла технологий оказывается полезным анализ динамики научных публикаций и патентов в данной области [7,8]. В частности, в [9] убедительно показано, что жизненный цикл совершенно разных технологий начиная от разработки и внедрения до вывода из производства, описывается общими закономерностями, которые могут быть определены из анализа динамики патентов в данной области. Очевидно, что кроме патентов обширную информацию об уровне развития технологии можно получить из анализа наукометрической информации [10,11]. В работах [12,13] было показано, что анализ публикационной активности, индексируемой наукометрической базе данных SCOPUS, позволяет выявлять, как уровень развития технологии в разных странах, так и организации, в которых формируются основные идеи данного направления, а также основные источники финансирования этих работ. В связи с этим представляется важным на примере двух телекоммуникационных технологий, одна из которых появилась значительно раньше, чем другая, проанализировать динамику их развития и выявить общие тенденции и дальнейшие перспективы развития данных направлений. Поэтому целью представленной работы является сравнительный анализ развития технологий WDM и SDM,

позволяющий выявить перспективные разработки, а также оценить уровень технологической зрелости этих технологий в разных странах. Для достижения поставленной цели используется методика наукометрического анализа публикаций подробно описанная в работах [7,12].

1. Технология WDM и SDM

Передача больших объемов данных на расстояниях свыше 50 км наиболее эффективно реализуется по ВОЛС. Скорости оптической передачи данных в ВОЛС экспоненциально увеличивались за последние несколько десятилетий. При этом важным аспектом всегда являлось использование определенных методов мультиплексирования для повышения пропускной способности, желательно без прокладки дополнительных волокон [14]. Наиболее распространенными способами мультиплексирования являются временное мультиплексирование (*TDM – Time-Division Multiplexing*), *WDM* и *SDM*. Рассмотрим более подробно технологии *WDM* и *SDM*.

1.1. Технология WDM

Технология *WDM* позволяет увеличить пропускную способность системы передачи по оптическому волокну и увеличить количество каналов, в результате уменьшения расстояния межканального интервала. Здесь работает принцип, чем больше каналов, тем больше скорость в одном оптическом волокне. К сожалению, каналные скорости прямо пропорционально зависят от расстояния между несущими длинами волн. В связи с этим мы не можем бесконечно нагружать близкие друг к другу каналы [15]. Структурная схема, поясняющая принцип технологии *WDM* представлена на рисунке 1.

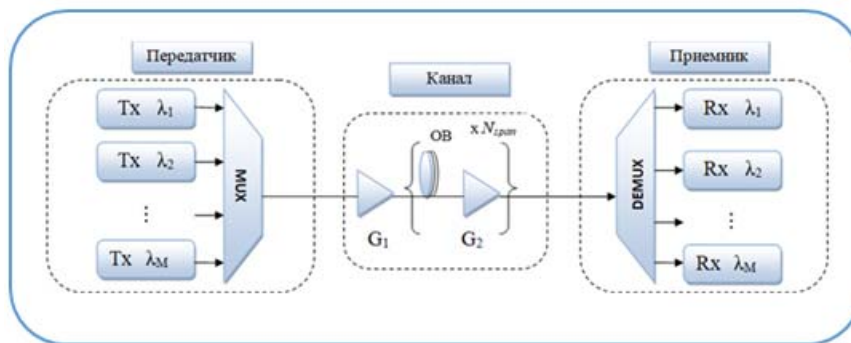


Рис. 1. Структура технологии WDM: Tx – передатчик, Rx – приемники

1.2. Технология SDM

Технологии *WDM* совместно с *SDM* являются многообещающим методом увеличения пропускной способности систем ВОЛС [16,17]. *WDM* системы в настоящее время приближаются к своему фундаментальному пределу. Следовательно, необходимы новые разработки, чтобы поддерживать рост трафика. *SDM* использует несколько сердцевин, практически при том же диаметре оптического волокна, для одновременной передачи независимых сигналов, поэтому пропускная способность может быть существенно увеличена. На рисунке 2 показана схема технологии *SDM* в сравнении с традиционными односердцевинными одномодовыми и многомодовыми волокнами.

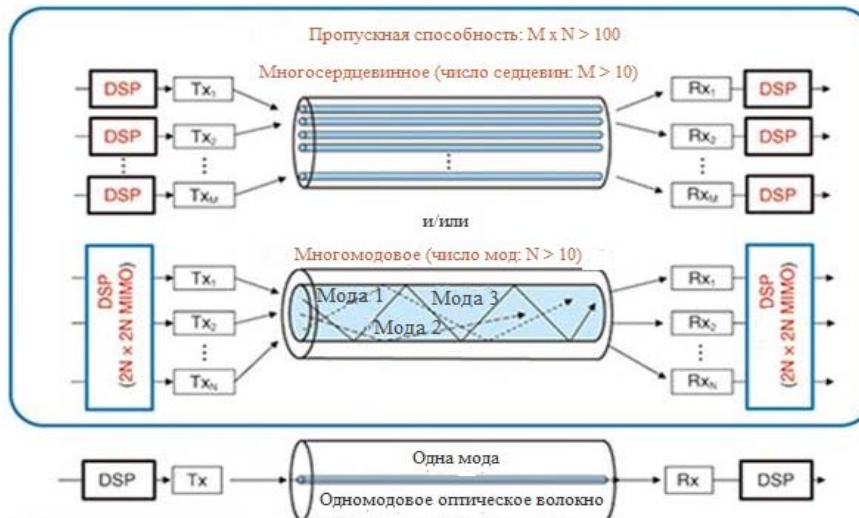


Рис. 2. Схема технологии *SDM* волоконно-оптических систем передачи:
DSP – цифровая обработка сигналов [17,18]

В качестве среды оптической передачи данных при применении технологии *SDM* используются многосерцевинные волокна (*MCF – Multi-Core Fiber*). Основная предпосылка *SDM* заключается в том, что если мы можем разместить N отдельных пространственных каналов в одном волокне, то общая пропускная способность волокна может быть увеличена в N раз [16, 19].

2. Методика исследований

В настоящее время для определения точек роста и основных тенденций в конкретной научной области с успехом применяются методики наукометрического анализа, основанные на данных наукометрической базы SCOPUS [8,12,20]. Подробное обоснование методики, которая использовалась в настоящей работе приведено в работе [6]. Для выявления актуальности исследований в области технологий *WDM* и *SDM* для ВОЛС и оценки общего потенциала этих направлений изучался поток научно-технической литературы за период 2017 по 2021 гг. с применением инструментов анализа базы данных Scopus.

С этой целью строились зависимости числа научных публикаций, и выдачи патентов от времени, после чего анализировалось распределение публикаций в выбранных областях: по странам; организациям, в которых проводились исследования; спонсорам, финансирующим работы в области *WDM* и *SDM*. Кроме этого, для оценки научного уровня публикаций анализировались журналы, в которых опубликованы результаты исследований по этим технологиям, а для выявления проникновения технологии в смежные области науки проводился анализ распределения публикаций по областям знаний, и изучались наиболее часто встречающиеся ключевые слова в публикациях, связанных с технологиями *WDM* и *SDM*.

3. Результаты исследований потока научно-технической литературы по технологиям *WDM* *SDM*

3.1. Динамика мирового потока публикаций по технологиям *WDM* и *SDM* за период 2017-2021 гг.

На рисунках 3 и 4 представлено количество публикаций по технологиям *WDM* и *SDM*, соответственно за период 2017-2021 гг. на основании базы данных Scopus согласно поисковому запросу: *TITLE-ABS-KEY (space AND division AND multiplexing) OR TITLE-ABS-KEY (multi AND core AND fiber)*. Результаты поиска показали, что в выборку попало 4 409 публикаций для технологии *WDM* и 2 807 публикаций для технологии *SDM*.

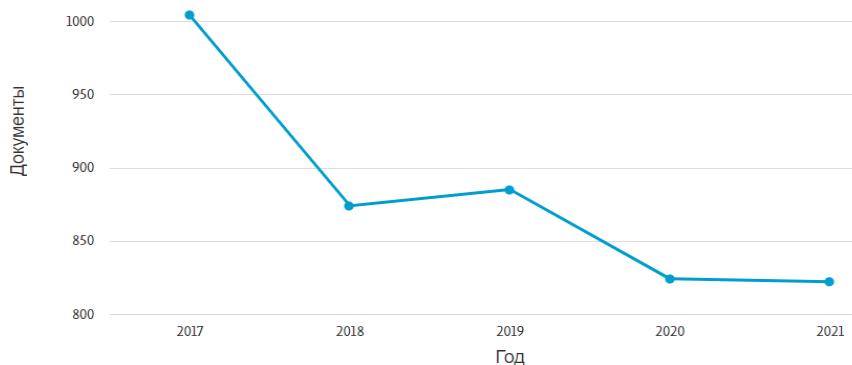


Рис. 3. Количество публикаций по технологии WDM за период 2017-2021 гг.

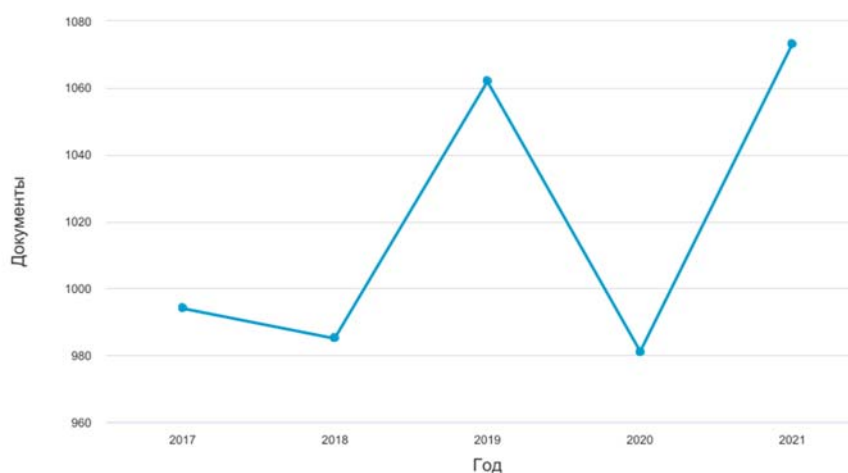


Рис. 4. Количество публикаций по технологии SDM за период 2017-2021 гг.

Как видно из рисунка 3 на интервале с 2017 по 2021 годы наблюдается падение интереса к исследованию в области технологии WDM. Это объясняется достаточно высокой проработкой технологии WDM, в результате чего дальнейшее развитие этих систем сдерживают два фактора: первый, состоит в том, что в традиционной технологии оптической передачи возникает ограничение спектральной эффективности из-за нелинейных эффектов в оптическом волокне. Вторым фактором является теоретический предел Шеннона. Традиционная теория связи утверждает, что верхний предел спектральной эффективности по отношению к фиксированному отношению сигнал-шум (SNR) определяется теоретическим пределом Шеннона [18].

Однако в волоконно-оптических системах передачи (ВОСП) увеличение мощности сигнала для пропорционального увеличения пропускной способности связи создает перекрестные помехи между сигналами WDM и искажает формы сигналов из-за нелинейных эффектов в оптическом волокне, тем самым предотвращая передачу сигналов на большие расстояния [22-28]. Эти нелинейные оптические эффекты в таких ВОСП дополнительно ограничивают спектральную эффективность сверх значения, определяемого пределом Шеннона. Наоборот, для технологии SDM (рис. 4) наблюдается значительный рост интереса к исследованию данной технологии с 2017 года, но в 2020 году произошел спад из-за пандемии Ковид-19.

Таким образом, наблюдаются колебания интереса к данной тематике, что характерно для достаточно новых направлений и технологий, которые находятся на стадии становления и выхода на рынок. В целом, поиск по ключевому слову SDM дает заниженные результаты по количеству статей, это обусловлено тем, что в статьях с MCF не всегда указывают ключевые слова *space division multiplexing*, поэтому мы расширили диапазон ключевых слов и добавили поиск по слову *multi core fiber*.

На сегодняшний день, именно многосердцевинные оптические волокна являются одним из самых перспективных направлений в пространственном уплотнении каналов [29-35]. Эта технология открывает новые возможности для одновременной реализации классических и квантовых каналов, что может быть особенно актуальным при создании протяженных квантовых сетей.

3.2. Распределение публикационной активности по источникам в области технологий WDM и SDM за период 2017-2021 гг. (таблица 1 и таблица 2)

Таблица 1

Распределение публикационной активности по источникам в области технологии WDM за период 2017-2021 гг.

Название журнала	Количество публикаций
<i>Journal of Lightwave Technology</i>	331
<i>Proceedings of SPIE the International Society for Optical Engineering</i>	248
<i>Optics Express</i>	167
<i>Journal of Optical Communications</i>	146
<i>IEEE Photonics Technology Letters</i>	109
<i>Optics Communications</i>	99
<i>IEEE Photonics Journal</i>	88
<i>Optical Engineering</i>	75
<i>Optik</i>	75
<i>Journal of Optical Communications and Networking</i>	74

Таблица 2

Распределение публикационной активности по источникам в области технологии SDM за период 2017-2021 гг.

Название журнала	Количество публикаций
<i>Journal of Lightwave Technology</i>	155
<i>Proceedings of SPIE The International Society for Optical Engineering</i>	115
<i>IEEE Access</i>	89
<i>Optics Communications</i>	66
<i>Optics Express</i>	60
<i>Journal Of Optical Communications</i>	50
<i>Optical Fiber Technology</i>	49
<i>IEEE Photonics Journal</i>	43
<i>IEEE Transactions on Vehicular Technology</i>	39
<i>Journal of Optical Communications and Networking</i>	34

Как видно из приведенных таблиц, журналы, в которых публикуются результаты исследований в области технологий WDM и SDM, являются достаточно престижными и ожидаемо, что списки базовых журналов для этих двух технологий во многом совпадают. Приведённым списком журналов, можно руководствоваться при поиске и представлении результатов исследования в области WDM и SDM технологий.

3.3. Распределение публикационной активности по странам в области технологий WDM и SDM за период 2017-2021 гг.

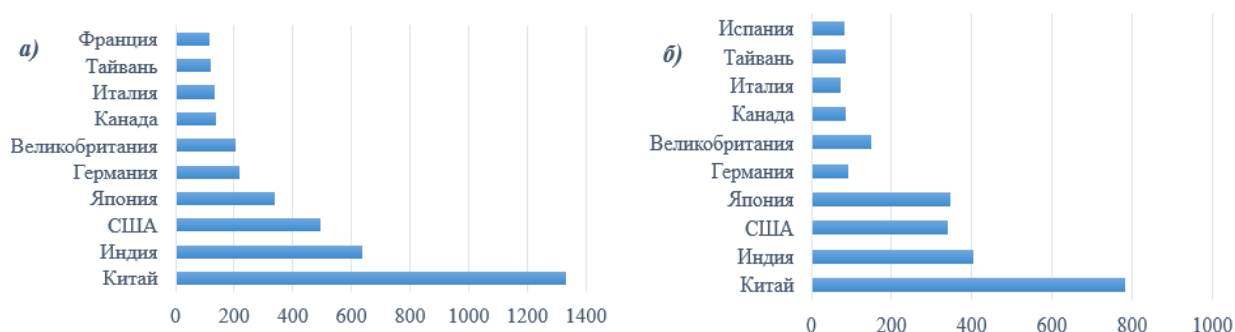


Рис. 5. Диаграмма распределения публикационной активности по странам на интервале времени период 2017-2021 гг.: а) технология WDM; б) технология WDM

На диаграммах, приведенных на рисунке 5, показаны страны-лидеры по публикационной активности в области телекоммуникационных технологий. Принадлежность публикации к той или иной стране определялась на основе аффилиации, которую указывают авторы, а если коллектив международный, то публикация засчитывается всем странам, присутствующим в списках аффилиаций данной статьи. Диаграммы на рисунке 5 показывают, что Китай и США пытаются развить все направления и не проходить мимо перспективных разработок в

телекоммуникациях. Значительный вклад в развитие телекоммуникационных технологий вносят исследовательские коллективы из Индии.

Важную составляющую в части публикаций по тематике *SDM* обеспечивает Япония, которая является одним из лидеров в производстве высокотехнологичных оптических волокон, которые формируют современную инфраструктуру ВОЛС. В частности, компания Fujikura, которая проводит ряд исследований с многосердцевидными оптическими волокнами и поставляет их в другие страны.

3.4. Распределение публикационной активности по финансирующему спонсору в области технологий *WDM* и *SDM* за период 2017-2021 гг. (таблица 3 и таблица 4)

Таблица 3

Распределение публикационной активности по финансирующему спонсору в области технологии *WDM* за период 2017-2021 гг.

Название организаций	Количество публикаций
<i>National Natural Science Foundation of China</i>	744
<i>National Key Research and Development Program of China</i>	162
<i>Fundamental Research Funds for the Central Universities</i>	135
<i>Horizon 2020 Framework Programme</i>	134
<i>Engineering and Physical Sciences Research Council</i>	82
<i>National Science Foundation</i>	79
<i>Japan Society for the Promotion of Science</i>	73
<i>Ministry of Science and Technology, Taiwan</i>	65
<i>European Commission</i>	52
<i>Beijing University of Posts and Telecommunications</i>	43

Таблица 4

Распределение публикационной активности по финансирующему спонсору в области технологии *SDM* за период 2017-2021 гг.

Название организаций	Количество публикаций
<i>National Natural Science Foundation of China</i>	465
<i>Horizon 2020 Framework Programme</i>	114
<i>National Key Research and Development Program of China</i>	85
<i>Fundamental Research Funds for the Central Universities</i>	73
<i>Japan Society for the Promotion of Science</i>	64
<i>National Institute of Information and Communications Technology</i>	55
<i>Engineering and Physical Sciences Research Council</i>	55
<i>National Science Foundation</i>	50
<i>Ministry of Science and Technology, Taiwan</i>	46
<i>Beijing University of Posts and Telecommunications</i>	43

Из приведённых в таблицах 3 и 4 данных, отображающих финансирование работ различными организациями, видно, что фонды Китая основательно финансируют исследования в области телекоммуникаций. При этом Китай целенаправленно поддерживает работы, как фундаментального, так и прикладного характера, которые направлены на увеличение пропускной способности оптических линий связи. Стоит заметить, что 90% организаций из таблицы 1 также присутствуют в таблице 2. Из этого можно сделать вывод, что данные организации занимаются исследованиями сразу по нескольким направлениям.

3.5. Распределение публикационной активности по отраслям знаний, связанных с технологиями *WDM* и *SDM* за период 2017-2021 гг.

На рисунке 6 приведены диаграммы, иллюстрирующие относительное распределение публикаций в период 2017-2021 гг. в отраслях научных знаний, определяющих развитие технологий *WDM* и *SDM*.

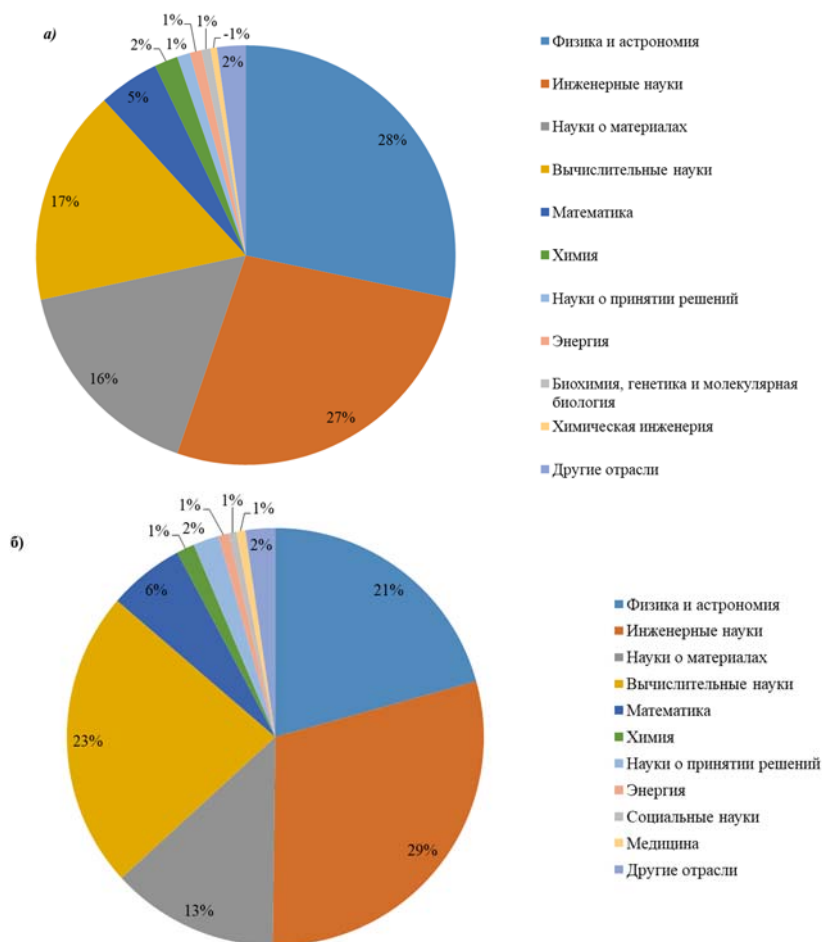


Рис. 6. Диаграмма распределения публикационной активности в период 2017-2021 гг. по отраслям знаний, связанным с развитием технологий: а) WDM; б) SDM

Заключение

На основе сравнительного анализа технологий WDM и SDM для ВОЛС установлено, что эти технологии находятся на разных стадиях жизненного цикла. В частности, пик научного интереса к технологии WDM уже пройден, а для технологии SDM он, еще не достигнут. Поэтому, прогноз развития технологии SDM и уровня вовлеченности научных коллективов в данную проблематику целесообразно проводить с учетом известной из прошлого динамики развития технологии WDM. В настоящее время большая доля исследований в области пространственного уплотнения оптоволокна приходится на научные группы из Китая, Индии, США и Японии, которые являются лидерами по числу публикаций в этой области.

С уверенностью можно сказать, что инициативу по пространственному уплотнению каналов перехватило многосердцевинное оптическое волокно. Оно занимает весомую часть фундаментальных и прикладных исследований в технологиях по увеличению пропускной способности ВОЛС, однако при этом следует заметить, что значительное количество работ по многосердцевинному волокну связано не с увеличением интернет трафика, а с применением многосердцевинного волокна в волоконно-оптических датчиках [24,25]. В мировом рейтинге стран, Россия по публикациям в области технологий WDM и SDM для ВОЛС не вошла в первую десятку, что связано во многом с низким финансированием фундаментальных и прикладных исследований по данным направлениям, что может привести к значительному технологическому отставанию в области создания высокоскоростных ВОСП.

Литература

1. Rabenandrasana J., Portnov E.L. The Choice of Optical Fiber with the Best Transmission Characteristics // IEEE, Systems of Signals Generating and Processing in the Field of on Board Communications. 2020, pp. 1-5.
2. Рабенандрасана Ж., Портнов Э.Л. Улучшения характеристик оптического волокна для передачи больших мощностей // Труды конференции «Телекоммуникационные и вычислительные системы» Международного форума информатизации МФИ-2019. МТУСИ, 2019. С.73-77.

3. Рабенандрасана Ж., Портнов Э.Л. Сравнение передаточных характеристик одномодовых оптических волокон для высокоскоростных передач // Международная научно-техническая конференция «Технологии информационного общества – ТИО-2020». М.: МТУСИ, 2020. С. 81-84.
4. Пушкарев А.В., Орлов В.Г. Эволюция телевидения высокой четкости // Телекоммуникации и информационные технологии. 2014. Т. 1. № 1. С. 26-28.
5. Winzer P.J., Neilson D.T., Chraplyvy A.R. Fiber-optic transmission and networking: the previous 20 and the next 20 years // Optics express, 2018, vol.26, no. 18, pp. 24190-24239.
6. Машиковцева Л.С., Болотов Д.В., Казанцев С.Ю., Колесников О.В., Миронов Ю.Б., Селюков А.С. Научометрический анализ публикаций по источникам одиночных фотонов для систем связи с квантовым распределением ключей // НТИ. Сер.1: Орг. и методика информ. работы, 2022. С. 22-31. DOI 10.36535/0548-0019-2022-01-3.
7. Андреев С. А., Казанцев С. Ю., Свистунова А. И., Тимовец П. А. Перспективы развития систем лазерной связи в атмосфере и гидросфере // Телекоммуникации и информационные технологии. 2022. Т. 9. № 1. С. 141-152.
8. Роза А.В. Возобновляемые источники энергии. Физико-технические основы: учеб. Пособие. – 2010.
9. Дымкова С.С. Новые принципы организации функционирования систем по продвижению результатов научных исследований // Системы синхронизации, формирования и обработки сигналов. 2017. Т. 8. № 1. С. 34-37.
10. Дымкова С.С. Повышение эффективности функционирования информационных систем и процессов в высшей школе // Методические вопросы преподавания инфокоммуникаций в высшей школе. 2017. Т. 6. № 2. С. 45-48.
11. Миронов 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. DOI 10.36724/2409-5419-2021-13-6-22-33.
12. Зайцев А.И., Коробов А.В., Рабенандрасана Ж. Источники одиночных фотонов: анализ и перспективных разработок для квантовых телекоммуникаций // Новые технологии высшей школы. Наука, техника, педагогика. Всероссийская научно-практическая конференция. М.: Московский политехнический университет, 2022. С. 145-150.
13. https://www.rp-photonics.com/space_division_multiplexing.html
14. Поляков А.В., Жуковский А.В. Нелинейные эффекты в оптоволоконных рециркуляционных запоминающих устройствах со спектральным уплотнением информационных каналов // проблемы физики, математики и техники № 2(11). Минск, 2012. С. 41-49.
15. Зайцев А. И., Зубилевич А. Л. Перспективы развития технологии пространственного уплотнения // Технологии информационного общества. 2021. С. 32-34.
16. Коришунов В.Н., Овчинникова И.А., Шишова Н.А. Применение многоканальных оптических волокон в инфокоммуникационном транспорте // REDS: Телекоммуникационные устройства и системы. 2021. Т. 11. № 4. С. 3-8.
17. Miyamoto Y., Takenouchi H. Dense Space-division-multiplexing Optical Communications Technology for Petabit-per-second Class Transmission // NTT Technical Review, 2014, vol. 12, no. 12, pp. 1-7.
18. Miyamoto Y., Kawamura R. Space Division Multiplexing Optical Transmission Technology to Support the Evolution of High-capacity Optical Transport Networks // NTT Technical Review, 2017, vol. 15, no. 6, pp. 1-7.
19. Машиковцева Л.С., Болотов Д.В., Казанцев С.Ю. [и др.]. Источники одиночных фотонов для систем связи с квантовым распределением ключей // XI международная конференция по фотонике и информационной оптике: Сборник научных трудов, Москва, 26–28 января 2022 года. М.: Национальный исследовательский ядерный университет "МИФИ", 2022. С. 547-548. EDN SEMOLM.
20. Васенин А.Г., Орлов В.Г. Компьютерная модель измерительного комплекса для ВОЛС // Фундаментальные проблемы радиоэлектронного приборостроения. 2010. Т. 10. № 1-1. С. 168-171.
21. Бондарь Д.В., Коришунов С.А., Дацов Ю.В. Перспективы применения волоконно-оптических датчиков физических величин в качестве средств измерения в системах обнаружения утечек // Наука и технологии трубопроводного транспорта нефти и нефтепродуктов. 2019. Т. 9. № 2. С. 191-199.
22. Крамаренко Б.А., Болдин А.П. Возможности статистического прогнозирования изменения общих и частных показателей безопасности дорожного движения при технической эксплуатации транспортных средств на основе использования программы STATISTICA-10-12 // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2021. № 3 (66). С. 3-12.
23. Лыгина Л.А., Строков Д.М., Котов А.А. Организация и безопасность движения в малых населенных пунктах сельского типа вдоль автомобильных дорог и режимы движения автомобилей // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2021. № 4 (67). С. 104-113.
24. Андронов В.Д. Надежность бетонных покрытий аэродромов после выполнения ремонтных работ // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2021. № 2 (65). С. 56-60.
25. Репин С.В., Шиманова А.А., Лутов Д.А. Методика повышения эксплуатационной надежности сложного технического объекта посредством анализа его структурной надежности // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2021. № 1 (64). С. 23-30.
26. Донченко В.В. Построение безопасных транспортных систем как идеология снижения смертности на автомобильных дорогах // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2021. № 3 (66). С. 114-121.
27. Бураков А.В., Соболев Е.Е. Развитие технологии санации деформационных швов и трещин жестких аэродромных покрытий // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2019. № 4 (59). С. 76-81.
28. Попов В.И., Чан Д.З. Эффективность применения пенополистирола в насыпях за интегральными устоями мостов // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2020. № 1 (60). С. 95-100.
29. Суладзе М.Д., Федулов В.К., Подопригора В.Г. Основы управления эксплуатационно-техническим состоянием аэродромных покрытий // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2020. № 4 (63). С. 64-71.

30. Ле Д.Л. Автомобилизация крупных агломераций и развитие транспортной инфраструктуры // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2020. № 1 (60). С. 3-10.
31. Корочкин А.В. Прогнозирование требуемой прочности дорожной одежды автомобильных магистралей // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2020. № 1 (60). С. 66-71.
32. Попов В.И., Чан Д.З. Современные способы снижения деформаций насыпей подходов к мостам // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2019. № 2 (57). С. 96-101.
33. Ушаков В.В., Горячев М.Г., Жустарёва Е.В. Конструктивные особенности нежестких дорожных одежд федеральных автомобильных дорог России со слоями из асфальтобетона // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2019. № 1 (56). С. 90-94.
34. Ларина Т.А., Зубарев Н.Р. Износ асфальтобетонных покрытий под воздействием колес легкового транспорта // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2019. № 3 (58). С. 55-61.
35. Немчинов М.В., Холин А.С. Принципы проектирования прочных и долговечных дорожных одежд // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). 2019. № 4 (59). С. 70-76.

APPLICATION OF SCIENTOMETRIC ANALYSIS TO IDENTIFY PROMISING DEVELOPMENTS IN THE FIELD OF OPTICAL FIBER OF COMMUNICATION LINE

Rabendasana Jocelin,

Assistant of the Department of DTE MTUCI, PhD, Moscow, Russia,

r.jocelin@mtuci.ru

Alexander I.Zaitcev

Postgraduate MTUCI, Moscow, Russia,

a.zaitcev96@yandex.ru

Abstract

The technologies of spectral multiplexing of communication channels and spatial multiplexing in fiber-optic communication lines are considered. The flow of scientific and technical literature in the field of research of technologies of spectral channel multiplexing and technologies of spatial multiplexing is analyzed. Based on the previously proposed methods for analyzing publication activity, an assessment of the technological involvement of various countries was carried out, and the stages of development of technologies for spectral and spatial multiplexing of communication channels were identified. Data on the dynamics of publications in the time interval of 2017-2022 are presented, which make it possible to predict the involvement of various teams in research on these topics. The groups of countries where the most competent research groups in the field of spectral division multiplexing and spatial multiplexing are working are determined. A list of organizations that most actively fund research on the considered technologies has been identified. The distribution of publication activity in the following fields of knowledge is presented: spectral multiplexing of communication channels and spatial multiplexing of communication channels. Based on a comparative analysis of the dynamics of the publication activity of the two directions, the main trends in the development of promising developments in the technology of spectral division multiplexing and spatial multiplexing are determined.

Keywords: *optical fiber of communication line, wavelength division multiplexing, space division multiplexing, capacity, optical fiber, scientometric analysis.*

МНОЖЕСТВЕННЫЙ ДОСТУП С РАЗДЕЛЕНИЕМ АБОНЕНТОВ ПО УРОВНЮ МОЩНОСТИ PD-NOMA

Петров Сергей Павлович,
аспирант МТУСИ, Москва, Россия,
serzh757@mail.ru

Бен Режеб Тауфик Бен Камель,
декан факультета РИТ МТУСИ, к.т.н., Москва, Россия,
t.benrejab@mtuci.ru

Аннотация

Неортогональный множественный доступ (NOMA) – это новый метод множественного доступа, который считается возможным кандидатом для 6-го и последующих поколений мобильной связи. Недавние исследования подтверждают, что метод обеспечивает больший прирост производительности в сравнении с используемыми в настоящее время методами ортогонального множественного доступа (OMA). В статье рассматривается производительность NOMA в нисходящей линии связи с использованием последовательного подавления помех SIC для демодуляции сигнала. Цель исследования – продемонстрировать эффективность демодуляции последовательного подавления помех в нисходящей линии связи в неортогональном множественном доступе. Приведены результаты математического моделирования помехоустойчивости на приемной стороне абонентского терминала в системе беспроводной связи с технологией NOMA.

Ключевые слова: частота битовых ошибок, PD-NOMA; неортогональный множественный доступ, последовательное подавление помех, SIC.

Введение

Интенсивный рост сотового трафика увеличил потребность в высокой пропускной способности сети, скорости передачи данных и пропускной способности [1]. С целью удовлетворения спроса на высокую пропускную способность, «3GPP» разработал несколько решений для систем мобильной связи 3.9 и 4-го поколения (4G), таких как *Long-Term Evolution (LTE)* и *LTE-Advanced*. Одним из таких решений является использование методов ортогонального множественного доступа (OMA – *Orthogonal Multiple Access*) для достижения высокой пропускной способности на системном уровне. Однако с увеличением объема мобильного трафика и внедрением таких технологий, как Интернет вещей (*IoT*), в настоящее время исследователи работают над дальнейшим развитием 4G в направлении систем мобильной связи 5-го и последующих поколений. Для дальнейшего усовершенствования технологий радиодоступа ученые изучают возможности новых методов множественного доступа. Одним из таких методов является неортогональный множественный доступ (NOMA) [3, 4], который считается многообещающим решением для достижения лучшей производительности системы, в сравнении с существующими методами множественного доступа. В NOMA по нисходящей линии связи абонентские терминалы (UE_s) для разделения сигналов нисходящей линии связи друг от друга используют область мощности (*Power Domain*).

Неортогональность этой схемы позволяет совмещать сигналы от разных пользователей для формирования единого сигнала передачи за счет использования области мощности [4]. Иначе говоря, передаваемая информация нескольких пользователей накладывается в области мощности путем неортогонального мультиплексирования по одному и тому же сигналу. Это дает совершенно новое решение для разделения сигналов разных абонентов и предоставления им доступа к базовой станции (*BS – Base Station*). Для правильного формирования сигнала неортогональным методом доступа необходимо использовать передовую технологию приема/передачи, например, последовательное подавление помех (*SIC – Successive Interference Cancellation*) [5, 6].

В данном исследовании использовался алгоритм SIC для получения информации в приемнике пользователя. Приемник абонента способен принимать два или более сигналов одновременно. Более сильный сигнал декодируется и вычитается из объединенного сигнала, затем демодулируется более слабый [7]. Чтобы осуществить последовательное подавление помех в NOMA на нисходящей линии связи, важно знать, какой пользователь будет выполнять SIC (это называется порядок декодирования). Вкратце, этот порядок определяется увеличением коэффициента усиления канала UE_s , который регулируется на стороне базовой станции [3]. Основываясь на этом порядке, пользователь, имеющий высокий коэффициент усиления канала, известный как ближний пользователь (*NU – near user*), декодирует сигнал пользователя с низким коэффициентом усиления канала, известного как дальний пользователь (*FU_s – far user*). Как только сигнал декодирован, NU отбрасывает нежелательную информацию с помощью SIC. Концепция SIC в нисходящей линии связи NOMA показана на рисунке 1.

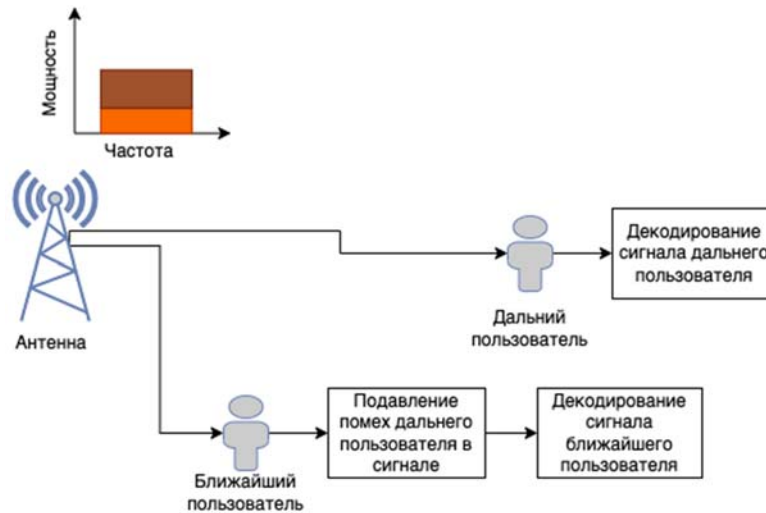


Рис. 11. NOMA нисходящей линии связи с SIC

На сегодняшний день было проведено несколько исследований по алгоритму SIC в NOMA [3], начиная с эволюции сотового множественного доступа от 3G до FRA (*Future Radio Access* – будущего радиодоступа). Результаты исследований отображены в таблице 1.

Таблица 2

Множественный доступ в сотовых связях от 3G до FRA

	3G	3.9/4G	FRA (Ожидается)
Вид множественного доступа	Non-orthogonal (CDMA)	Orthogonal (OFDMA)	Non-Orthogonal (NOMA)
Форма сигнала	Single Carrier	OFDM	OFDM
Адаптация связи	Быстрый TCP	AMC	AMC + Power Allocation
Изображение			

Соответствующее исследование было проделано [4] с применением концепции разделения сигнала антенны базовой станции на сектора для сопряжения пользователей и выполнения алгоритма SIC.

Большая часть исследовательских работ, проведенных в отношении последовательного подавления помех в NOMA, рассматривали идеальные условия, предполагающие, что помехи от дальних пользователей полностью устраняются на приемнике ближайшего пользователя, другими словами, ближнему пользователю известен сигнал дальнего пользователя.

В предлагаемой статье рассмотрим не идеальные условия для SIC. Эти условия означают, что информация о сигнале дальнего пользователя у ближнего пользователя искажается Рэлеевскими замираниями и шумами в канале связи. Результаты моделирования представлены в графиках частоты битовых ошибок (*BER*) на приемной стороне как для ближнего пользователя, так и для дальнего.

Модель

Рассмотрим структурную схему модели. Схема системы с передатчиком и приемником показана на рисунке 2. В рамках исследования рассматривается случай, когда базовая станция передает сигнал для j пользовательских устройств (UE_s) с мощностью передачи P_j . Тогда передаваемый сигнал можно выразить так:

$$s = \sum_{j=1}^N S_j \sqrt{P_j}, \tag{1}$$

где: s_j – кодированные суперпозицией сигналы пользователей j и $j \in [1, \dots, N]$.

В нашем случае, для упрощения рассмотрим $N = 2$, где $j = 1$ – пользователь с высоким коэффициентом усиления (ближний пользователь), а $j = 2$ является пользователем с низким коэффициентом усиления (дальний пользователь).

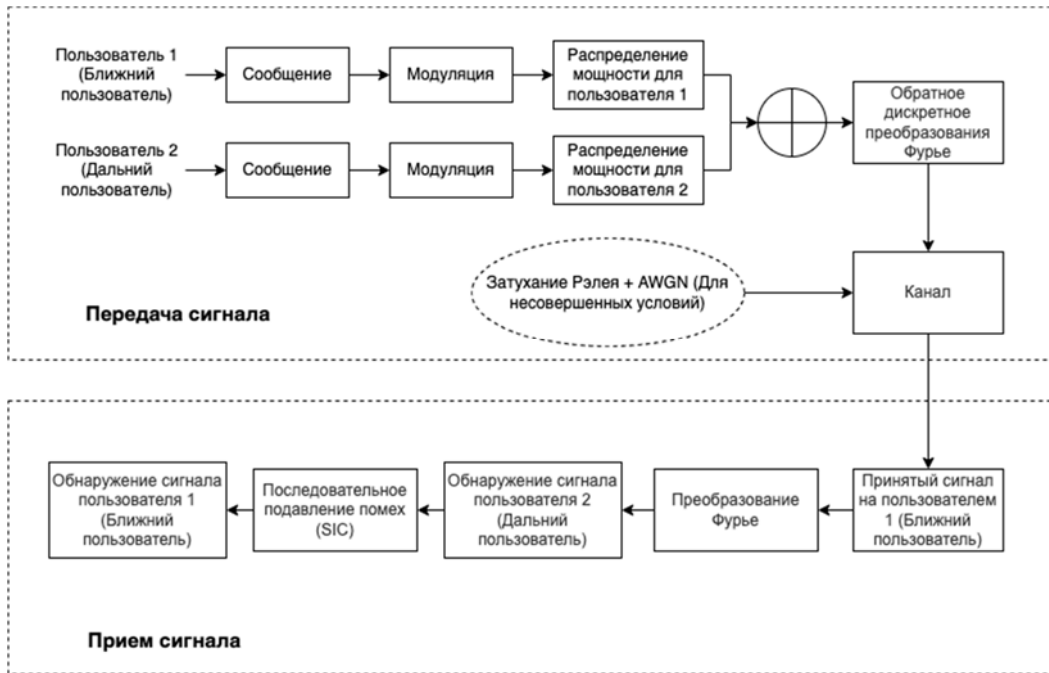


Рис. 12. Структурная схема передачи и приема сигнала

Принятый сигнал для j -пользователя представлен в виде:

$$y_j = h_j \times s + n_j, \tag{2}$$

$$y_j = h_j \times \left(\sum_{j=1}^N s_j \sqrt{P_j} \right) + n_j, \tag{3}$$

где h_j – комплексный коэффициент передачи канала между базовой станцией и j пользователями, n_j – аддитивный белый гауссовский шум (AWGN) с нулевым средним значением и единичной дисперсией, а мощность передачи P_j определяется как:

$$P_j = \frac{\alpha_j \times P_{BS}}{N_{SB}}, \tag{4}$$

где α_j обозначает коэффициент распределения мощности j -пользователя, P_{BS} – мощность передачи обслуживающей антенны, N_{SB} – количество поддиапазонов в системе.

В нисходящей линии связи NOMA последовательное подавление помех реализовано на приемнике абонента, поэтому важно знать необходимый порядок декодирования принятого сигнала. Этот порядок определяется согласно значениями комплексного передаточного коэффициента канала, нормированного к значению шума $|h_j|^2 / N_{0,j}$. Исходя из этого, любой пользователь с высоким коэффициентом усиления канала может правильно декодировать сигналы пользователей, появляющихся перед его собственным сигналом. Если учитывать, что в общем сигнале пользователей есть некоторый пользователь 1 с высоким коэффициентом усиления канала, то, пользователь 1 может декодировать сигналы других пользователей, чей нормированный коэффициент канала $|h_j|^2 / N_{0,j}$ ниже, чем нормированный коэффициент канала другого пользователя $|h_1|^2 / N_{0,1}$. В этом случае сигнал пользователя 1 будет представлен в следующем виде:

$$y_a = h_a \times s_a \sqrt{P_a} + h_a \times \left(\sum_{j=a+1}^N s_j \sqrt{P_j} \right) + n_j \cdot \quad (5)$$

Желаемый сигнал
Необходимо удалить с помощью SIC

Далее будут приведены результаты моделирования как для ближнего пользователя, так и для дальнего пользователя в нисходящем канале с рэлеевским некоррелированным замираниями при использовании приемника SIC.

Параметры математического моделирования

В этом разделе статьи приведено моделирование на уровне канала связи с целью оценки производительности ближнего и дальнего пользователей в нисходящем канале с рэлеевским некоррелированным замираниями при использовании приемника SIC. Для простоты предполагается сценарий с двумя пользователями. Параметры канала представлены в таблице 2

Таблица 3

Параметры канала

Канал/Шум	Рэлеевский с некоррелированными замираниями/AWGN
Модуляция	BPSK
Схема доступа	PD-NOMA
Мощность U1/U2	0.8/0.2
Дистанция до антенны U1/U2	1000/500
Диапазон SNR (dB)	0:40

Результаты математического моделирования

Очевидно, что если рассматриваемая система представляет собой двухпользовательскую нисходящую систему для NOMA, то график зависимости BER от SNR (dB) с использованием метода модуляции BPSK в рэлеевском канале будет таким, как показано на рисунке 3.

Мощность канала в нашей системе равняется единице, поэтому для каждого пользователя устанавливается свой коэффициент мощности: пользователь 1 (дальний) - 0.8 и пользователь 2 (ближний) – 0.2.

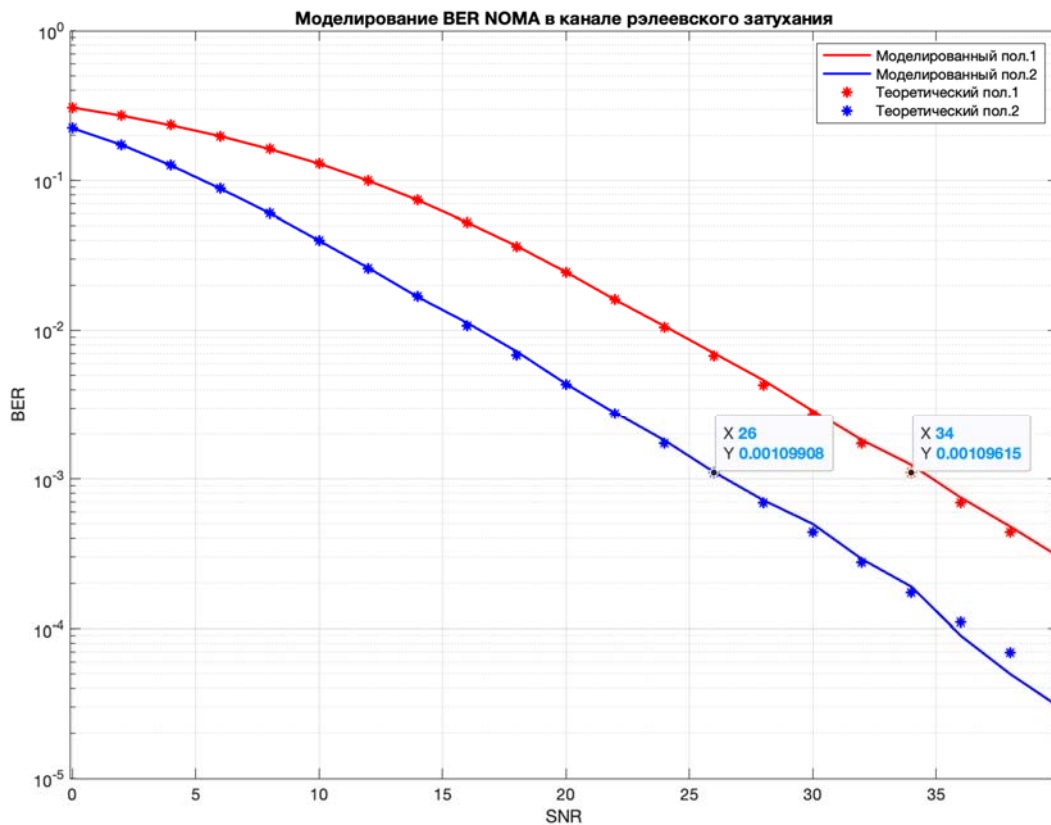


Рис. 3. Соотношение BER/SNR для нисходящего канала NOMA с модуляцией BPSK для двух пользователей

Результаты моделирования показывают, что BER для двухпользовательской системы NOMA меньше, чем у двухпользовательской системы OMA [9].

Также NOMA обеспечивает лучшую производительность, чем OMA, при меньшем использовании ресурсов, так как система с тремя пользователями TDMA (OMA) занимает три временных интервала для передачи их информации, а в NOMA требуется только один слот, так как информация всех пользователей передается одновременно.

Большая часть исследований нисходящей линии связи NOMA сосредоточена на передаче сигнала, а исследований на стороне приема крайне мало. Это обусловлено, прежде всего, сложностью, с которой приходится сталкиваться при приеме сигнала ближнего пользователя, особенно когда речь идет о случаях с двумя или более пользователями, так как ближний пользователь должен выполнять последовательное подавление помех для нескольких пользователей с низким коэффициентом усиления канала, которые накладываются на один и тот же сигнал. Подобные исследования целесообразно дополнить условиями кластеризации пользователей для получения результатов с наложением сигналов.

Заключение

В представленной работе исследован метод неортогонального множественного доступа (NOMA) и его реализация. Сравнение с методом ортогонального множественного доступа (OMA) и преимущества метода NOMA подробно освещены во многих предшествующих работах.

В статье рассмотрена модель с несколькими пользователями. Для простоты математического расчета взят случай с двумя пользователями и искажением информации в сигнале пользователя путем добавления эффекта канала затухания Рэлея.

Различные типы NOMA и их реализация с помощью алгоритма последовательного подавления помех (SIC) были разработаны с использованием математических представлений NOMA. Эти реализации могут быть весьма полезны для создания сетей 6G, а также сетей будущих поколений с высокой спектральной эффективностью, высокой скоростью передачи данных, низкой задержкой и высокой пропускной способностью.

На последующем этапе исследований стоит обратить особое внимание на реализацию эффективного алгоритма распределения ресурсов с целью кластеризации пользователей и оптимального распределения мощности между ними для обеспечения наиболее оптимальной производительности при минимальных требованиях к ресурсам.

Литература

1. *Usman M.R., Shin S.Y.* Channel allocation schemes for permanent user channel assignment in wireless cellular networks // IETE Journal of Research, vol. 62, no. 2, pp. 189-197, 2016.
2. Cisco Systems, "Cisco visual networking index: Global mobile data traffic forecast update, 2015-2020," White paper, San Jose, CA, USA, Feb. 2016.
3. *Saito Y., Kishiyama Y., Benjebbour A., Nakamura T., Li A., Higuchi K.* Non-orthogonal multiple access (NOMA) for cellular future radio access // IEEE 77th Vehicular Technology Conference (VTC Spring), pp. 1-5, June 2013.
4. *Li A., Lan Y., Chen X., Jiang H.* Non-orthogonal multiple access (NOMA) for future downlink radio access of 5G // China Communications, vol. 12, no. Supplement, pp. 28-37, 2015.
5. 3GPP TR36.814 (V9.0.0), "Further advancements for E-UTRA physical layer aspects," Mar. 2010. [Электронный ресурс]. – URL: <http://www.qtc.jp/3GPP/Specs/36814-900.pdf> (Дата обращения: 20.03.2022).
6. *Otao N., Kishiyama Y., Higuchi K.* Performance of non-orthogonal access with SIC in cellular downlink using proportional fair-based resource allocation // Wireless Communication Systems (ISWCS), 2012 International Symposium on, pp. 476-480, 2012.
7. *Souvik S., Santhapuri N., Choudhury R.R., Nelakuditi S.* Successive interference cancellation: a back-of-the-envelope perspective // Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks, p. 17, 2010.
8. *Yan C., Harada A., Benjebbour A., Lan Y., Li A., Jiang H.* Receiver Design for Downlink Non-Orthogonal Multiple Access (NOMA) // Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st, pp. 1-6, 2015.
9. *Reddy P.V., Reddy S., Reddy S., Sawale R.D.* Analytical Review on OMA vs. NOMA and Challenges Implementing NOMA // Conference: 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), 2021.

MULTIPLE ACCESS WITH SEPARATION OF SUBSCRIBERS BY POWER LEVEL PD-NOMA

Sergey P. Petrov,
Postgraduate MTUCI, Moscow, Russia,
serzh_757@mail.ru

Ben Rejeb Taoufik Ben Kamel,
Dean of the Faculty of R&T MTUCI, PhD, Moscow, Russia,
t.benreheb@mtuci.ru

Abstract

Non-orthogonal Multiple Access (NOMA) is a new method of multiple access, which is considered a possible candidate for the 6th generation of mobile communications and subsequent. Recent studies confirm that the method provides a greater performance gain compared to the currently used orthogonal multiple access (OMA) methods. The article discusses the performance of NOMA in a downlink using serial interference suppression SIC for signal demodulation. The aim of the study is to demonstrate the effectiveness of demodulation of sequential interference suppression in a downlink in non-orthogonal multiple access. The results of mathematical modeling are demonstrated, namely, the results of noise immunity on the receiving side of the subscriber terminal in a wireless communication system with NOMA technology.

Keywords: bit error rate; PD-NOMA; non-orthogonal multiple access; sequential interference suppression; SIC.

ТЕНДЕНЦИИ РАЗВИТИЯ ТЕХНОЛОГИИ T-SDN

Деарт Владимир Юрьевич,
доцент кафедры СС и СК МТУСИ, к.т.н., Москва, Россия,
vdeart@mail.ru

Лысых Полина Олеговна,
магистрант МТУСИ, Москва, Россия,
polinalysyh@gmail.com

Аннотация

Транспортные SDN (T-SDN) постепенно занимают все более значимое место в транспортных сетях операторов. Это объясняется тем, что T-SDN расширяют возможности программно-конфигурируемых сетей, ориентированных на уровень IP/Ethernet, дополнительным функционалом управления оптическим уровнем. Использование управляемых оптических соединений позволяет более эффективно реализовывать такие сервисы как предоставление пропускной способности по требованию (BoD) для ЦОД или создание наложенных оптических сетей. В статье рассматриваются принципы построения T-SDN, приводится классификация решений T-SDN, наиболее важные решения рассматриваются подробно. Приведено текущее состояние стандартизации T-SDN. Сделана оценка текущих проблем внедрения технологии T-SDN в современные транспортные сети. Приведены пути дальнейшего развития технологии T-SDN в связи с построением сетей 5G и другими аспектами развития транспортных сетей.

Ключевые слова: SDN, T-SDN, OpenFlow+, программно-определяемая сеть, оптическая транспортная сеть, виртуализация сети NVF

Определение T-SDN

Transport SDN (T-SDN) — это основанная на SDN архитектура для контроля и управления транспортными сетями, которая обеспечивает управление как уровнем L3/L2 (IP/Ethernet), так и оптическим уровнем (OTN/WDM).

Форум *Optical Internetworking Forum (OIF)* определяет транспортную SDN (T-SDN) как подмножество функций архитектуры SDN, включающее соответствующие компоненты транспортной сети: маршрутизаторы, коммутаторы, фотонные коммутаторы, оптические мультиплексоры (ROADM).

В этом смысле мы определяем T-SDN как подмножество архитектуры SDN, которое включает в себя расширения абстракций, интерфейсов, протоколов и элементов плоскости управления, чтобы справиться с особенностями транспортных сетей и преодолеть ограничения OpenFlow в оптической области.

Общий план архитектуры T-SDN

SDN — это архитектура для проектирования, построения и управления сетями. SDN меняет каждый аспект современного сетевого мира и способствует прорывным инновациям во всех сетевых секторах — коммутация пакетов, сеть беспроводного доступа, корпоративная сеть, центр обработки данных, облачные вычисления, виртуализация, оптическая коммутация и транспортные сети.

Как правило, SDN состоит из трех основных плоскостей с определенными функциями и интерфейсами, которые представлены на рисунке 1 [1,2].

Ниже приведено описание архитектуры T-SDN на рисунке 1 в направлении от нижней к верхней плоскостям:

1) Уровень данных находится в нижней части архитектуры SDN, он отвечает за обработку пакетов в пути данных на основе политик, полученных от уровня управления. Плоскость данных состоит из физических или виртуальных сетевых элементов, пересылающих и обрабатывающих трафик, таких как коммутаторы, маршрутизаторы и промежуточные устройства. Плоскость данных пересылает, отбрасывает и модифицирует пакеты в соответствии с политиками, установленными плоскостью управления. Это возможно благодаря определению соответствующих интерфейсов, называемых южными интерфейсами (SBI). Через SBI плоскость данных предоставляет плоскости управления видимость и контроль над своими возможностями обработки и пересылки. В данном примере рассмотрены три домена, входящие в T-SDN: два пакетных домена, управляемых с помощью протокола OpenFlow и оптический домен, управляемый GMPLS.

2) Плоскость управления (CP): SDN перемещает плоскость управления из прошивки сетевых элементов и реализует ее в виде программного обеспечения. Плоскость управления на основе ПО обеспечивает программный доступ к сетевым ресурсам и политикам переадресации, а также делает сетевое администрирование быстрым и гибким. Плоскость управления состоит из логически централизованного контроллера сетевой операционной

системы (*NOS*) или *SDN*. В рассматриваемом случае, контроллер имеет три разных южных интерфейса: два с протоколом *OpenFlow*, для управления пакетными доменами и отдельный интерфейс с протоколом *GMPLS*. Плоскость управления абстрагирует низкоуровневые детали плоскости данных и предоставляет пул *API*-интерфейсов, называемых северными интерфейсами (*NBI*) или *NB-API*, для плоскости приложений.

Плоскость приложений— это место, где находятся приложения, определяющие правила сетевой переадресации (программное обеспечение). Приложения получают упрощенное (абстрактное) представление сети от контроллера *T-SDN* через *REST API* и на его основе реализуют логику управления для принятия решений, которые будут преобразованы контроллером. Приложения могут сами предоставлять другой уровень абстрактного управления сетью.

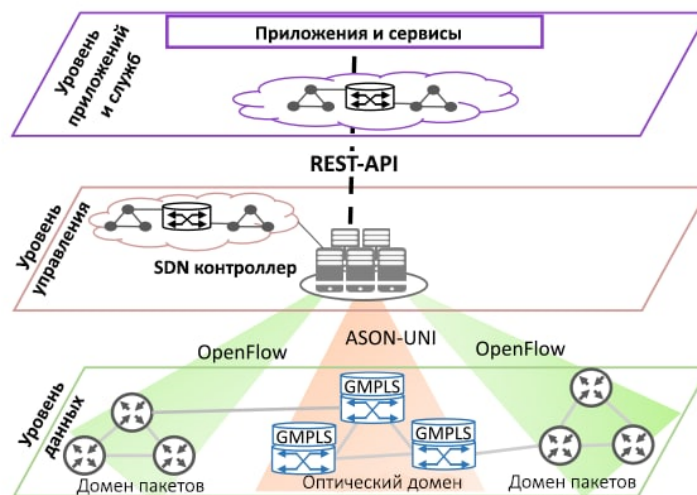


Рис. 1. Архитектура *T-SDN*

Благодаря программируемости, предлагаемой *SDN*, инновации возможны на уровне приложений. Существует широкий спектр приложений, уже предложенных и протестированных для различных сетевых доменов, которые были классифицированы как:

- управление трафиком,
- мобильность и беспроводная связь,
- измерение и мониторинг,
- безопасность и центр обработки данных.

Плоскость приложений представляет собой один из ключевых аспектов *SDN*, поскольку она предлагает возможность писать код на разных языках (которые используют сетевые *API*) для выполнения бизнес-аналитики, оптимизации и предоставления новых услуг. Такой программно-сетевой интерфейс не был доступен до *SDN*, и он меняет всю сетевую экосистему:

- 1) Сетевые операторы теперь имеют возможность избежать привязки к поставщику, разрабатывая свои собственные приложения, не зависящие от поставщика (или повторно используя приложения с открытым исходным кодом);
- 2) Сторонние вендоры имеют возможность выйти на рынок *SDN* с приложениями для оркестрации, аналитики, мониторинга и оптимизации;
- 3) Для крупных поставщиков (например, Cisco, Nokia, Juniper, Ciena, NEC и др.) плоскость приложения представляет собой еще одну основу для создания ценности для своих клиентов и дифференциации по сравнению с конкурентами.

Классификация *T-SDN*

На рисунке 2 представлена классификация решений *T-SDN* по их архитектуре плоскости управления: монолитные *T-SDN* (*SDON*), иерархический *T-SDN* (*HT-SDN*) и плоский/ячеистый *T-SDN* (*FT-SDN*).



Рис. 2. Классификация решений T-SDN на основе архитектуры плоскости управления

Монолитная архитектура (*SDON*): *SDON* была предложена исследовательскими усилиями и стала первым шагом в развитии транспортных *SDN*. В литературе мы можем найти термин *SDON (Software Defined Optical Networking)*, который относится к:

- одному контроллеру *SDN* в одном оптическом домене, основанному на расширениях, которые обеспечивают *SDN* на оптическом уровне 0, включающем программно-определяемые приемопередатчики, пути и *OXC*s вместе с расширениями к плоскости управления *SDN* и южные интерфейсы, например, *OpenFlow*.
- единый контроллер *SDN* в многоуровневой сети, обеспечивающий единую плоскость управления (*UCP*) *IP* и оптического уровней. Благодаря оптическому уровню с поддержкой *SDN* *SDON* может использовать преимущества *UCP* для *IP* и оптических уровней.

Иерархическая архитектура (*HT-SDN*): Органы стандартизации, занимающиеся *SDN* и транспортными сетями (в основном *ONF* и *OIF*), согласовали иерархическую архитектуру контроллеров для транспортных *SDN (HT-SDN)*. Иерархическая архитектура лучше подходит для многодоменной природы транспортных сетей, где несколько контроллеров домена (на основе *SDN* и устаревших) управляются либо родительским контроллером или оркестратором транспортной сети. Контроллер *SDON* становится контроллером домена в архитектуре *HT-SDN*.

Плоская или ячеистая архитектура (*FT-SDN*). Плоская архитектура уровня управления (*FT-SDN*) состоит из нескольких контроллеров домена с одноранговой координацией. Таким образом, в отличие от *HT-SDN*, который использует северный и южный интерфейсы для связи между контроллерами, *FT-SDN* использует интерфейсы Восток/Запад для однорангового взаимодействия между контроллерами *SDN*.

Плоские архитектуры плоскости управления не были в центре внимания на ранней стадии разработки *T-SDN*. Стандартизация интерфейсов восток/запад намного отстает от достижений в интерфейсах северного и южного направлений. Ожидается, что одноранговые отношения вызовут больший интерес для:

- кластеризации плоскости управления, которая поддерживается последней версией контроллера *OpenDaylight (ODL)* и контроллерами *ONOS (Open Network Operating System)* [3], но недостаточно изучена в литературе;
- координации между поставщиками услуг, при которой ожидается создание плоских архитектур среди поставщиков услуг.

Использование транспортных сетевых технологий

Транспортные сети используют множество различных технологий на нескольких уровнях:

- уровень 3 и уровень 2: *IP*, *Ethernet*, *MPLS* и *MPLS-TP*, которые обеспечивают статистическое мультиплексирование на пакетном уровне (*L2*).
- уровень 1 и уровень 0: *OTN* уровня 1, который поддерживает электрическое мультиплексирование с временным разделением *ODUk (TDM)*, оптическое мультиплексирование с разделением по длине волны (*WDM*) на уровне 0 и новые технологии гибких сетей.

Традиционно функции маршрутизации, сигнализации и защиты размещались на уровне *IP*, а оптический уровень обеспечивал статическую связь для устройств уровня 3. Однако гибкость и динамические возможности современных оптических устройств позволяют избежать пошаговой обработки *IP* за счет эффективной и динамической адаптации оптических соединений к требованиям трафика и обхода уровня *IP* всякий раз, когда это необходимо. Оптический байпас позволяет избежать энергопотребления, затрат, задержек и сложности поэтапной обработки *IP*-адресов.

Монолитная архитектура: *SDON*

Исследовательская деятельность началась в плоскостях управления *T-SDN* в 2009 году в Стэнфордском университете с так называемых расширений *Packet and Circuit Convergence (PAC.C)* для протокола *OpenFlow*. Первой задачей было теоретически и экспериментально доказать жизнеспособность и полезность перехода на *SDN*. *PAC.C*, решение, нацеленное на полностью централизованную архитектуру с одним контроллером *SDN*, основанным на встроенной поддержке *OpenFlow* в сетевых элементах и расширениях *OpenFlow* для управления потоками каналов и потоками пакетов. *PAC.C* приводит к слиянию доменов

пакетов и каналов в сеть с коммутацией потоков с *UCP*, которая выигрывает от видимости и контроля над *IP* и оптическими доменами.

В 2009 году *PAC.S* установила базовый подход к транспортной *SDN*. До 2014 года большая часть исследовательских усилий преследовала общую цель: обеспечить прямое управление плоскостью оптических данных с помощью контроллера *SDN/OpenFlow*.

Классифицируем решения, характеризующиеся использованием одного контроллера *SDN* для управления многоуровневыми транспортными сетями, как *SDON*. Основываясь на типе *SBI*, мы дополнительно классифицируем решения *SDON* на:

Единый южный интерфейс:

- Расширенный *OpenFlow+* (*OF+*) с собственной поддержкой.
- *OF+* с агентской поддержкой.
- *NETCONF* с поддержкой на основе агента.

Гибридные интерфейсы *SDN/GMPLS*

- Унаследованные интерфейсы для *GMPLS CP* (*GMPLS - Generalized Multi-Protocol Label Switching*) [4].
- *OF+* в направлении *GMPLS CP*.
- Устаревшие интерфейсы (*RSVP-TE* и *OSPF*) и *OF+* на основе агента в направлении *GMPLS CP*.

Иерархическая архитектура: *HT-SDN*

После успешной проверки концепции *SDON* акцент сместился на иерархические архитектуры контроллеров, которые мы называем в этой работе *HT-SDN*.

Иерархические архитектуры повышают масштабируемость и обеспечивают лучшую интеграцию разнородной среды доменов/уровней транспортных сетей за счет размещения специализированных контроллеров для каждого домена и уровня в иерархической структуре. Домен относится к автономной области сети, определяемой конкретной технологией уровня, поставщика, плоскости данных или плоскости управления. Оптический уровень может состоять из нескольких доменов, заданных поставщиками, гетерогенной плоскости данных и плоскости управления *SDN* и *GMPLS*.

Иерархическая архитектура *T-SDN* представлена на рисунке 3. Контроллеры домена отвечают за расчет внутримоментного пути и управление сложностью оптического домена. Каждый контроллер домена предоставляет родительскому контроллеру или оркестратору абстрактное представление своего домена. Родительский контроллер или приложение оркестратора транспортной сети координирует работу контроллеров домена, отвечает за расчет междоменных путей и создание сквозных служб.

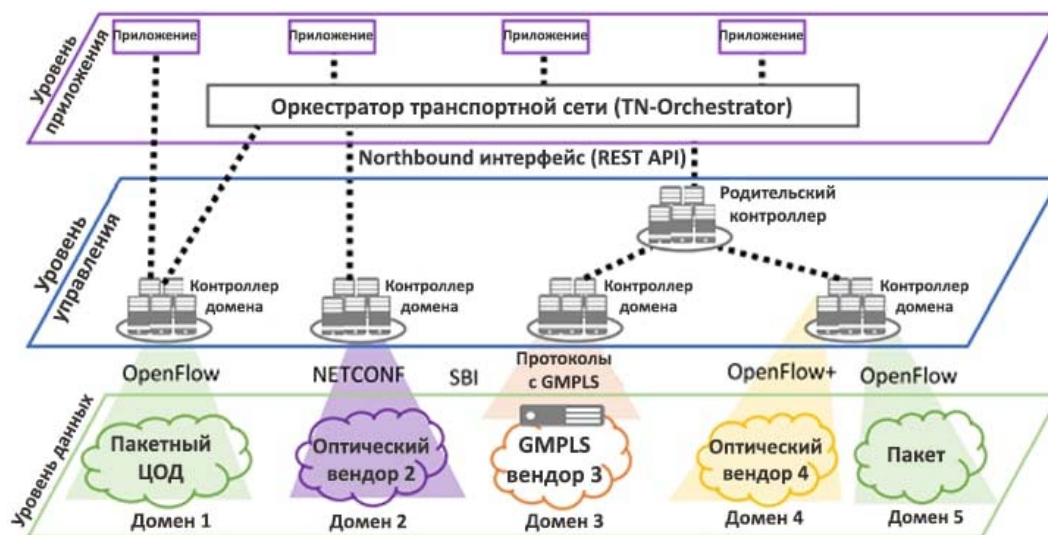


Рис. 3. Иерархическая архитектура *T-SDN*

Предлагаемая классификация *HT-SDN* на основе интерфейса, используемая для взаимодействия между иерархией контроллеров:

- Иерархия контроллеров *OF+*;
- Иерархия с сохранением состояния элемента вычисления пути (*PCE – Path Computation Element*) (*SH-PCE*);
- Гибридная иерархия *SDN/GMPLS*:
- Родительский контроллер *SDN* над контроллером домена на основе *PCE*;

- Контроллер/оркестратор *SDN* на гетерогенных контроллерах домена;
- *Orchestrator* на основе протокола *Orchestration Protocol (COP)* [5] через гетерогенные контроллеры домена.

Стандартизация

Только после 2013 года Организация по разработке стандартов начала выпускать документы, относящиеся к *T-SDN*. Основные организации по разработке стандартов (*SDO*), задействованные в *T-SDN*:

Open Networking Foundation (ONF) [6] поддерживает *T-SDN*, полностью основанную на *OpenFlow*, и уже начал вводить основные оптические функции в спецификации *OpenFlow ONF Transport API* (далее *ONF-TAPI*) является основной организацией для обеспечения функциональной совместимости в многодоменных, многоуровневых и многовендорных средах.

IETF также внесла свой вклад в разработку нескольких протоколов и технологий, которые можно повторно использовать для достижения *T-SDN*, например, *GMPLS*, протокола элементов вычисления пути (*PCEP*) и *NETCONF*. Еще одним важным вкладом *IETF* является многодоменная архитектура оркестровки *SDN* на основе *PCE* для сетевых операций на основе приложений (*ABNO*) [7]. Язык моделирования данных *YANG RFC 6020* становится важным компонентом для стандартизации моделей данных контроллеров, а также для создания стандартных северных и южных интерфейсов.

API для контроля и управления транспортными сетями являются основной темой, интересующей сетевых провайдеров, поскольку они способствуют снижению капитальных и эксплуатационных расходов их многоуровневой и мультивендорной транспортной инфраструктуры. Усилия по стандартизации транспортных *API (TAPI)* можно разделить на два типа:

- Общесетевые модели данных для северных *API*: предоставляются плоскостью управления и стандартизируются *IETF, ONF* и *OIF*;
- Модели данных сетевых элементов для южных *API*: предоставляются непосредственно транспортными сетевыми устройствами и стандартизированы проектом *Open Reconfigurable Optical Add-drop Multiplexers (ROADM)* [8] и рабочей группой *OpenConfig*.

Сетевые платформы с открытым исходным кодом

Еще одним важным компонентом эволюции *T-SDN* являются сетевые платформы с открытым исходным кодом. Сосредоточимся на основных проектах с открытым исходным кодом, которые поддерживают реальные решения *T-SDN*. Два основных контроллера операторского класса с открытым исходным кодом:

- *OpenDaylight*: становится общей платформой для решений поставщиков и новаторских демонстраций;
- *Открытая сетевая операционная система (ONOS)* – более молодой игрок, специализирующийся на поставщиках услуг и быстро набирающий популярность на рынке *T-SDN*.

Такие контроллеры вместе со спецификациями *ABNO* (от *IETF*) заполняют пробел медленного процесса стандартизации технологий *T-SDN*. Другие сетевые компоненты с открытым исходным кодом, которые также являются частью программно-определяемой трансформации:

- *OPEN-Orchestrator (OPEN-O)*: первый проект сквозного оркестратора служб с открытым исходным кодом, поддерживающий интеграцию как *NFV*, так и *SDN*;
- Архитектура расширенного контроля, оркестровки, управления и политики (*ECOMP*) с открытым исходным кодом. Ее цель – поддержка полной автоматизации и постепенное снижение зависимости от устаревшей *OSS*;
- Открытая платформа сетевой автоматизации (*ONAP*). Проект: Слияние Open Source *ECOMP* [9] и *OPEN-O* [10];
- Открытая платформа для *NFV (OPNFV)* [11]: разработка и развитие компонентов *NFV* в различных экосистемах с открытым исходным кодом;
- Платформа с открытым исходным кодом для аналитики сетевых данных (*PNDA*) [12] – платформа аналитики больших данных для сетей и сервисов.

T-SDN и виртуализация сетевых функций (NFV)

В то время как *SDN* отделяет управление от уровня данных, *NFV* отделяет программное обеспечение от аппаратного обеспечения. *NFV* – это парадигма сетевой архитектуры, которая использует методы виртуализации для динамического предоставления виртуализированных сетевых функций (*VNF*). *VNF* – это программные реализации физических сетевых функций (*PNF*), включая функции данных, контроля и управления, которые необходимы для работы сети. *VNF* можно динамически создавать в средах облачных вычислений с использованием готового коммерческого оборудования (*COTS*) вместо использования оборудования, зависящего от функций и поставщиков. *NFV* позволяет масштабировать ресурсы в зависимости от требований. Свобода размещения вычислительных ресурсов и ресурсов хранения в наиболее удобных местах и в наиболее удобные часы позволяет повысить эффективность работы. В этом подразделе дадим краткое введение в *NFV* и подведем итоги

исследований по *T-SDN* и *NFV*.

1. Праймер *NFV*: согласно Европейскому институту телекоммуникационных стандартов (*ETSI*), архитектура *NFV* состоит из трех основных элементов, изображенных на рисунке 4: инфраструктуры виртуализации сетевых функций (*NFVI*), *VNF*, управления и оркестровки *NFV* (*MANO*). *MANagement and Orchestration (MANO)* [8] с открытым исходным кодом – это сообщество под руководством оператора, которое предоставляет стек *MANO*, согласованный с архитектурами и информационными моделями *ETSI NFV*.

2. *T-SDN* и *NFV*: *NFV* можно реализовать без технологий *SDN*. Фактически, первая попытка внедрения *NFV* в транспортных сетях использовала не плоскость управления *SDN*, а плоскость управления *GMPLS/PCE*.



Рис. 4. Архитектура виртуализации сетевых функций

Архитектура *NFV* вместе с архитектурами виртуализации *T-SDN* использовались для виртуализации клиентских контроллеров *SDN* в облаке. Таким образом, виртуальные сети *T-SDN* могут быть предоставлены динамически по запросу, контроллер может быть перемещен из-за изменений в требованиях или для восстановления после аварии.

Оркестратор *T-SDN* и *NFV* для многопользовательской передачи сети. Этот оркестратор использует виртуализацию *NFV*, предоставляемую гипервизором оптической сети.

Архитектура *SDN-NFV* была распространена на мобильную сеть. Сеть радиодоступа (*RAN*) подключается к объектам центра обработки данных через транспортную сеть.

Транспортная сеть состоит из нескольких доменов с различными технологиями транспортных сетей. Была продемонстрирована виртуализация транспортной сети с использованием контроллеров *SDN*, подобных *VNF*, и расширенного пакетного ядра. Такая виртуализация может быть широко использована в мобильной сети, чтобы справиться с колебаниями трафика.

Проблемы транспортного *SDN*

SDN был специально определен для сетей с коммутацией пакетов на уровнях 3 и 2. Сегодня стандартизация *SDN* на основе *OpenFlow* активно поддерживается *ONF*, и существует растущий рынок коммерческих решений *SDN* на основе *OpenFlow*.

T-SDN предполагает поддержку уровней 3 и 2 и дополнительную поддержку сетей с коммутацией каналов на уровнях 1 (*SONET/SDH* и *OTN*) и 0 (оптический), что влечет за собой значительные проблемы по сравнению с решениями *SDN*, которые сосредоточены только на уровнях 3 и 2. Особую проблему представляет собой то, что на современных оптических сетях большинство установленных мультиплексоров *ROADM* не могут управляться протоколом *OpenFlow* и другими управляющими протоколами.

Безопасность — большая проблема в *T-SDN*. Несколько арендаторов могут использовать одну и ту же инфраструктуру на разных уровнях. Это все еще открытая область для анализа того, как принципы *SDN* могут применяться для повышения безопасности арендаторов. Например: сегменты сети должны обеспечивать степень изоляции, основанную на соглашениях об уровне обслуживания, и услуги аутентификации для клиентов. Введение в архитектуру *SDN* обеспечивает как потоки, так и возможности решения по безопасности благодаря централизованному уровню управления. Транспортные сети обладают высокой пропускной способностью и охватывают большие расстояния, поэтому они являются мишенью крупномасштабных атак.

Два основных препятствия, которые могут поставить под угрозу быстрое развертывание *T-SDN*, можно резюмировать следующим образом:

- внедрение *SDN* требует, чтобы устройства уровня данных были с поддержкой *SDN*. Поэтому это представляет собой важную инвестицию для операторов, которым необходимо обновить свое оборудование. Такие инвестиции в некоторых случаях может быть трудно поддерживать;
- неоднородность и сложность оборудования, в частности, фотонных коммутаторов, которые поддерживали бы протокол управления *OpenFlow+* или другой протокол управления оптическим уровнем.

Эти две проблемы сопровождаются отсутствием стандартизации *NBI*.

Переход от устаревших транспортных сетей к *T-SDN* все еще остается открытой проблемой, с которой сегодня сталкиваются поставщики телекоммуникационных услуг. Ожидается, что гибридное развертывание *T-SDN* будет наиболее многообещающим подходом, поскольку оно позволяет использовать устаревшие решения плоскости управления (такие как архитектуры *GMPLS/ASON* и *PCE*) без замены старого оборудования и использования полностью новых доменов.

Развитие *T-SDN*

Внедрение контроллера транспортного сегмента станет важным для максимального повышения операционной эффективности по мере перехода *CSP* к сетям *E2E 5G*. Под этим понимается преимущество обеспечения работы для гибридных развертываний *4G/5G* (то есть развертываний без прерывания услуг *4G*). Управление транспортным сегментом *5G* также даст возможность лучше и более согласованно обеспечивать выполнение соглашений *SLA* во всех реализациях сегмента транспортной сети. Контроллер транспортного сегмента обеспечит использование преимуществ *SDN* и автоматизации сети на основе реализаций намерений (*intent based*), а также программируемости сети при разделении транспортной сети на сегменты, улучшая соблюдение *SLA* на уровне каждого сегмента.

Устаревшая *OSS* будет постепенно заменена новой оркестровкой и автоматизацией *NFV-SDN*. *NFV* и *T-SDN* будут и впредь способствовать динамичному и масштабируемому развертыванию сервисов с дальнейшей интеграцией облачной и сетевой инфраструктуры. Операторы связи и поставщики услуг будут использовать больше проектов *Open Networking* для создания контроллеров, приложений, оркестраторов и даже *API*-интерфейсов. И последнее, но не менее важное: большие данные и машинное обучение будут постепенно обеспечивать сетевую аналитику и интеллектуальные функции на уровне приложений (или на уровне оркестровки) для повышения эффективности принятия решений по автоматизации транспортных сетей и услуг.

Фактически можно говорить о двух тенденциях развития технологии *T-SDN*:

1. Использование *T-SDN* для создания слайсов (срезов) в транспортных сетях мобильных сетей *5G* [14];
2. Срастание транспортной сети *T-SDN* с внутренней сетью ЦОДов, построенных на оптических коммутаторах *OCS* [15].

Заключение

Технология *T-SDN* уже применяется в современных сетях и обеспечивает следующие преимущества:

- Упрощение операций по работе с транспортной сетью.
- Оптимальное использование пропускной способности как на уровне пакетной сети, так и на уровне оптической сети.

Будущее *T-SDN* будет определяться иерархической и гетерогенной архитектурой, в которой элементы плоскости данных и плоскости управления используют и предоставляют четко определенные общие интерфейсы. В плоскости передачи данных транспортных сетей устаревшие устройства будут работать дольше, чем в других сегментах, но все чаще будет появляться новая волна устройств типа «белый ящик». Плоскость управления будет постепенно объединять все домены в сети, включая программно-определяемую систему *5G*.

Разделение транспортной сети на сегменты (*slicing*) станет фундаментальной основой предоставления различных услуг *E2E* поверхединой физической транспортной сети. Такой функционал имеет решающее значение для развития сетей *5G*, поскольку гарантирует операционную гибкость и эффективность при предоставлении услуг без вмешательства пользователя и персонала и с поддержкой соглашений об уровне обслуживания *SLA*, особенно в части, касающейся пропускной способности, величины задержки и уровня надежности. Контроллер транспортного среза становится ключевым строительным блоком для автоматизации создания, оптимизации и обеспечения гарантии безопасности транспортных сегментов в контексте мультидоменных/мультисегментных, мультитехнологичных и мультивендорных сетей *E2E 5G*.

Разделение плоскости управления и данных позволяет организовывать сквозные сервисы между доменами с использованием абстрактных представлений, предоставляемых северными *API*-интерфейсами плоскости управления. *HT-SDN* также является многообещающим решением для сосуществования *T-SDN* с другими устаревшими, но широко распространенными реализациями плоскости управления, такими как *GMPLS*. Таким образом, это также ключ к ускорению развертывания *T-SDN*.

Будущее *T-SDN* будет определяться иерархической и гетерогенной архитектурой. В плоскости передачи данных транспортных сетей устаревшие устройства будут работать дольше, чем в других сегментах, но все чаще будет появляться новая волна устройств.

Литература

1. Comprehensive Survey on T-SDN: Software-Defined Networking for Transport Networks. Rodolfo Alvizu, Guido Maier, Navin Kukreja, Achille Pattavina, Roberto Morro, Alessandro Capello, Carlo Cavazzoni // IEEE Communications Surveys & Tutorials. Volume 19, Issue 4, 2017, pp. 2232-2283.

2. Чернов И.В., Орлов В.Г. Особенности программно-конфигурируемых сетей // Телекоммуникации и информационные технологии. 2018. Т. 5. № 1. С. 21-25.
3. Смелянский Р.Л., Антоненко В.А. Концепции управления и виртуализации сетевых сервисов в современных сетях передачи данных. Учебное пособие. М.: КУРС, 2020. 160 с.
4. Mannie E. Generalized multi-protocol label switching (GMPLS) architecture // Internet Eng. Task Force, Fremont, CA, USA, RFC 3945, Март. 2013. [Электронный ресурс]. Режим доступа: <https://rfc-editor.org/rfc/rfc3945.txt>
5. Vilalta R. et al. The need for a control orchestration protocol in research projects on optical networking // Proc. Eur. Conf. Netw. Commun. (EuCNC), Париж, Франция, Июнь 2015. С. 340-344.
6. Open Networking Foundation [Электронный ресурс]. Режим доступа: <https://osm.etsi.org/>
7. Aguado A., Lopez V., Marhuenda J., de Dios O.G., Fernandez-Palacios J.P. ABNO: A feasible SDN approach for multi-vendor IP and optical networks // Proc. Opt. Fiber Commun. Conf., San Francisco, CA, USA, 2014. С. 1-3.
8. Open ROADMSA [Электронный ресурс]. Режим доступа: <http://openroadm.org/home.html>
9. Open ECOMP [Электронный ресурс]. Режим доступа: <http://www.openecomp.org/>
10. OPEN-Orchestrator Project [Электронный ресурс]. Режим доступа: <https://wiki.open-o.org/>
11. Open Platform for NFV (OPNFV) [Электронный ресурс]. Режим доступа: <https://www.opnfv.org/>
12. Open Source Platform for Network Data Analytics (PNDA) [Электронный ресурс]. Режим доступа: <http://pnda.io/>
13. Open Source MANO [Электронный ресурс]. Режим доступа: <https://osm.etsi.org/>
14. Коган С.С. Стандартизация решений и сегментирование транспортного уровня сети 5G // Первая миля, №2, 2021.
15. Ryohei Urata, Hong Liu, Kevin Yasumura, Erji Mao, Jill Berger, Xiang Zhou, Cedric F. Lam, Roy Bannon, Darren Hutchinson, Daniel Nelson, Leon Poutievski, Arjun Singh, Joon Ong, Amin Vahdat. Mission Apollo: Landing Optical Circuit Switching at Datacenter Scale. CoRR abs/2208.10041 (2022).

TRENDS IN THE DEVELOPMENT OF T-SDN TECHNOLOGY

Vladimir Y. Deart,

*Associate Professor of the Department of CN@CS MTUCI,
PhD, Moscow, Russia,
vdeart@mail.ru*

Polina O. Lysykh,

*Graduate MTUCI, Moscow, Russia,
polinalysyh@gmail.com*

Abstract

Transport SDN (T-SDN) are gradually taking an increasingly significant place in the transport networks of operators. This is due to the fact that T-SDN extend the capabilities of software-defined networks oriented to the IP/Ethernet layer with additional optical layer management functionality. The use of managed optical connections allows for more efficient implementation of services such as the provision of on-demand bandwidth (BoD) for data centers or the creation of superimposed optical networks. The article discusses the principles of T-SDN construction, provides a classification of T-SDN solutions, the most important solutions are considered in detail. The current state of T-SDN standardization is given. The assessment of the current problems of T-SDN technology implementation in modern transport networks is made. The ways of further development of T-SDN technology in connection with the construction of 5G networks and other aspects of the development of transport networks are given.

Keywords: SDN, T-SDN, OpenFlow+, software defined networking, optical transport network, network virtualization NFV

СРАВНЕНИЕ МЕТОДОВ ОПРЕДЕЛЕНИЯ УГЛОВЫХ КООРДИНАТ В СИСТЕМАХ РАДИОМОНИТОРИНГА

Носов Иван Юрьевич,
студент МТУСИ, Москва, Россия,
thevanes@mail.ru

Мирошниченко Антон Валерьевич,
ассистент кафедры МТС МТУСИ,
a.v.miroshnichenko@mtuci.ru

Аннотация

В статье рассматриваются и описываются методы определения угловых координат источника радиоизлучения в системах радиомониторинга. Приведено математическое моделирование определения потенциальной точности каждого метода и влияния угла раскрытия между двумя лучами на точность обнаружения источника. Получены выражения для потенциальной точности различных методов определения угловых координат источников радиоизлучения, в зависимости от ширины главного лепестка диаграммы направленности радиопеленгатора. Определены оптимальные, по критерию максимума точности угловых координат, углы раскрытия диаграмм направленности для равносигнального метода и метода минимума.

Ключевые слова: угловые координаты, диаграмма направленности, радиосигнал, метод максимума, метод минимума, метод сравнения, пеленгационная характеристика.

Введение

В настоящее время измерение угловых координат источника радиоизлучения имеет большое практическое применение. В радиолокации [1- 4] с помощью измерения угловых координат источника радиоизлучения определяют направление на объект, осуществляют слежение за объектом и его сопровождение. В системах связи [5-7] определяют местоположение передающего устройства в пространстве, а также местоположение источника помех. В навигации с помощью измерения угловых координат определяют собственное местоположение относительно радиомаяков и спутников. Таким образом, все вышеперечисленные области, связанные с контролем и обнаружением различных объектов объединяет одна и та же задача – определение направления прихода радиосигнала в пространстве и отсчет его угловой координаты.

Прежде, чем реализовывать на практике различные методы измерений угловых координат, необходимо провести их математическое моделирование. Такое моделирование позволит осуществить различные эксперименты, так как почти все параметры математической модели могут быть заданы в процессе исследований. Благодаря этому имеется возможность последовательно сравнить между собой различные методы радиопеленгации и их различные параметры, например, пеленгационные характеристики. Это позволит определить точность пеленгации и провести некоторые математические эксперименты путем изменения какого-то определенного параметра и нахождения зависимостей других параметров от изменяемого.

Методы определения угловых координат в системах радиомониторинга

Известны три основных метода определения угловых координат применяемые в системах радиомониторинга [10-12]: метод максимума, метод минимума, метод сравнения (равносигнальный метод).

Метод максимума. При измерении угловой координаты используется одна диаграмма направленности (рис. 1). Измерение угловой координаты θ_0 объекта происходит, когда выходное напряжение приемного устройства достигает амплитудного значения, то уровень сигнал окажется максимальным (рис. 2). Для этого диаграмму направленности антенны специально ориентируют своим максимумом на объект.

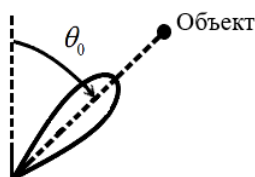


Рис. 1. Ориентация диаграммы направленности при измерении угловой координаты θ_0 объекта методом максимума

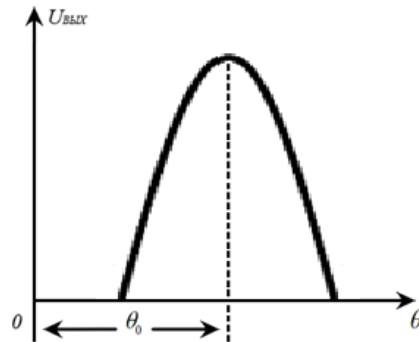


Рис. 2. Зависимость амплитуды выходного напряжения приемного устройства от текущего значения угла θ при использовании метода максимума

При измерении угловой координаты методом максимума обеспечивается высокий уровень принимаемого сигнала, но точность такого измерения не высокая.

Метод минимума. По названию метода понятно, что он в каком-то смысле является противоположным методу максимума. При измерении угловых координат методом минимума используются две диаграммы направленности.

Суть метода заключается в том, что две диаграммы направленности приёмной антенны принудительно ориентируются до того момента, когда минимум обеих (зачастую – просто нуль) диаграмм будет направлен на объект (рисунок 3). В этот момент амплитуда принимаемого радиосигнала окажется минимальной (рис. 4), что позволяет провести измерение угловой координаты θ_0 объекта.

При измерении угловой координаты методом минимума обеспечивается высокая точность измерения, но уровень принимаемого сигнала минимален.

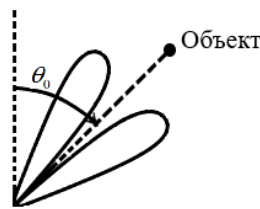


Рис. 3. Расположение диаграмм направленности в момент измерения угловой координаты θ_0 объекта методом минимума

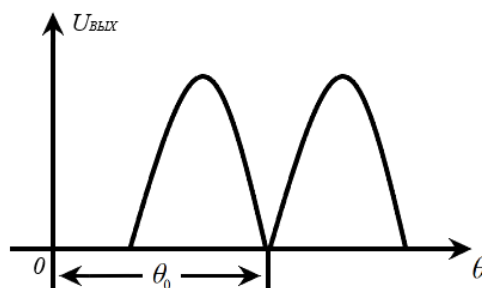


Рис. 4. Зависимость амплитуды выходного напряжения приемного устройства от текущего значения угла θ при использовании метода минимума

Равносигнальный метод. При измерении угловых координат равносигнальным методом так же, как и при методе минимума, используются две диаграммы направленности (рис. 5).

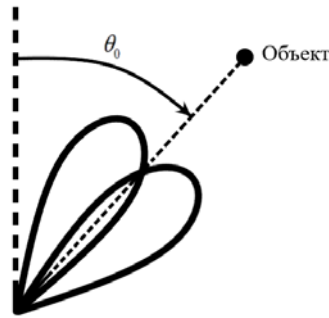


Рис. 5. Расположение диаграмм направленности в момент измерения угловой координаты θ_0 объекта равносигнальным методом

Измерение угловой координаты θ_0 происходит тогда, когда объект оказывается на направлении, при котором амплитуды принятых от объекта радиосигналов будут одновременно одинаковы для обеих диаграмм направленности. Математически это достигается путем решения уравнения

$$A(t)G_1(\Theta) = A(t)G_2(\Theta), \tag{1}$$

где $A(t)$ – сигнал источника радиоизлучения; G_1 и G_2 – усиление диаграмм направленности в направлении радиопеленгатора.

Выражение (1) можно представить в виде:

$$A(t)(G_1(\Theta) - G_2(\Theta)) = U_{\text{вых}}(\Theta) = 0. \tag{2}$$

Приняв в выражении (2) $A(t) = 1$, получим:

$$U_{\text{вых}}(\Theta) = G_1(\Theta) - G_2(\Theta). \tag{3}$$

Выражение (3) называют пеленгационной характеристикой для равносигнального метода, типовой график которой приведён на рисунке 6. [8,9].

Направление, при котором пеленгационная характеристика равна нулю, называется равносигнальным. Диаграммы направленности при этом пересекаются между собой на уровне 0.7 от их максимума. При измерении угловых координат данным методом обеспечивается высокая точность измерения, а сам сигнал имеет достаточно высокий уровень мощности [13].

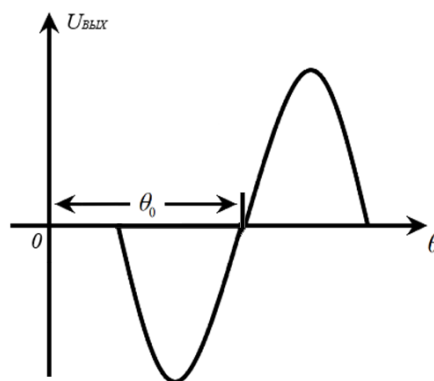


Рис. 6. Зависимость амплитуды выходного напряжения приемного устройства от текущего значения θ , при использовании равносигнального метода

Технически реализация данного метода может быть различной. Метод может быть реализован путем применения двух рупоров с фазовращателями на одной параболической антенне, путем применения двух облучателей или путем применения фазированных антенных решеток.

На рисунке 7 приведена структурная схема стандартной моноимпульсной радиолокационной станции, использующей равносигнальном методе пеленгации [17].

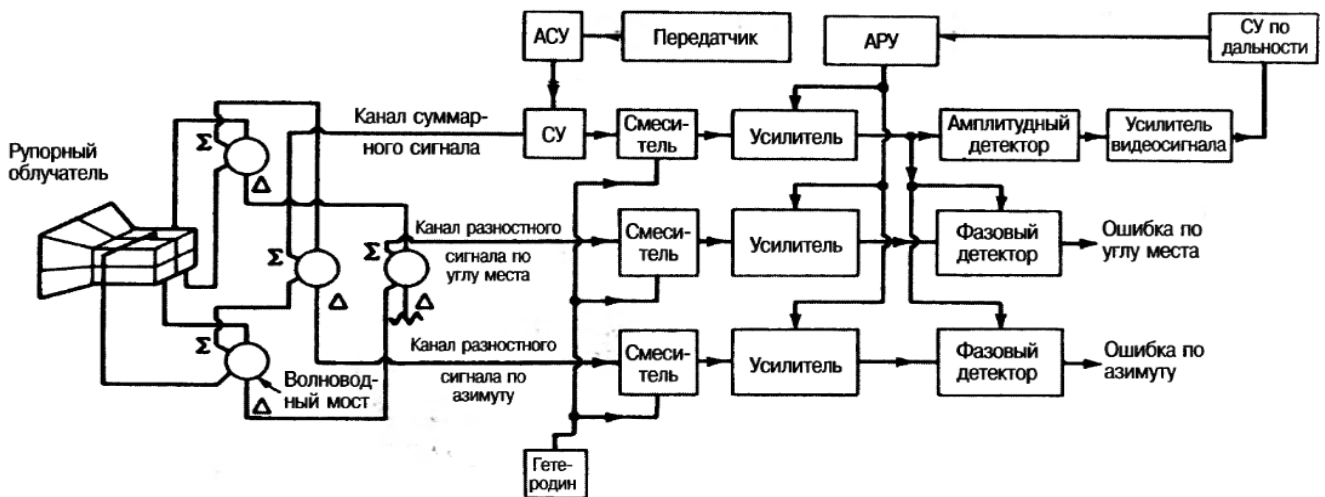


Рис. 7. Структурная схема стандартной моноимпульсной радиолокационной станции с равносигнальным методом пеленгации

Математическое моделирование методов определения угловых координат

Проведем математическое моделирование для определения точности различных методов радиопеленгации.

В соответствии с рекомендацией МСЭ-R [14], в случае, если аналитическое выражение для описания диаграммы направленности антенны (ДНА) радиолокатора неизвестно, и уровень первых боковых лепестков не превышает величину 20 дБ, следует описывать ДНА радиолокатора выражением:

$$G(\theta) = \frac{G_0 \sin(k\theta)}{k\theta}, \quad (4)$$

где G_0 – максимальное усиление антенны; k – масштабирующий коэффициент; θ – угол поворота антенны.

Величина $k\theta = \pi$ определяет ширину главного лепестка диаграммы направленности антенны.

Мощность сигнала на входе приемника радиопеленгатора задается как

$$P_{пр}(\theta) = P_0 R G(\theta), \quad (5)$$

где P_0 – мощность источника радиоизлучения; R – величина затухания сигнала при распространении;

G – усиление антенны; θ – угол между максимумом центром диаграммы направленности радиопеленгатора и направлением на источник радиоизлучения.

Приняв для простоты анализа $G_0 = 1$, $P_0 = 1$ Ватт, $R = 1$, из выражений (4) и (5) получим:

$$P_{пр}(\theta) = G(\theta) = \frac{\sin(k\theta)}{k\theta}, \quad (6)$$

где k – масштабирующий коэффициент; θ – угол поворота антенны.

Преобразуем выражение (6) для различных методов радиопеленгации:

для метода максимума:

$$P_{пр.макс}(\theta) = \left| \frac{\sin(k\theta)}{k\theta} \right|, \quad (7)$$

где k – масштабирующий коэффициент; θ – угол поворота антенны.

Для метода минимум:

$$P_{пр.мин}(\theta) = \left| \frac{\sin(k(\theta - 0.5\varphi_0))}{k(\theta - 0.5\varphi_0)} + \frac{\sin(k(\theta + 0.5\varphi_0))}{k(\theta + 0.5\varphi_0)} \right|, \quad (8)$$

где k – масштабирующий коэффициент; θ – угол поворота антенны; φ_0 – угол между двумя максимумами диаграммы направленности.

для равносигнального метода:

$$P_{пр.равн}(\theta) = \frac{\sin(k(\theta - 0.5\varphi_0))}{k(\theta - 0.5\varphi_0)} - \frac{\sin(k(\theta + 0.5\varphi_0))}{k(\theta + 0.5\varphi_0)}, \quad (9)$$

где k – масштабирующий коэффициент; θ – угол поворота антенны; φ_0 – угол между двумя максимумами диаграммы направленности.

В качестве показателя точности пеленгации можно взять общепринятую величину по «уровню 0.9», которую математически можно описать как:

$$P_{пр.макс}(\theta) \geq 0.9 \text{ - для метода максимума;} \quad (10)$$

$$P_{пр.мин}(\theta) \leq 0.1 \text{ - для метода минимума;} \quad (11)$$

$$|P_{пр.равн}(\theta)| \leq 0.1 \text{ - для равносигнального метода.} \quad (12)$$

На рисунке 8 приведены примеры пеленгационных характеристик для различных методов пеленгации, построенных по формулам (7-9).

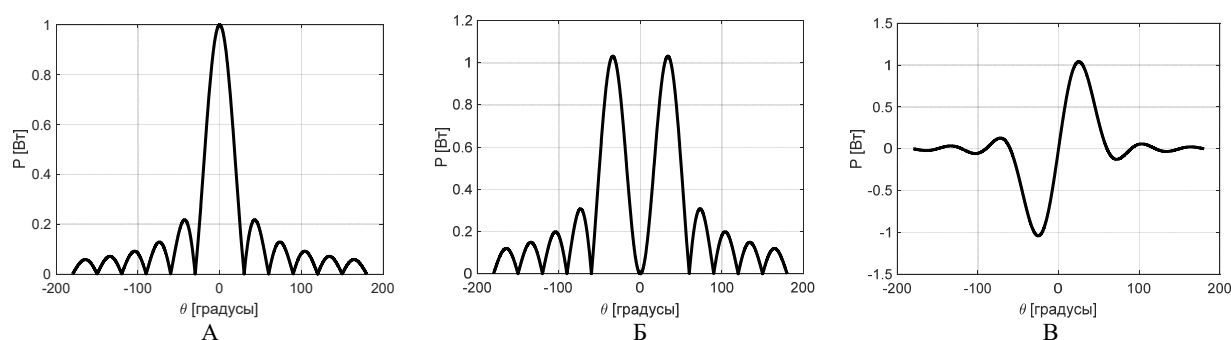


Рис. 8. Пример пеленгационных характеристик: А – метода максимума; Б – метода минимума; В – равносигнального метода

В [10, 11] показано, что оптимальное значение величины φ_0 по критерию максимальной точности пеленгации равно ширине главного лепестка диаграмм направленности антенн пеленгатора для метода минимума, и - около половины ширины главного лепестка диаграмм направленности антенн для равносигнального метода

В таблице 1 приведены значения точности пеленгации в градусах по критериям, принятым в выражениях (10), (11) и (12) для разных методов в зависимости от ширины диаграммы направленности пеленгатора

На рисунке 9 приведена визуализация точности пеленгации в зависимости от ширины диаграммы направленности антенны радиопеленгатора.

Таблица 1

Точность пеленгации в градусах для разных методов, в зависимости от ширины диаграммы направленности пеленгатора

Ширина ДНА пеленгатора (град)	5°	10°	15°	20°	25°	30°	35°	40°
метод «Максимума»	1.2°	2.4°	3.6°	4.9°	6°	7.4°	8.4°	9.4°
метод «Минимума»	1.1°	2.2°	3.4°	4.4°	5.6°	6.6°	7.8°	9°
метод «Равносигнальный»	0.2°	0.4°	0.5°	0.72°	0.96°	1.06°	1.3°	1.52°

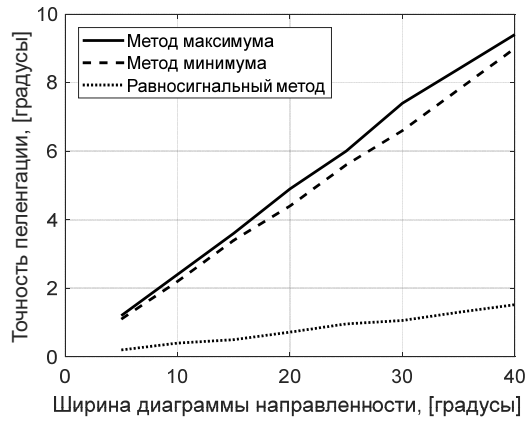


Рис. 9. Точность пеленгации для разных методов, в зависимости от ширины диаграммы направленности антенны пеленгатора

В таблице 2 приведены значения относительной точности пеленгации по критериям (10-12), в зависимости от отношения ϕ_0 /ширина ДНА.

Таблица 2

Относительная точность пеленгации, в зависимости от отношения (ϕ_0 /ширина ДНА)

Отношение ϕ_0 /ширина ДНА	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3
метод Минимума	-	-	-	-	-	4.2	1.5	1.2	1	1.8	2.5	2.9
метод Равносигнальный	2.4	1.6	1.3	1.2	1	1	1.1	1.2	1.5	1.9	2.7	4.9

На рисунке 10 приведены результаты визуализация относительной точности пеленгации в зависимости от отношения ϕ_0 /ширина ДНА.

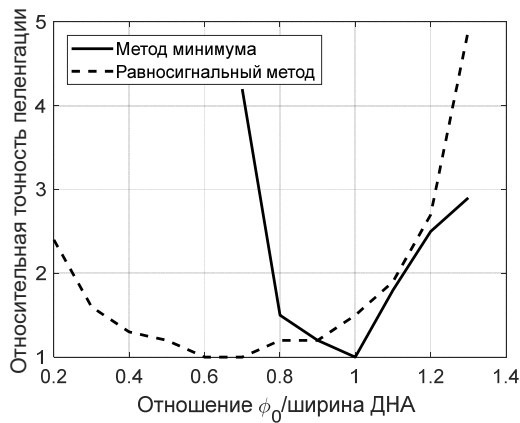


Рис. 10. Относительная точность пеленгации в зависимости от отношения ϕ_0 /ширина ДНА

Анализ результатов математического моделирования

Проведенное математическое моделирование позволило определить зависимость точности радиопеленгации от ширины диаграммы направленности антенны радиопеленгатора (табл. 1 и рис. 9) в одной из плоскостей. Для всех трех методов эта зависимость близка к линейной и аппроксимируется выражением

$$\Delta\Theta \approx \begin{cases} 0.25L_{\text{дна}} - \text{метод максимума} \\ 0.22L_{\text{дна}} - \text{метод минимума} \\ 0.04L_{\text{дна}} - \text{равносигнальный метод} \end{cases} \quad (13)$$

В общем случае, измерение угловых координат источника радиоизлучения проводится в двух ортогональных плоскостях: азимут φ и угла места β (рис. 11.).

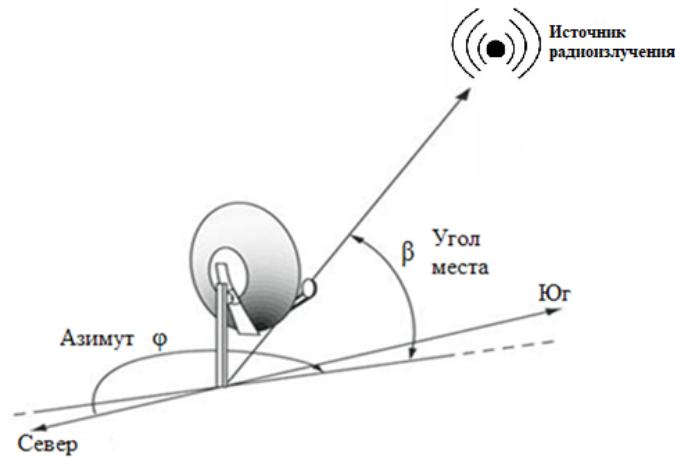


Рис. 11. Определение угловых координат источника радиоизлучения

Так как плоскости азимута и угла места являются ортогональными, то точность измерения каждой из угловых координат можно считать независимой и определяемой только шириной главного луча диаграммы направленности антенны радиопеленгатора.

В случае симметрии антенны радиопеленгатора в азимутальной и угломестной плоскости (что характерно, например, для бортовых радиолокационных станций некоторых самолетов) можно принять точность определения координат φ и β одинаковой. Однако в случае, если антенна радиопеленгатора имеет несимметричную форму, например, форму эллипса, то точность определения угловых координат будет отличаться в виду различий в ширине диаграммы направленности антенны в азимутальной и угломестных плоскостях.

Если в качестве примера радиопеленгатора задать параболическую антенну, характерную для 80-х годов XX века, симметричную в угломестной и азимутальных плоскостях, имеющую геометрические размеры антенны около 0.7 м и работающую на частоте 10 ГГц, то ширина диаграммы направленности ее антенны составляет около 3 градусов [14, 15]. Тогда, согласно выражению (13), можно получить предельную точность пеленгации для бортовой радиолокационной станции самолета:

$$\Delta\Theta_{сам} \approx \begin{cases} 0.25 \cdot 3 = 0.75^\circ - \text{метод максимума} \\ 0.22 \cdot 3 = 0.66^\circ - \text{метод минимума} \\ 0.04 \cdot 3 = 0.12^\circ - \text{равносигнальный метод} \end{cases} .$$

Полученные по критерию точности радиопеленгации результаты (табл. 2, рис. 9) подтвердили, что наибольшая точность достигается при выборе раскрыва антенны равным ширине диаграммы её направленности для метода минимума, и при выборе раскрыва равного 0.6-0.7 – для равносигнального метода. При этом отклонение от этих оптимальных значений на 0.1 величины ведет к почти двукратному ухудшению точности для метода минимума.

Равносигнальный метод является менее чувствительным к точности раскрыва лучей диаграммы направленности антенны и с некоторым ухудшением в точности (не более чем в 1,5 раза) может использоваться при отклонении величины раскрыва на величину ± 0.4 от оптимального.

Как видно из выражения (13) основным способом улучшения точности пеленгации является уменьшение ширины главного луча диаграммы направленности радиопеленгатора, что может быть достигнуто за счет увеличения геометрических размеров антенны радиопеленгатора, что не во всех случаях представляется возможным.

Другим повышением точности радиопеленгации может стать поиск и разработка новых методов, обеспечивающих рост точности пеленгации при сохранении габаритных параметров приёмных антенны [16-18].

Выводы

Полученные выражения и расчёты показывает, что точность равносигнального метода определения угловых координат примерно в пять раз выше, чем метода максимума. В тоже время, данный метод радиомониторинга требует создания многолучевой антенной системы, что не всегда целесообразно.

Метод минимума дает несколько лучшую точность по сравнению с методом максимума, но в тоже время, обладает рядом недостатков, например, отсутствием возможности оперативно определить пропадание пеленгуемого сигнала, что исключает целесообразность применение данного метода.

Литература

1. Егоров В.В., Попов В.А., Селезнев А.Г., Мирошниченко А.В., Татарчук И.А. Концепция создания авиационного боевого комплекса с ведущим пассивным каналом // Четвертая всероссийская научно-техническая конференция «Моделирование авиационных систем». Сборник тезисов докладов. 2020. С. 61-62.
2. Грибанов П.В., Сперанский В.С. Современные тенденции развития систем наблюдения за воздушной обстановкой // Телекоммуникации и информационные технологии. 2018. Т. 5. № 1. С. 16-20.
3. Ионов В.В., Пестряков А.В. Современное состояние систем космического мониторинга за подвижными объектами на территории России // Телекоммуникации и информационные технологии. 2019. Т. 6. № 1. С. 24-27.
4. Соколов В.Б., Калинин И.В., Киселев А.В. Авиационные комплексы радиолокационного дозора и наведения. Москва: Научно-информационный центр НИИАС, 1983. 82 с.
5. Егоров В.В., Мирошниченко А.В., Татарчук И.А. Измерение расстояний в системах навигации, наблюдения и связи // Четвертая всероссийская научно-техническая конференция «Моделирование авиационных систем». Сборник тезисов докладов. 2020. С. 165-166.
6. Григорьев И.Д., Орлов В.Г. Особенности метода доступа к среде в VDL MODE 4 // XI Международная отраслевая научно-техническая конференция «Технологии информационного общества». Сборник трудов. 2017. С. 455.
7. Фадеев А.Н., Орлов В.Г. Возможные подходы к повышению пропускной способности системы "ОВЧ ЛПД режима 4" // Фундаментальные проблемы радиоэлектронного приборостроения. 2016. Т. 16. № 5. С. 309-311.
8. Проскурин В.И., Ягольникова С.В., Шевчук В.И. Радиолокационное наблюдение. Методы, модели и алгоритмы. Москва: Радиотехника, 2017. 368 с.
9. Ширман Я.Д., Голиков В.Н. Теоретические основы радиолокации. Москва: Советское радио, 1970. 560 с.
10. Вакин С.А., Шустов В.Н. Основы радиопротиводействия и радиотехнической разведки. М.: Сов. Радио, 1968. 448 с.
11. Справочник по радиолокации. Под ред. Меррилла И. Скольника. Перевод с английского под общей редакцией В.С. Вербы. Т. 1. М.: Техносфера, 2015. 672 с.
12. Карташкин А.С. Авиационные радиосистемы. М.: РадиоСофт, 2015. - 302 с.
13. Справочник по радиолокации / под ред. М.И. Скольника. В 2 книга. Книга 1. М.: Техносфера, 2014. 672 с.
14. Рекомендация МСЭ-R М.1851-1 Математические модели диаграмм направленности антенн радиолокационных систем радиопредельная для использования при анализе помех. 2018. 26 с.
15. Ротхаммель К. Энциклопедия современных антенн. Москва: ДМК-Пресс, 2016. 812 с.
16. Егоров В.В., Попов В.А., Селезнев А.Г., Мирошниченко А.В., Татарчук И.А. Оценка точностных характеристик многопозиционной малобазовой системы высокоточного определения координат источников радиоизлучения методами математического моделирования // Четвертая всероссийская научно-техническая конференция «Моделирование авиационных систем». Сборник тезисов докладов. 2020. С. 19-21.
17. Мирошниченко А.В. Исследование потенциальных точностных характеристик малобазовых систем, построенных на триангуляционном способе определения координат источников радиоизлучения // Первая миля. 2020. № 6. С. 38-43.
18. Комарова Н.В., Мирошниченко А.В. Способ определения местоположения и вектора скорости полета летательного аппарата // В книге: Авиационные системы в XXI веке. Тезисы докладов юбилейной Всероссийской научно-технической конференции., 2022. С. 112-113.

ANALYSIS AND COMPARISON OF METHODS FOR DETERMINING ANGULAR COORDINATES IN RADIO MONITORING SYSTEMS

Ivan Y. Nosov,
Student MTUCI, Moscow, Russia
thevanes@mail.ru

Anton V. Miroshnichenko,
Assistant of the Department of MTS MTUCI, Moscow, Russia
a.v.miroshnichenko@mtuci.ru

Abstract

The article discusses methods for determining the angular coordinates of a radio emission source in radio monitoring systems. A description is attached for each method. Mathematical modeling was carried out to determine the potential accuracy of each method and the effect of the opening angle between the two beams on the detection accuracy. Expressions are obtained to determine the potential accuracy of various methods for determining the angular coordinates of radio sources, depending on the width of the main lobe of the directional pattern of the radio direction finder. The optimal angles of the radiation patterns for the equi-signal method and the minimum method are determined according to the criterion of maximum accuracy of determining angular coordinates.

Keywords: angular coordinates, radiation pattern, radio signal, maximum method, minimum method, comparison method, direction finding characteristic.

«Схемотехника и технологические решения»

УПРАВЛЯЕМЫЙ СВЧ АВТОГЕНЕРАТОР НА КОМПЛЕМЕНТАРНОЙ ПАРЕ МОП-ТРАНЗИСТОРОВ

Фан Туан Ань,
аспирант МТУСИ, Москва, Россия,
alexpta96@gmail.com

Логвинов Василий Васильевич,
доцент кафедры РОС МТУСИ, к.т.н., Москва, Россия,
adlerbasil@rambler.ru

Аннотация

В статье рассмотрен СВЧ автогенератор, реализованный по схеме Колпица на комплементарной паре МОП транзисторов с технологией 0.35мкм. Предложенное схемотехническое решение использует источник с пониженным напряжением питания (3 В) и минимальным числом применяемых компонентов. Управление частотой генерации достигается включением варактора в схему контура. Изменение емкости варактора под действием управляющего напряжения с выхода цифрового тракта приемника радиотехнической системы позволяет изменять частоту автогенератора в диапазоне принимаемых частот стандарта LTE.

Ключевые слова: СВЧ-автогенератор, комплементарная пара, генератор, управляемый напряжением, варактор, схема Колпитца.

Введение

Радиотракты приемников цифровых систем передачи данных, реализованных с использованием прямого преобразования принятого сигнала супергетеродинной структуры или с цифровой промежуточной частотой, включают тракт синтеза частот. Автогенераторы, выполненные как опорные стабилизированные кварцем генераторы, являются основной частью синтезатора частот, формирующего сетку частот. Они обладают высокой стабильностью генерируемой частоты, и используя управление напряжением, обеспечивают прием сигнала в любой части выделенного рабочего диапазона. Относительная нестабильность частоты формируемых гармонических сигналов на выходе генераторов, управляемых напряжением (ГУН) даже для абонентских терминалов составляет $\delta_{\text{ГУН}} = 10^{-6} \dots 10^{-7}$. Применение термостатирования опорного генератора на базовых станциях систем подвижной связи позволяет повысить стабильность частоты опорных генераторов до значений $\delta_{\text{ОГ}} = 10^{-7} \dots 10^{-8}$. Применением корреляционных методов обработки сигнала требует обеспечения синхронности и синфазности несущей (средней) частоты принятого сигнала и колебания, создаваемого ГУНом. Для этого требуется включить в структуру тракта синтеза частот генератор, частота которого может изменяться под действием регулирующего напряжения, создаваемого в цифровой части приемника [1]. Такие генераторы, управляемые напряжением VCO (Voltage Control Oscillator) должны изменять генерируемую частоту не только между фиксированными значениями сетки частот, но и во всей полосе принимаемых частот, выделенных конкретной системе передачи данных, что определяется планировщиком сети и алгоритмом динамического распределения частот.

Конструктивно генераторы СВЧ диапазона, управляемые напряжением, выполняются по схеме индуктивной или емкостной трех-точки с использованием в качестве активных элементов (АЭ) МОП транзистора с каналом n - типа. В последнем случае (схема Колпица) схемная реализация на единой подложке оказывается технологически более простой и компактной. В качестве общих требований, предъявляемых к ГУН, которые используются в приемниках абонентских терминалов (АТ), является применение общего низковольтного источника питания и возможно более широкий диапазон перестройки. Это особенно важно при построении многофункциональных приемников, способных работать в нескольких стандартах. Электронная перестройка ГУН обеспечивается с помощью варактора или МОП транзистора. При этом без применения дополнительных мер достигается перекрытие 18 -20% рабочего диапазона частот [3] при однопортовом или дифференциальном выходе. Для расширения диапазона перекрываемой полосы, например, для приема сигналов нескольких систем (GSM, DCS, PCS, WCDMA) применяют блок коммутируемых конденсаторов [2].

Одним из способов расширения диапазона регулирования является построение управляемых генераторов, формирующих гармоническое колебание на субгармонике основной частоты (обычно равной половине рабочей частоты). Такая конструкция генератора особенно актуальна в гомодинных приемниках [1], когда на входах смесителя требуется обеспечивать равенство частот входного воздействия и ГУН. Результатом этого является

сильное взаимное влияние источников вследствие просачивания напряжения гетеродина на вход смесителя. Это создает неконтролируемое напряжение смещения на его входе, что создаёт опасность излучения приемником на частоте сигнала и др. Применение преобразователей с ГУНом на субгармонике [4] решает многие из перечисленных проблем.

Применение генераторов, по схеме Колпица, достаточно просто реализуются в области относительно низких частот, но даже в этом случае требуется включение дополнительных цепей, обеспечивающих необходимое напряжение смещения и заданный режим АЭ по постоянному току. Это приводит к снижению показателей генератора при переходе на работу в диапазоне СВЧ из-за влияния паразитных связей в его схеме [6], включая связи внутри АЭ.

Анализ свойств генератора на комплементарной паре МОП транзисторов во временной и частотной области

В качестве альтернативы традиционной однопортовой схеме Колпица на МОП транзисторе с n-каналом в работе предложена базовая схема [5] на комплементарной паре МОП транзисторов с n- и p-каналом (рис. 1), моделирование работы которой приводится далее.

Принципиальная схема автогенератора (рисунок 1б) обладает минимальным числом компонентов: комплементарная пара и колебательный контур, включающий индуктивность ($L1$) и конденсаторы $C1$ и $C2$, которые при работе в СВЧ диапазоне, могут быть заменены конденсаторами C_{gs1} и C_{gs2} , величина емкости которых определяется емкостью между затвором и истоком выбранных транзисторов (BSS84V и BSS83PH) из библиотеки программы MicroCap 12 [7]. Включение в схему дросселя $L2$ позволило снизить напряжение источника питания ($V3 = 3В$) и обеспечить мягкий режим самовозбуждения генератора в диапазоне рабочих частот стандарта LTE. Распределение напряжения и тока генератора показано на рисунке 2. Результаты моделирования генератора для указанных условий во временной области приведены на рисунке 3, а спектр напряжения выходного сигнала на рисунке 4.

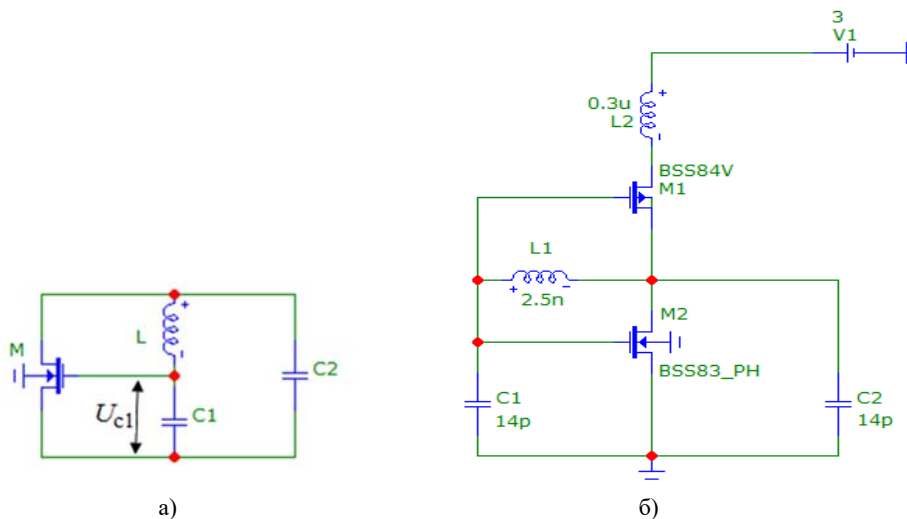


Рис 1. Принципиальная схема генератора Колпица: а) на одном МОП транзисторе с n-каналом; б) на комплементарной паре n-МОП и p-МОП транзисторов

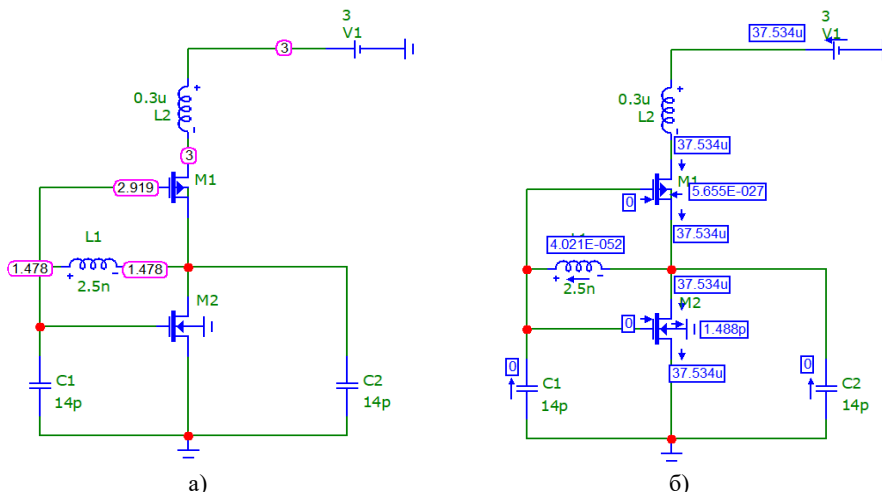


Рис 2. Значения напряжения и тока в узлах и ветвях схемы генератора: а) напряжение в узлах; б) ток в ветвях

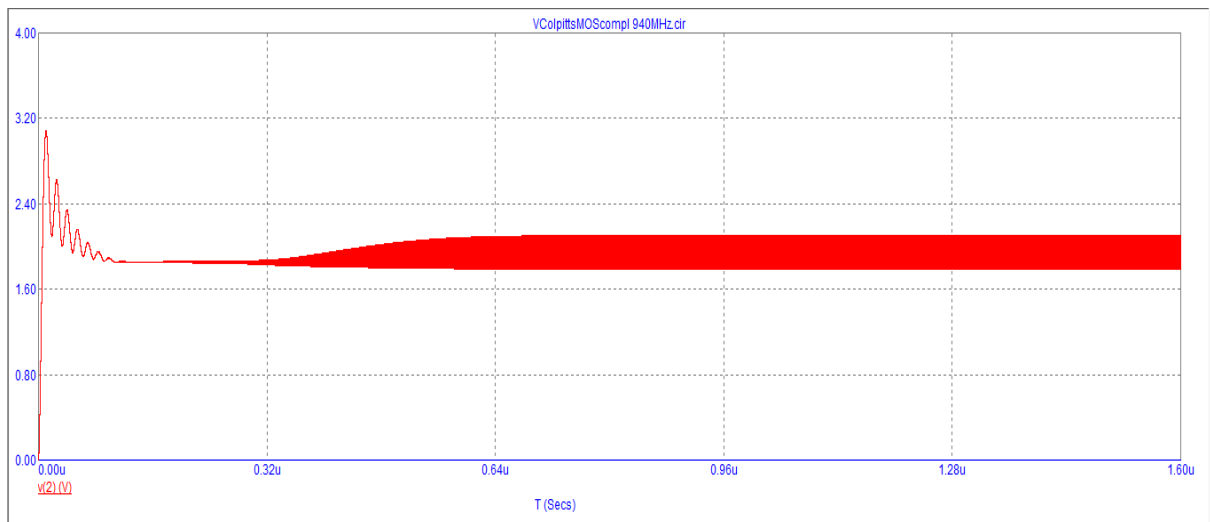


Рис. 3. Сигнал на выходе автогенератора

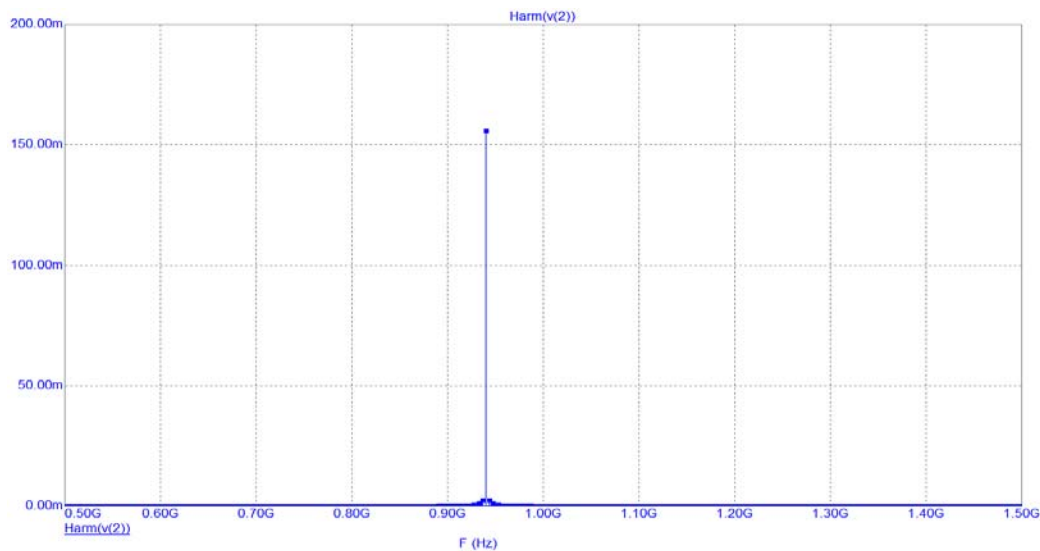


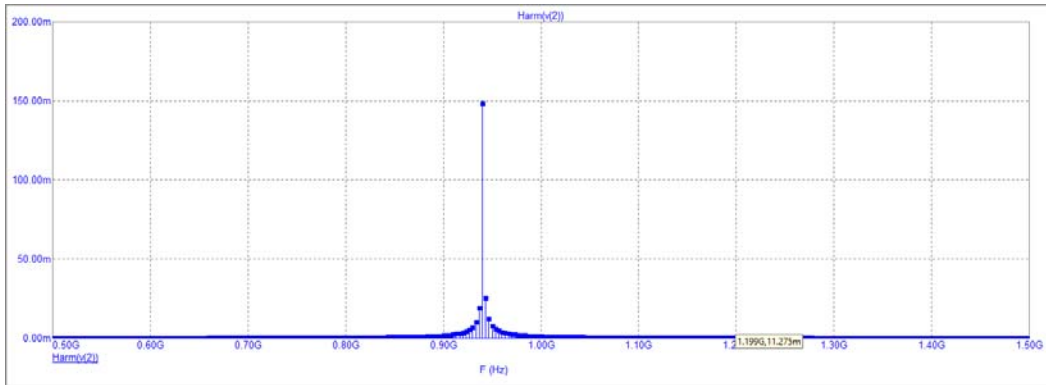
Рис. 4. Спектр выходного сигнала для установившегося режима генератора

Как следует из полученных результатов (рис. 3 и 4), генерируемые колебания обладают достаточно узким спектром и амплитудой первой гармоники $U_{m1} = 160 \text{ mV}$ при низкой фильтрующей способности эквивалентного колебательного контура.

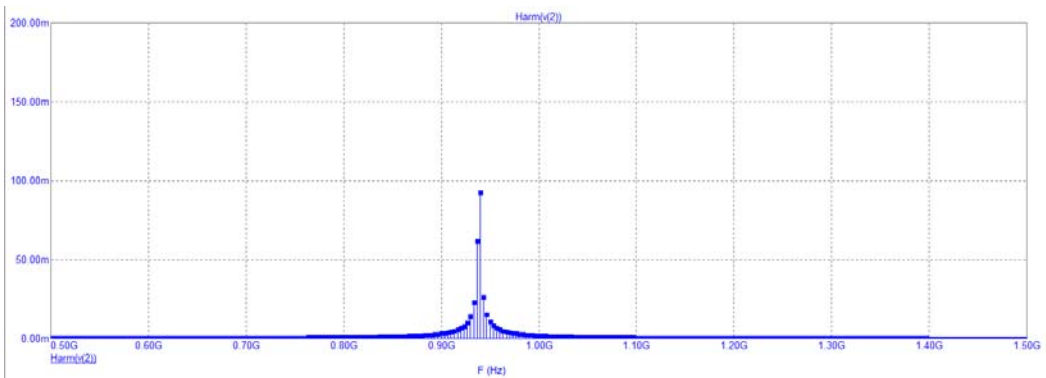
Введение положительной обратной связи в цепь генератора (рис. 1а) изменяет известное свойство резонансного усилителя, нагруженного на параллельный одиночный колебательный контур, что улучшает его добротность, а значит, и фильтрующие свойства. Увеличение индуктивности $L1$ (рис. 1б) в этой цепи (при сохранении частоты генерации) не приводит к увеличению добротности цепи (рис. 5а,б), а лишь уменьшает величину напряжения обратной связи U_{c1} (рис. 1а) и снижает фильтрующую способность контура.

Сравнение значений амплитуды генерируемого напряжения при различных значениях индуктивности цепи показывает, что увеличение индуктивности $L1$ приводит к уменьшению амплитуды генерируемых колебаний. Коэффициент шума остался на фиксированном значении $NF = -196,7 \text{ дБ}$. Это указывает на необходимость выбора глубины обратной связи при настройке схемы для получения необходимой частоты генерации и режима постоянного тока МОП-транзисторов, обеспечивающего мягкий режим самовозбуждения.

Вариации генерируемой частоты при изменении температуры окружающей среды - $20^\circ\text{C} \dots 50^\circ\text{C}$ оценивались при исходном значении $f_{\text{ген}} = 940 \text{ МГц}$ при $T = 23^\circ\text{C}$.



а) $L1 = 1 \text{ нГн}$



б) $L1 = 5,64 \text{ нГн}$

Рис. 5. Спектральные характеристики сигнала генератора при различных значениях индуктивности $L1$ цепи на фиксированной частоте ($f_{ген} = 940 \text{ МГц}$)

Анализ характеристик выходного сигнала от температуры внешней среды указывает на сильную зависимость частоты генерации от температуры среды. Так при изменении температуры от -20°C до 50°C частота генерации уменьшается на 24 МГц (рис. 6).

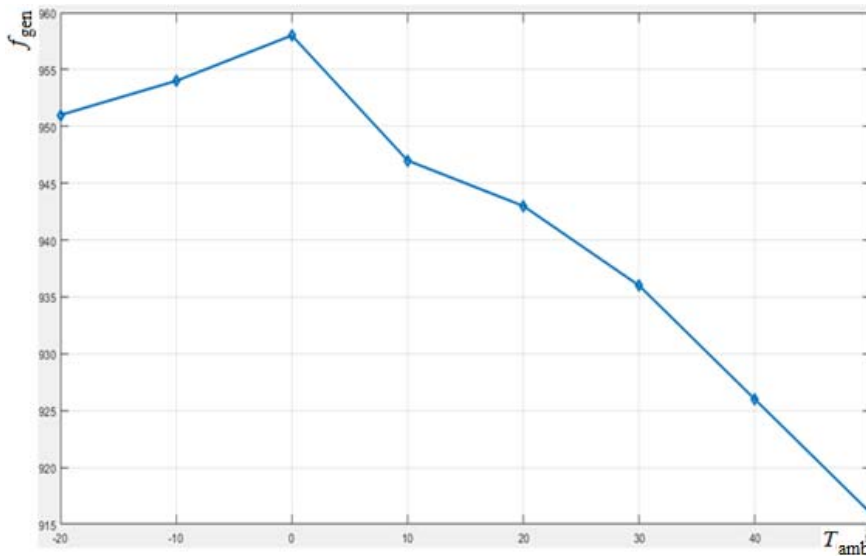


Рис. 6. Зависимость генерируемой частоты от температуры окружающей среды

Как следует из рисунка 7, температура окружающей среды оказывает существенное влияние на коэффициент шума генератора.

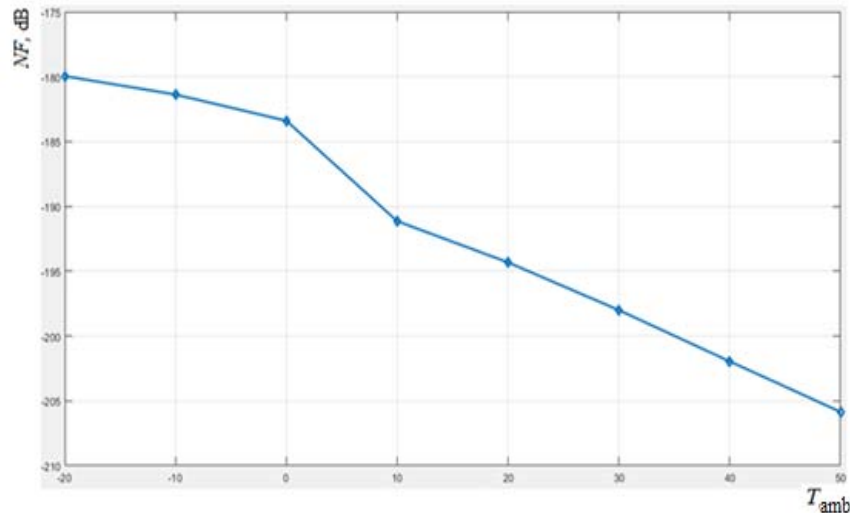


Рис. 7. Зависимость величины коэффициента шума при изменении внешней температуры

Так, при изменении температуры от -20°C до $+50^{\circ}\text{C}$ значение шума уменьшается примерно на 25 дБ. При этом максимальное значение $NF = -180$ дБ (наихудший сценарий), что значительно меньше (примерно на 30 дБ) коэффициента шума резисторного дифференциального каскада на МОП-транзисторах на аналоговых частотах [7].

Управление частотой генерации и шумовые показатели генератора

Потребность в перестройке ГУН обусловлена, прежде всего, требованием обеспечивать работу АТ в полосе частот, выделенной конкретной сети, необходимость обеспечивать роуминг и хэндовер, а также возможностью принимать сигналы других стандартов и частотных диапазонов. В описываемой схеме это достигается подключением варактора (рис. 8), емкость которого изменяется под действием управляющего напряжения, вырабатываемого в цифровой части приемника АТ, как результат обработки сигналов синхронизации, поступающих от БС.

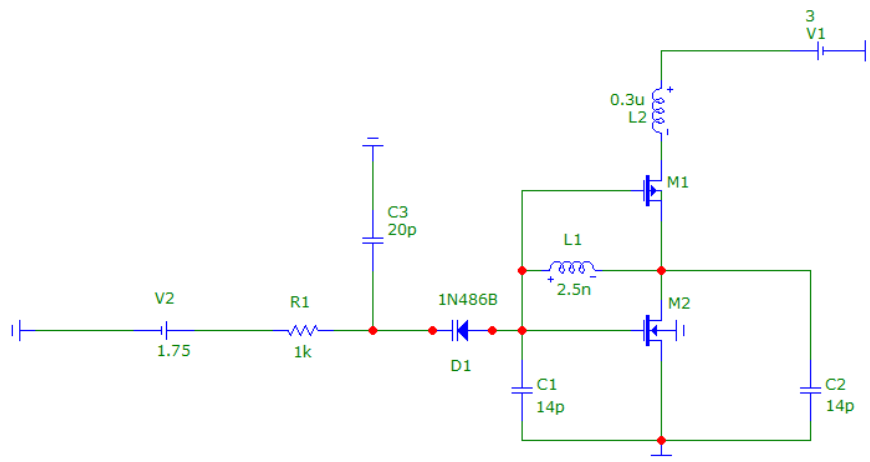


Рис. 8. Принципиальная схема генератора на N-МОП и P-МОП транзисторах с управлением частотой генерации

Моделирование выполнялось для МОП транзисторов с длиной канала $L = 0,35$ мкм и варактора $D1$ (1N4868). Результаты моделирования, приведенные на рисунке 9, показывают, что для достижения изменения частоты генерации в полосе частот, выделенной системе LTE (925...960 МГц), достаточно изменять управляющее напряжение $V2$ на варакторе в пределах 1,6-2,1 В.

В этом интервале значений $u(B)$, частота генерации изменяется по закону, близкому к линейному, что повышает эффективность и точность её регулирования.

Расчеты интегрального коэффициента шума (рисунок 10), обусловленного физическими процессами в транзисторе, амплитудно-фазовой конверсией и дрожанием фазы, не превышают -193 дБ, что существенно ниже коэффициента шума, например, смесителя.

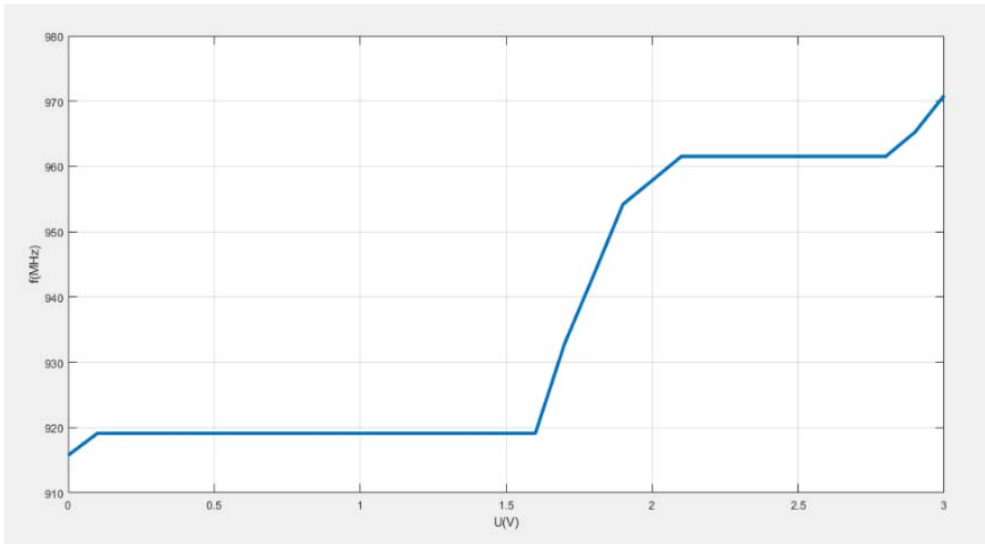


Рис. 9. Генерируемая частота в зависимости от управляющего напряжения на варакторе

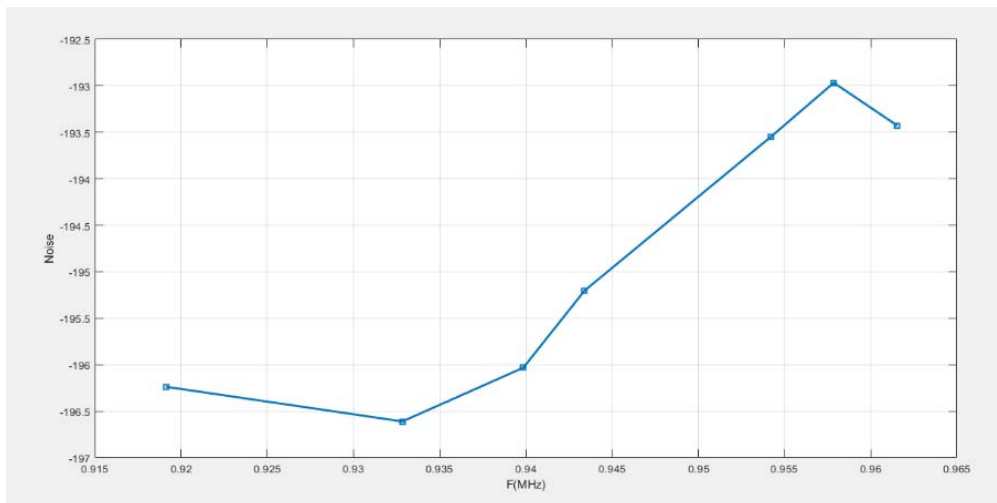
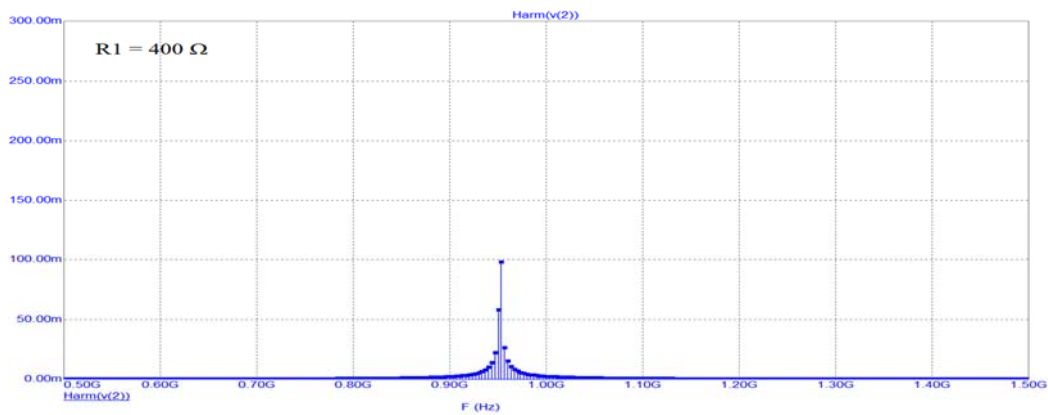


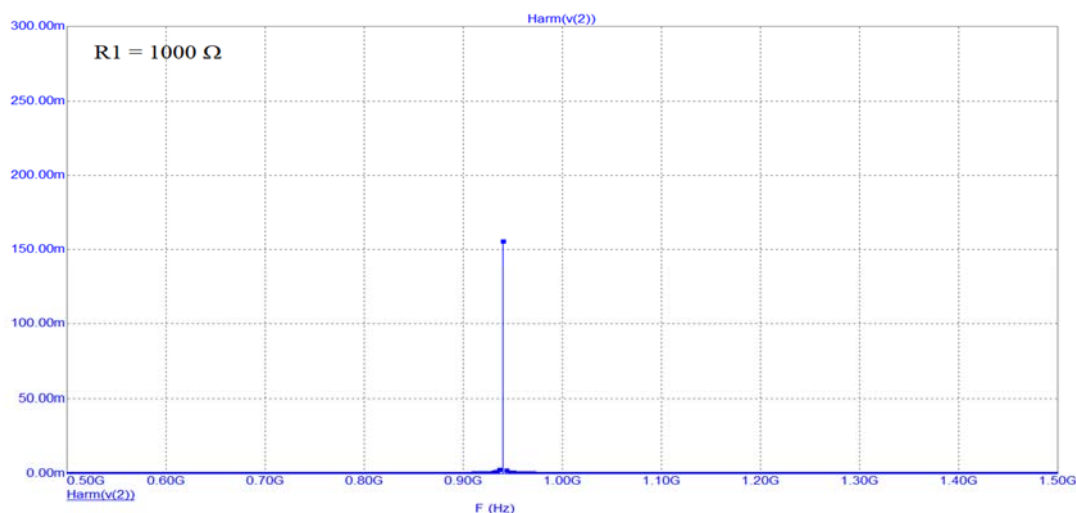
Рис. 10. Коэффициент шума генератора как функция напряжения на варакторе

Фильтр нижних частот R1C3 в схеме на рисунке 8 предназначен для снижения влияние параметров источника управляющего напряжения на варакторе, что обусловлено, в основном, его постоянной времени.

Подключение варактора к колебательной системе оказывает значительное влияние на спектральные характеристики генератора (рис. 11).



а) $R1 = 400 \text{ Ом}$



б) R1 = 1000 Ом

Рис. 11. Спектр выходного сигнала для различных параметров фильтра варактора

Как следует из рисунка 11б увеличение постоянной времени ФНЧ (R1C3) с увеличением сопротивления резистора R1, снижает его шунтирующее влияние на эквивалентный колебательный контур и улучшает фильтрующие свойства.

Выводы

Выполненное моделирование генератора СВЧ на комплементарной паре МОП транзисторов позволяет применять низковольтное напряжение питания при малой потребляемой мощности, что делает предпочтительным его использование в качестве ГУН в АТ мобильных систем передачи данных.

Долговременную нестабильность частоты генерации, обусловленную влиянием температуры окружающей среды достаточно легко устранить корректированием напряжением регулирования на варакторе, формируемым в цифровом тракте.

Отсутствие идентичности МОП транзисторов, обладающих разным типом проводимости канала, может быть устранено использованием инверсной технологии при формировании структуры транзистора [8,9].

Литература

1. Logvinov V.V., Smolskiy S.M. Radio Receivers for Systems of Fixed and Mobile Communications, Springer, 2022.
2. Manetakis K., Jessie D., Narathong C. A CMOS VCO with 48% tuning range for modern broadband systems // Proc. IEEE Custom Integrated Circuits Conf, 2004, pp. 265-268.
3. Berny A.D., Ninejad A.M., Meyer R.G. A 1.8GHz LC VCO with 1.3 GHz tuning range and digital amplitude calibration // IEEE J. of Solid-State Circuits, 2005, vol. 40, №4, pp. 909-917.
4. Cornwel G., Gupta Ch. Investigate Wideband Frequency Converters. Microwaves & RF, April 2016, pp. 90-97.
5. Patent Application Publication Nov. 8, 2007 Sheet 1 of 6 US 2007/0257742 A1.
6. Razavi B. RF Microelectronics. Upper Saddle River, NJ: Prentice-Hall, 1998.
7. <http://www.spectrum-soft.com> (Accessed: 10 April 2022).
8. Bertling K., Racis A.D., Yeou Y.T., O'Brien C.J., Domyo H. SOS Junctionless MOSFETs vs. Inversion Channel MOSFETs: Doubling the Device Speed Without Changing the Technology.
9. Logvinov V.V., Anh P.T. Controlled Microwave Autogenerator on a Complementary Pair of MOSFET Transistors // 2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), 2022, pp. 1-4.

**CONTROLLED MICROWAVE AUTOGENERATOR ON A COMPLEMENTARY PAIR
OF MOSFET TRANSISTORS**

Phan Tuan Anh,
Postgraduate MTUCI, Moscow, Russia,
alexpta96@gmail.com

Vasily V. Logvinov,
Associate Professor of the Department of CER MTUCI,
PhD, Moscow, Russia,
adlerbasil@rambler.ru

Abstract

A microwave oscillator is proposed, implemented according to the Colpitts scheme on a complementary pair of MOS transistors with 0.35 μm technology. This design uses a source with a reduced supply voltage (3 V) and a minimum number of components used. Generation frequency control, providing a fixed value of the converted frequency at the mixer output, which is achieved by including a varactor in the tank. Variation of the capacitance of the varactor under the action of a control voltage from the output of the digital path of the receiver of the radio engineering system allows you to change the frequency of the oscillator controlled by the voltage (VCO) in the operating frequency band of the receiver's input signal.

Keywords: *microwave oscillator, complementary pair, oscillator controlled by the voltage, capacitance of the varactor, Colpitts scheme.*

РАЗРАБОТКА ЛАБОРАТОРНОГО СТЕНДА ДЛЯ НАГЛЯДНОГО ОТОБРАЖЕНИЯ И ИЗУЧЕНИЯ СТОЯЧИХ ВОЛН

Борисов Михаил Денисович,
студент МТУСИ, Москва, Россия,
j0ystake@yandex.ru

Волков Дмитрий Владиславович,
студент МТУСИ, Москва, Россия,
BB.translator@yandex.ru

Яновский Алексей Сергеевич,
заведующий лабораторией кафедры ТиЗВ МТУСИ, Москва, Россия,
JanovskyAleksy@gmail.com

Аннотация

Для формирования и распространения качественного аудиоматериала требуется серьезный подход к его созданию, важнейшим этапом которого является сведение аудиоконтента. Сведение проводится в контрольных комнатах – помещениях с идеальной акустикой. При проектировании данных помещений важно исключить все факторы появления искажений сигнала. Стоячие волны – одна из наиболее важных причин их возникновения. Статья посвящена разработке трубы Кундта – прибора для отображения и изучения стоячих волн в пространстве. В рамках данной работы был создан стенд, который будет использоваться в образовательном процессе для освоения студентами вуза лабораторного курса по основам акустики.

Ключевые слова: акустика, стоячие волны, резонанс, лабораторная установка, Труба Кундта, преобразование Фурье, длина волны.

Введение

Как увидеть невидимое? Этим вопросом задался немецкий ученый Август Кундт и в 1866 году изобрел прибор, позволяющий пронаблюдать явление акустической стоячей волны в цилиндре. Для наглядного представления волны Кундт использовал тальк, пробковую пыль, или ликоподий (сухие споры растений). В качестве генератора сигнала был использован металлический стержень, при натирании тканью которого, тот издавал сигнал постоянной частоты. С помощью данного изобретения Кундт не только проанализировал поведение стоячей волны, но и смог измерить скорость звука в воздухе и в водороде.

Аналогом трубы Кундта является труба Рубенса. Принцип работы данных приборов очень схож, однако Рубенс использовал перфорированную металлическую трубу, заполненную горючим газом. После подачи в трубу сигнала, газ поджигался. Высота столпов пламени, исходящих из отверстий трубы, менялась в зависимости от давления, тем самым визуализируя стоячую волну [8].

В обычном восприятии людей стоячие звуковые волны ассоциируются с тем, от чего требуется избавляться. Это не удивительно, так как стоячие волны и места их сосредоточения (моды) создают неравномерности распределения звука в помещении, в котором возникает "гул" на резонансных частотах до тех пор, пока не произойдет полное затухание мод. Это мешает прослушивать и наслаждаться музыкальными произведениями, а главное – профессионально заниматься процедурой сведения (созданием единого звучания элементов микса) и мастерингом (подстройкой звучания под различные площадки и условия прослушивания) [5,7].

Для борьбы со стоячими волнами нужно в первую очередь правильно оборудовать помещение: использовать необходимые материалы для отделки поверхностей, соблюдать оптимальные соотношения сторон и избегать прямых углов [8]. Самое благоприятное помещение для стоячих волн – это равносторонняя комната, так как в модах будут образовываться волны сразу нескольких порядков. Правильное расположение акустических систем в помещении также играет существенную роль в создании качественного звучания.

Для борьбы со стоячими волнами необходимы акустические диффузоры. Они позволяют рассеивать звуковую волну, которая попадет в диффузор, и тем самым предотвращают амплитудные волновые искажения. Образец акустического диффузора представлен на рисунке 1.



Рис. 1. Акустический диффузор

Основы теории сведения звуковых треков

Сведение является важнейшим этапом в подготовке и записи музыкальных программ [6]. Оно обеспечивает баланс звучания всех инструментов, объем и глубину звуковой картины, устранение интонационные огрех и тд. Возникающая при звучании в помещении стоячая волна представляет собой периодическое во времени синфазное колебание с характерным распределением амплитуды – чередованием узлов (нулей) и пучностей (максимумов) [1-3]. Для наглядного примера возьмем синусоидальную стоячую волну, которая представлена аналитическим выражением:

$$F(x) = A \cos(kx - \varphi) \tag{1}$$

где A , k , φ – постоянные.

В точках, где $F(x) = 0$ – амплитуда звуковых колебаний наименьшая и в них образуются узлы стоячей волны. В точках, где функция $F(x)$ имеет максимум амплитуда колебания наибольшая, и эти точки называются пучностями. Данное явление изображено на рисунке 2.

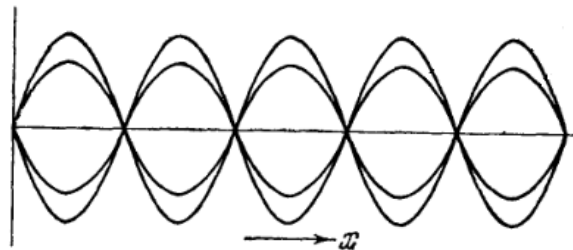


Рис. 2. Синусоидальная стоячая волна

В случае трубы Кундта стоячая волна визуализируется с помощью материала, которым она заполнена. Это показано на рисунке 3, где А – это стоячая волна, а Б – распределение материала заполнения трубы относительно узлов и пучностей.

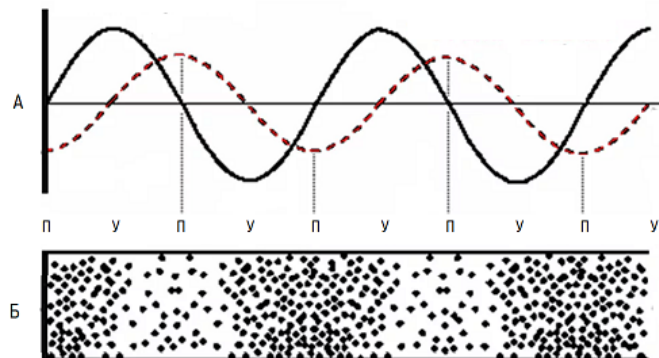


Рис. 3. Распределение материала в трубе относительно узлов и пучностей звуковой волны

Стоячая волна, отражаясь от поверхности трубы, порождает явление резонанса. Под резонансом понимается совпадение частоты внешнего воздействия с собственной резонансной частотой объекта. Частот резонанса может быть несколько, расчёт их производится по формуле:

$$f_p = \frac{c}{2l} \quad (2)$$

где: c – скорость звука; l – длина трубы [3].

Зная частоту резонанса, можно найти длину волны по формуле (3).

$$\lambda = \frac{c}{f} \quad (3)$$

Длина стоячей волны является половиной этого значения.

Скорость звука в разных средах различна. Она зависит от упругости и плотности среды распространения, а также от температуры и влажности. С увеличением температуры среды скорость увеличивается, а с уменьшением – убывает. Например, скорость звука в воздухе, при температуре 20°C, равна 343 м/с, а в водороде, при 0°C – 1284 м/с.

График зависимости скорости звука от температуры и давления воздуха представлен на рисунке 4. [3]

Для анализа спектра звукового сигнала путём разложения его на гармонические составляющие используется дискретное преобразование Фурье (ДПФ) (4).

$$S_x(n) = \sum_{k=0}^{N-1} s(k) \left[\cos\left(\frac{2\pi}{N}nk\right) - j\sin\left(\frac{2\pi}{N}nk\right) \right], n = 0, \dots, N-1 \quad (4)$$

ДПФ обеспечивает представление сигнала набором комплексных синусоид, неизменных на времени анализа и соотносимых со спектральными составляющими.

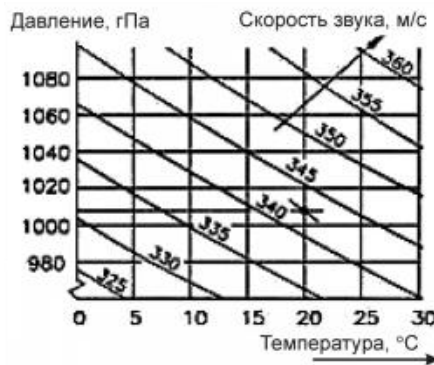


Рис. 4. График зависимости скорости звука от температуры и давления воздуха

Спектральное разрешение ДПФ, т. е. полоса, в пределах которой оценивается энергия соответствующим коэффициентом, рассчитывается по формуле (5),

$$df = bF_d/N \quad (5)$$

где df – спектральное разрешение; F_d – частота дискретизации; b – коэффициент, характеризующий увеличение ширины полосы оценки в зависимости от типа окна.

Шаг сетки частот ДПФ принято называть бином; при b равном единице (прямоугольное окно), бином совпадает с df .

Дискретное преобразование Фурье содержит алгоритм быстрого вычисления – быстрое преобразование Фурье (БПФ), с помощью которого вычисляется большая часть других ортогональных преобразований [4].

Пример спектра сигнала на одной из частот резонанса лабораторного стенда для изучения стоячих волн приведён на рисунке 5.

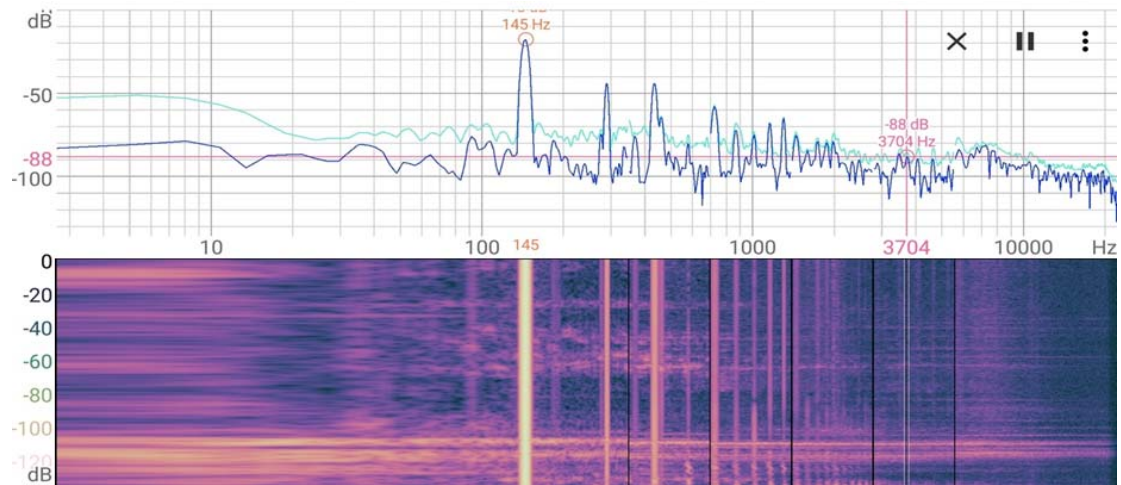


Рис. 5. Спектр сигнала на частоте 145 Гц

Структурная схема лабораторного стенда

Структурная схема лабораторного стенда представлена на рисунке 6.

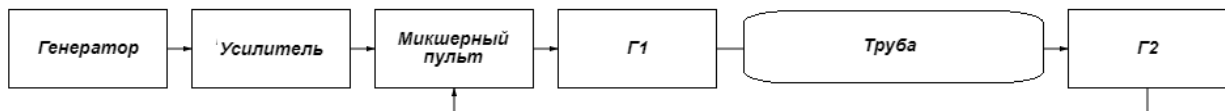


Рис. 6. Структурная схема стенда

Лабораторный стенд состоит из следующих элементов: к трубе из закаленного стекла подведены источники звука – громкоговорители. Они подключены к генератору синусоидального сигнала стабильной частоты с регулятором. Следующим элементом схемы является усилитель. Его нужно использовать в том случае, если генератору не хватает мощности сгенерировать сигнал нужного уровня. Требуемый уровень сигнала зависит от характеристик используемой трубы: толщины, длины и материала. Также в схеме присутствует микшерный пульт, с его помощью, возможно, регулировать баланс двух громкоговорителей ($G1$ и $G2$), что требуется для достижения наилучшей визуализации.

Принцип работы лабораторного стенда

Принцип работы стенда состоит в следующем: с генератора на громкоговоритель через усилитель и микшерский пульт подается синусоидальный сигнал. При изменении частоты сигнала до частоты резонанса трубы в ней возникает стоячая волна, которая создает явление резонанса за счёт отражения звуковой волны от внутренней поверхности прозрачной трубы. Настроив сигнал с генератора на частоту резонанса, можно получить визуализацию точек максимума и минимума звуковой волны за счёт движения сыпучего материала, которым заполнена труба. При этом в местах пучностей звуковой волны наглядно видно скопление сыпучего материала, которым заполнена прозрачная труба, а в узлах – его уменьшение (фото на рисунке 7).



Рис. 7. Фото трубы с сыпучим материалом

Выбор приборного состава лабораторного стенда

Первым элементом лабораторной установки является генератор постоянного сигнала Г6-51 (рис. 8). В генераторе имеется возможность формирования на выходе сигналов треугольной, прямоугольной и синусоидальной форм. Для эксперимента используется синусоидальный сигнал. С помощью ручек грубой и плавной настройки, выбрав нужный диапазон, можно добиться точной, вплоть до десятой доли герца, настройки на частоту резонанса трубы. Уровень сигнала может регулироваться от -40 дБ до 0 дБ. В генераторе имеется цифровой дисплей, что значительно упрощает индикацию результатов его настройки.



Рис. 8. Генератор постоянного синусоидального сигнала

С выхода генератора сигнал поступает на усилитель, который в свою очередь подключен к микшерному пульта фирмы "Mackie 1402 VLZ" [6]. С помощью усилителя регулируется напряжение, подаваемое на громкоговорители. Максимальное напряжение составляет 12,5 В. Микшерный пульт, представленный на рисунке 9, необходим для регулировки баланса громкоговорителей.



Рис. 9. Микшерный пульт

К микшерному пульта подключены два громкоговорителя, которые плотно прилегают к стеклянной трубе. При разработке стенда было протестировано более десятка различных конфигураций источников воспроизводимого сигнала. После тщательного анализа выбор пал на коаксиальные громкоговорители фирмы "Урал", которые представлены на рисунке 10. Особенность их конструкции заключается в том, что содержащиеся в них динамики высоких и низких частот расположены на одной оси и, следовательно, излучаемый ими сигнал исходит из одной точки пространства. Технические характеристики данной модели динамиков: максимальная мощность – 170 Вт, сопротивление – 4 Ом, чувствительность – 90 дБ, собственная частота – 56 Гц.

Главным конструктивным компонентом стенда является труба, выполненная из закалённого стекла. Длина трубы составляет 1500 мм, диаметр – 80 мм, толщина стекла – 8 мм. Данные параметры выбраны исходя из того, что длина трубы определяет её минимальную резонансную частоту, а диаметр – максимальную.



Рис. 10. Коаксиальные громкоговорители

Особенности сборки стенда

Первым этапом сборки является подготовка трубы. Концы трубы обрабатываются армированной лентой, что очень важно для повышения её прочности в местах соприкосновения с каркасом. Данный этап очень важен, так как труба должна иметь повышенную прочность в местах соприкосновения с каркасом. Поверх армированной ленты необходимо наложить звукоизоляционную ленту. Вторым этапом является сборка каркаса лабораторного стенда. Каркас состоит из деревянных несущих конструкций с алюминиевыми вставками. Стойки для трубы укреплены ребрами жесткости для лучшей устойчивости конструкции. Труба помещается между двумя стойками и заполняется материалом на выбор.

Был проведен ряд экспериментов с выбором материала для визуализации стоячей волны: пробковая пыль, тальк, пенополистирол и деревянные опилки. В ходе сравнения данных материалов было принято решение использовать пенополистирольные частицы. Их размер также влияет на качество визуализации: более крупные частицы менее подвержены электризации и слипанию, чем мелкие. После выбора материала для визуализации, труба обрабатывается антистатическим веществом. Это необходимо для предотвращения налипания частиц пенополистирола на стенки стеклянной трубы из-за их электризации. Далее следует этап герметизации всей конструкции, после которого можно крепить акустические громкоговорители. После подключения всех элементов схемы стенд готов к работе. Фото смонтированного лабораторного стенда для изучения звуковых стоячих волн приведено на рисунке 11.



Рис. 11. Лабораторный стенд

Методика выполнения лабораторной работы

В ходе выполнения лабораторной работы перед студентами будет поставлена следующая задача: настроить трубу на частоту резонанса и измерить параметры возникшей в ней стоячей звуковой волны.

Рассмотрим порядок выполнения лабораторной работы учащимися:

1. Включить лабораторный стенд;

2. С помощью микшерного пульта настроить баланс сигнала между двумя громкоговорителями;
3. Настроить воспроизводимый сигнал на частоту резонанса трубы;
4. Регулятором усиления установить уровень сигнала, достигнув корректной визуализации;
5. Измерить расстояние между узлами и посчитать длину волны;
6. Повторить пункты 3-5 для всех частот резонанса;
7. Полученные данные занести в таблицу и составить отчет.

В ходе испытаний лабораторного стенда было выявлено, что частоты резонанса трубы составляют: 115, 268, 445, 507, 765. Гц, следовательно, учащимся необходимо провести пять измерений.

Заключение

Лабораторные работы являются неотъемлемой частью учебного процесса. В ходе их выполнения студенты активно погружаются в изучаемую тему. Однако, степень погружения в тему лабораторного практикума определяется не только их заинтересованностью в результатах обучения, но и возможностями лабораторного оборудования. При разработке лабораторного стенда авторы руководствовались такими его параметрами как внешний вид, надежность конструкции, функциональность и наглядность получаемых результатов эксперимента. В ходе работы на стенде учащиеся получают базовые знания в области акустики, в том числе о формировании, распространяются и характеристиках стоящих звуковых волн.

Литература

1. Горелик Г.С. Колебания и волны. Введение в акустику, радиофизику и оптику. 2008. 3-е издание. С. 157-159.
2. Стрелков С.П. Механика: учебник. 2005. 4-е издание С. 481-488.
3. Ефимов А.П., Никонов А.В., Сапожков М.А., Шоров В.И. Акустика. Справочник. Радио и связь. 2-е издание. 1989. С. 17-18.
4. Попов О.Б., Рихтер С.Г., Терехов А.Н. Командирование сигналов в канале звукового вещания. Учебное пособие для вузов. Под ред. профессора С. Г. Рихтера. М.: Горячая линия – Телеком, 2021. С. 113-114.
5. Ньюэлл Ф. Звукозапись: акустика помещений. Пер. с англ. А. Кравченко; Под ред. А. Кравченко; Предисл. А. Кравченко. М.: 2004. С. 18-19.
6. Акимочев Е.А., Орлов В.Г. Принципы построения и функционирования аппаратно-вещательного комплекса радиостанции "Радио 1" // Фундаментальные проблемы радиоэлектронного приборостроения. 2018. Т. 18. № 4. С. 1053-1056.
7. Стоячие волны: что это и как с ними бороться. 2022. [Электронный ресурс]. Режим доступа: <https://pop-music.ru/articles/stoyachie-volny-cto-eto-i-kak-s-nimi-borotsya/> (Дата обращения: 07.11.2022).
8. The Ruben's Tube: Soundwaves in Fire! [Электронный ресурс]. Режим доступа: <https://www.instructables.com/The-Rubens--Tube:-Soundwaves-in-Fire!/>. (Дата обращения: 07.11.2022).

CREATION OF A LABORATORY STAND FOR VISUAL REPRESENTATION AND RESEARCH OF STANDING WAVES

Mikhail D. Borisov,
Student MTUCI, Moscow, Russia,
j0ystake@yandex.ru

Dmitriy V. Volkov,
Student MTUCI, Moscow, Russia,
BB.translator@yandex.ru

Aleksey S. Janovsky,
Head of the Laboratory of Department of T&ZV MTUCI, Moscow, Russia,
JanovskyAleksey@gmail.com

Annotation

The distribution of high-quality audio material requires serious approach to its creation, the most important part of which is mixing. Mixing is carried out in control rooms – rooms with perfect acoustics. When designing such rooms, it is important to exclude all factors, which may lead to signal being distorted. Standing waves are one of the most important causes of distortions. The article is dedicated to the creation of the Kundt's tube – device for displaying and studying standing waves in space. A stand was created for the article, which will be used in the future to teach students the basics of acoustics.

Keywords: acoustics, standing waves, resonance, laboratory setup, Kundt's tube, Fourier transform, wavelength.

ЦВЕТОВОЕ КОНТРАСТИРОВАНИЕ В РЕНТГЕНОВСКОМ КОНТРОЛЕ

Иванчев Василий Витальевич,
аспирант МТУСИ, Москва, Россия,
demidra18@gmail.com

Аннотация

В статье приведён анализ современного состояния рентгенографических систем с определением ключевых параметров изображений, получаемых с помощью этих систем. Также проведен анализ цветowych пространств и методов цветового контрастирования, и обоснован вывод о необходимости создания нового метода цветового контрастирования, учитывающего теплоточность цветowych оттенков.

Ключевые слова: рентгенография, цветowych пространства, обработка динамического диапазона, цветочное контрастирование.

Введение

В настоящее время рентгеновские снимки применяются для решения различных задач во множестве областей промышленности, от контроля качества производства частей авиационной техники [1] до электроники [2] и контроля сварных соединений [3]. Рентгенография позволяет увидеть скрытые дефекты выпускаемой продукции, такие как: полости, трещины и иные повреждения внутри конструкций. Однако, рентгеновские снимки в современных системах промышленной рентгенографии обладают высоким динамическим диапазоном, поэтому для восприятия таких снимков оператором, изображения, полученные с рентгенографических систем, необходимо предварительно обработать. Одним из методов такой обработки, позволяющей оператору воспринять рентгенографическое изображение без потери информативности, является цветочное контрастирование.

Современное состояние рентгенографии

Компьютерная томография применяется для решения различных производственных и технологических задач во многих отраслях промышленности, а также широко используется в медицине и здравоохранении. Свое распространение компьютерная томография получила благодаря возможности создания рентгеновского трехмерного изображения объекта с возможностью послойного анализа.

Рентгеновский снимок получается путем детектирования гамма-лучей, сгенерированных в рентгеновской трубке. Разогнанные электроны бомбардируют мишень в трубке, что порождает рентгеновское излучение с тормозным спектром электронов и характеристическим спектром атомов мишени (рисунок 1) [4]. На графике изображен рентгеновский спектр вольфрамового анода при ускоряющих напряжениях в диапазоне $U_a = 80-140\text{kV}$. K_α и K_β – всплески на графике спектральных составляющих интенсивности излучения при переходе электрона из L- и M-орбиты в K-орбиту.

Системы рентгенографии для различных отраслей производства существенно различаются по параметрам. Например, по динамическому диапазону, что обуславливается различным коэффициентом затухания интенсивности излучения для различных материалов [4]:

$$\mu = \frac{\rho N_A}{A} \sigma_{tot} = n \sigma_{tot} \quad (1)$$

где ρ – плотность, N_A – число Авогадро, A – атомный вес материала, n – число атомов в мишени, σ_{tot} – полное поперечное сечение для рассеяния или поглощения.

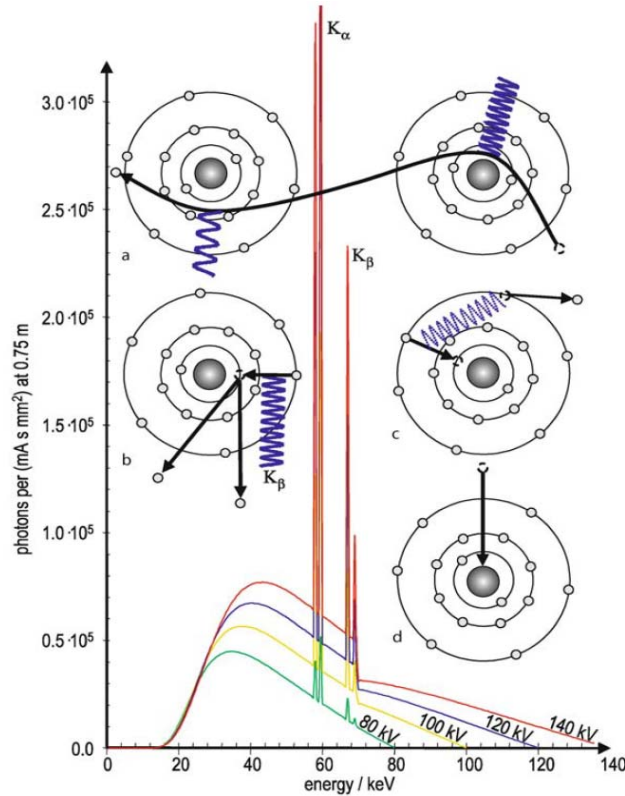


Рис. 1. Процесс получения рентгеновского снимка:

a – тормозное излучение, b – характеристическое излучение, c – Оже-процес, d – столкновение электронов с ядрами)

Существует несколько видов рассеяния, зависящих от мощности рентгеновского излучения и происходящих при этом процессах [4]:

1. Релеевское или Томпсоновское рассеяние.

Полное поперечное сечение для Томпсоновского рассеяния:

$$\sigma_{Thomson} = \frac{8\pi r_e^2}{3} * \frac{\omega^4}{(\omega^2 - \omega_0^2)} \quad (2)$$

$$r_e = \frac{1}{4\pi\epsilon_0} * \frac{e^2}{m_e c^2} \quad (3)$$

где ϵ_0 – диэлектрическая проницаемость свободного пространства, ω - частота рентгеновского излучения, ω_0 – собственная частота электронов связанного атома, m_e – масса электрона, c – скорость света, e – электрический заряд, r_e – классический радиус электрона.

2. Фотоэлектрическое рассеяние.

Линейный коэффициент фотоэлектрического рассеяния α [4]:

$$\alpha = k \frac{\rho}{A} * \frac{Z^4}{(h\nu)^3} \quad (4)$$

где k – константа, которая зависит от орбиты электрона, h – постоянная Планка, ν – частота излучения, Z – атомный номер вещества.

3. Комптоновское рассеяние

Полное поперечное сечение для Комптоновского рассеяния [4]:

$$\sigma_{Compton} = 2\pi r_e^2 \left(\left(\frac{1+\varepsilon}{\varepsilon^2} \right) + \left(2 \frac{(1+\varepsilon)}{1+2\varepsilon^2} - \frac{\ln(1+2\varepsilon)}{\varepsilon} \right) + \frac{\ln(1+2\varepsilon)}{2\varepsilon} - \frac{1+3\varepsilon}{(1+2\varepsilon)^2} \right) \quad (5)$$

где ε – приведенная энергия падающего фотона [4]:

$$\varepsilon = \frac{h\nu}{(m_e c^2)} \quad (6)$$

4. Образование пары.

Полное поперечное сечение при образовании пары [4]:

$$\sigma_{pair\ production} = \alpha r_e^2 Z^2 \left(\frac{28}{9} \ln(2\varepsilon) - \frac{218}{27} + \frac{6.45}{\varepsilon} \right) \quad (7)$$

где α – постоянная тонкой структуры.

Некоторые виды рассеивания могут происходить одновременно. Также, исследуемый образец может состоять из разных материалов, что создает сразу несколько коэффициентов затухания. В таких случаях, коэффициенты затухания суммируются, и итоговая интенсивность вычисляется по закону [4]:

$$I(\eta) = I_0 * e^{-\mu^* \eta} \quad (8)$$

где $I(\eta)$ – интенсивность рентгеновского излучения после прохождения через исследуемый образец, I_0 – начальная интенсивность рентгеновского излучения, η – расстояние, пройденное через образец.

Примеры графиков коэффициентов затухания в зависимости от энергии падающих на материал фотонов приведены на рисунках 2-4. Графики представлены для отдельных материалов, часто использующихся в электронной промышленности: оксида кремния, кремния и припоя.

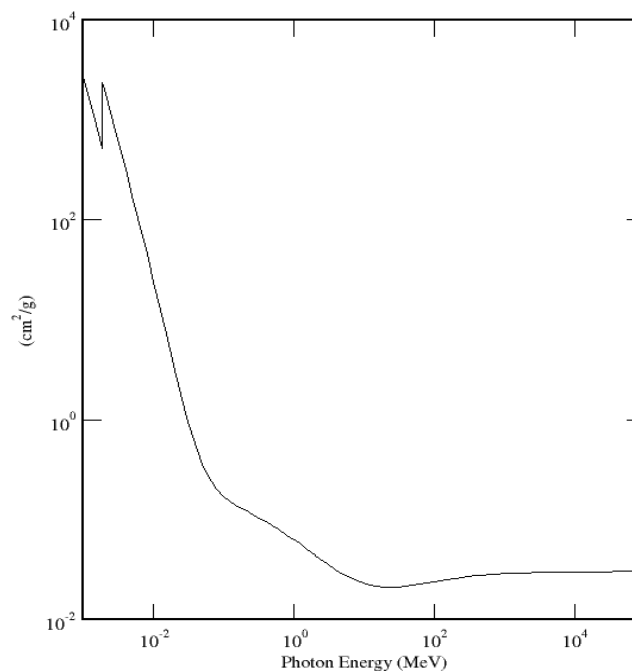


Рис. 2. Суммарный коэффициент затухания для оксида кремния

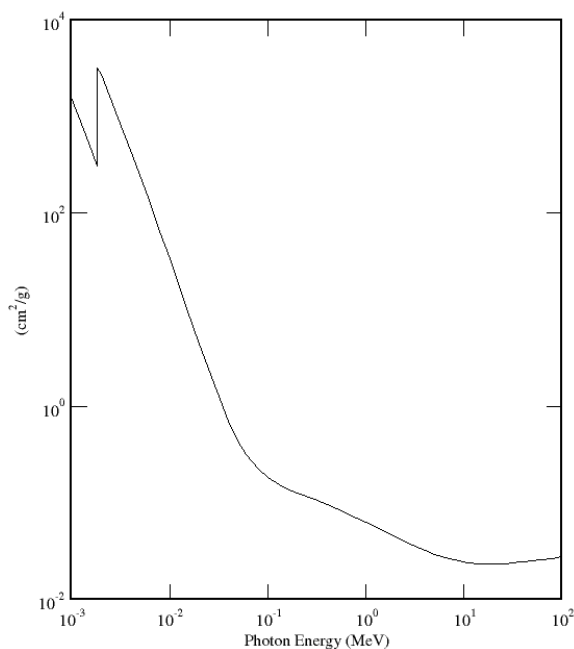


Рис. 3. Суммарный коэффициент затухания для кремния

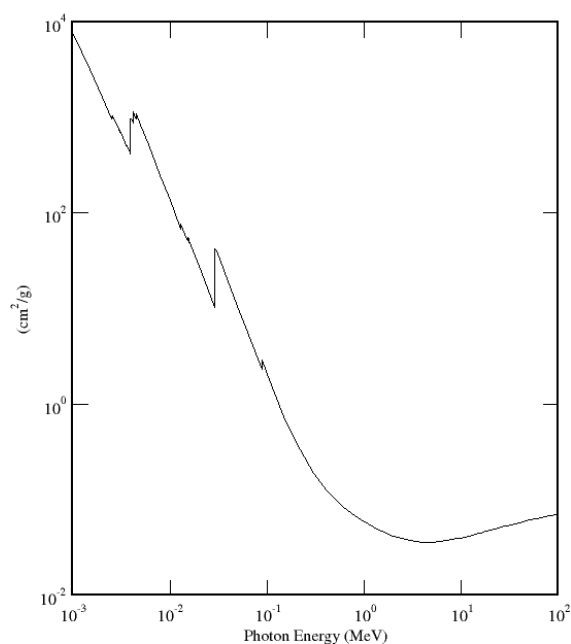


Рис. 4. Суммарный коэффициент затухания для припоя ПОС-90

На сегодняшний момент различимость деталей у современных плоскочувствительных цифровых детекторов составляет порядка 200 нм. Динамический диапазон около 10000 к отношению уровня шума [5-7]. Однако, человек может различать около 200 градаций яркости [8]. Таким образом, оператору для контроля образца через необработанное изображение необходимо выбрать требуемый диапазон градаций. С помощью обработки рентгеновских изображений можно согласовать динамический диапазон рентгеновского изображения с возможностями человеческого зрения. Одним из способов такого согласования является цветовое контрастирование [9].

Изменение ЧД в устройствах ЦАП и АЦП

Для цветового контрастирования необходимо выбрать подходящее цветовое пространство. В настоящее время существует множество цветовых пространств, которые предназначены для разных целей. В телевидении используются трехцветные цветовые пространства по типу СIE 1931. Однако у этого цветового пространства есть фактор, существенно влияющий на решении поставленной задачи. Этим фактором являются эллипсы МакАдама [10].

Эллипсы МакАдама – это области на цветовом пространстве, внутри которых цветовые характеристики не меняются. В условиях поставленных целей амплитудного цветового контрастирования это означает, что нескольким градациям яркости будет соответствовать один и тот же цветовой оттенок [11]. В итоге, информационная составляющая обработанного изображения снизится.

Для решения этой проблемы необходимо выбрать цветовое пространство, в котором эллипсы МакАдама либо отсутствуют, либо имеют минимальную площадь, то есть отвечают условию равноконтрастности. Для этого необходимо рассмотреть существующие цветовые пространства.

Модель RGB

Цвета получаются в результате смешения красного, зеленого и синего цветов (рис. 5). Сама модель выглядит в виде куба, в вершинах которого лежат базовые цвета. Сами же вершины имеют координаты, равные единице.



Рис. 5. Цветовое пространство RGB

Модель HIS

Модель HSI предназначена для удобной работы с ней человеком. Тут цветовой тон (H) задается углом относительно оси красного цвета, насыщенность (S) – расстоянием от центра, интенсивность (I) – расстоянием от черного цвета (рис. 6).

Формулы преобразования от RGB в HSI [10]:

$$\left. \begin{aligned}
 H &= \begin{cases} \theta, B \leq G \\ 360 - \theta, B > G \end{cases}, \theta = \arccos \left(\frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right) \\
 S &= \frac{3}{(R+G+B)} \min(R, G, B) \\
 I &= \frac{1}{3}(R+G+B)
 \end{aligned} \right\} \quad (9)$$

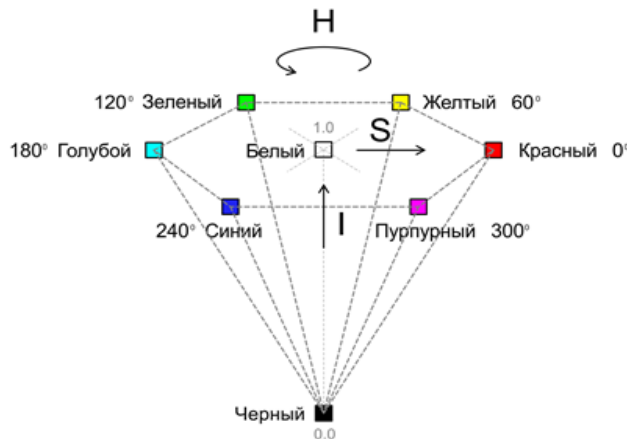


Рис. 6. Цветовое пространство HSI

Модель МКО-1931

Эта модель основана на стандартных кривых сложения $X(\lambda), Y(\lambda), Z(\lambda)$. График восприятия основных цветов приведён на рисунке 7, [8]. Полученные на основе кривых сложения координаты называют стандартными цветовыми координатами.

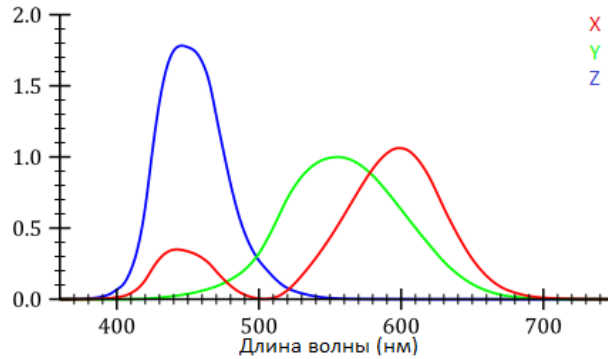


Рис. 7. Стандартные кривые сложения

При переводе стандартных цветовых координат в относительные цветовые координаты [10] можно получить цветовой треугольник МКО (рис. 8).

$$\begin{cases} x = \frac{X}{X+Y+Z} \\ y = \frac{Y}{X+Y+Z} \\ z = \frac{Z}{X+Y+Z} \end{cases} \quad (10)$$

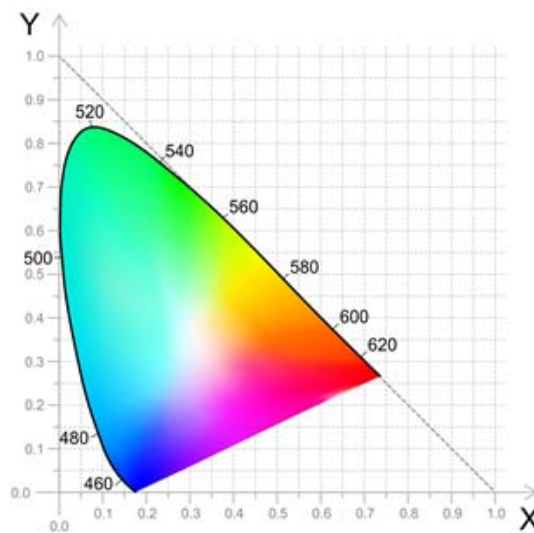


Рис. 8. Срез цветового пространства CIE 1931

Модель CIE 1931 широко используется в телевизионных системах. Однако, у этой модели есть недостаток в виде эллипсов МакАдама. Это области в цветовом пространстве (рис. 9), в которых цвет не меняется [10]. То есть, модель CIE XYZ обладает нелинейностью изменения цвета по координатам. На рисунке 9 эллипсы МакАдама для наглядности увеличены.

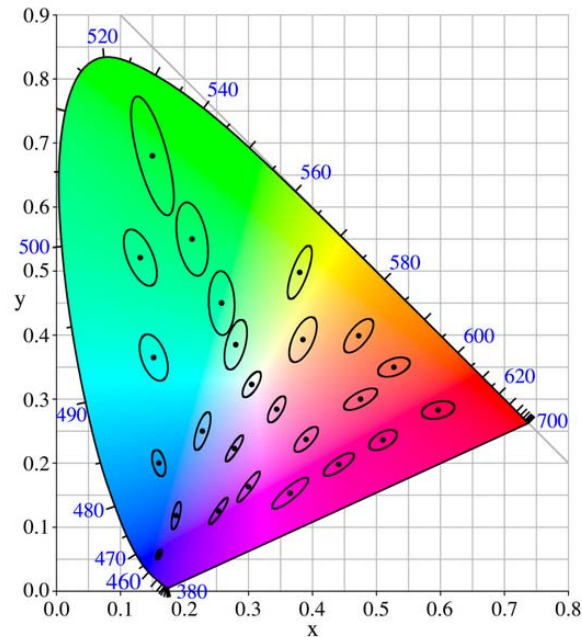


Рис. 9. Срез цветового пространства CIE 1931 с эллипсами МакАдама

Модель CIE LAB

Данное цветовое пространство было разработано в целях устранения нелинейности модели CIE XYZ. L означает светлоту (от 0 до 100), A – позиция между зеленым и пурпурным цветами, B – позиция между синим и желтым цветами.

Перевод из CIE XYZ в CIE LAB осуществляется по следующим формулам [12]:

$$\begin{cases} L = 166f\left(\frac{Y}{Y_n}\right) - 16 \\ A = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \\ B = 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \end{cases} \quad (11)$$

где (X_n, Y_n, Z_n) – координаты точки белого в пространстве CIE XYZ, а

$$f(x) = \begin{cases} \sqrt[3]{x} \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 x + \frac{4}{29} \end{cases} \quad (12)$$

Эллипсы МакАдама при этом преобразуются в окружности одинакового размера (рис. 10), [12].

Таким образом, последнее из приведенных на рисунке 10 цветовых пространств отвечает поставленным условиям, поэтому целесообразно для дальнейшей разработки алгоритма выбрать именно его.

Недостатком CIE LAB является сложность в работе с координатами отдельного оттенка. Это играет важную роль, потому что, как будет видно далее, для оценки теплоты/холодности цветов и оттенков необходимо сформировать выборку из точек. Однако, пространство CIE LAB, как и его срезы, имеет сложную форму (рис. 11), что затрудняет формирование выборки цветов.

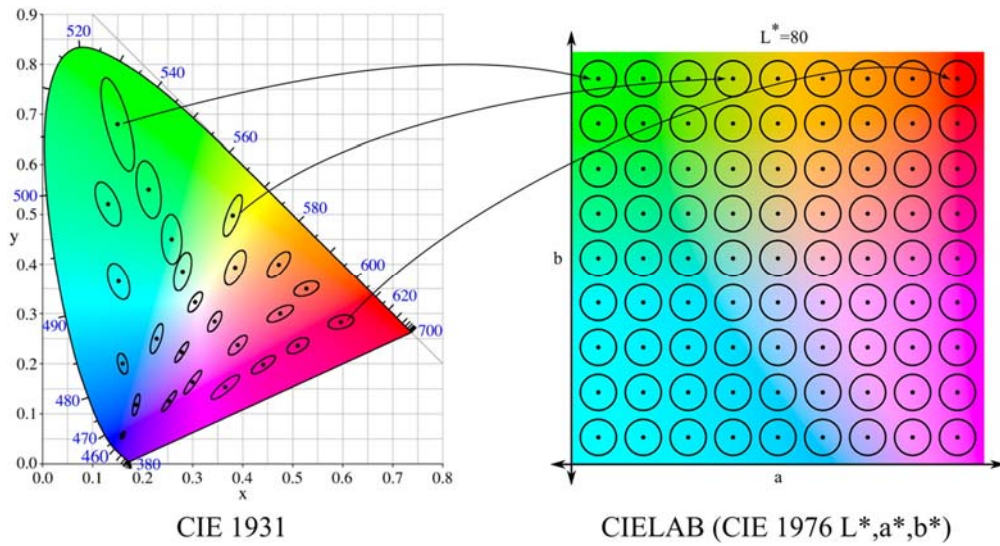


Рис. 10. Слева – срез цветового пространства CIE 1931 с эллипсами МакАдама, Справа – фрагмент среза цветового пространства CIE LAB с эллипсами МакАдама

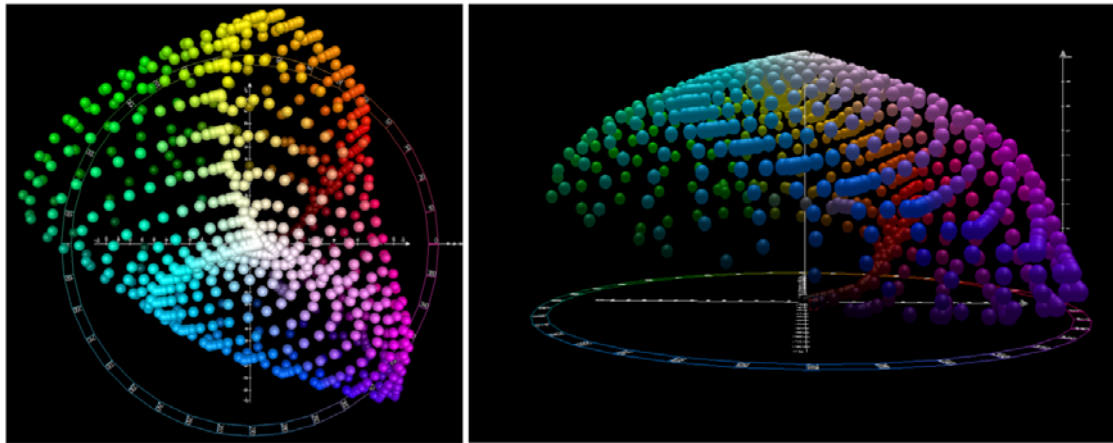


Рис. 11. Цветовое пространство CIE LAB: слева – вид сверху, справа – вид спереди

Исходя из этого, необходимо выбрать такую цветовую модель, которая будет проста и для задачи формирования выборки, и для перевода в CIE LAB. Таким пространством является CIE LCH (рис. 12), [13].

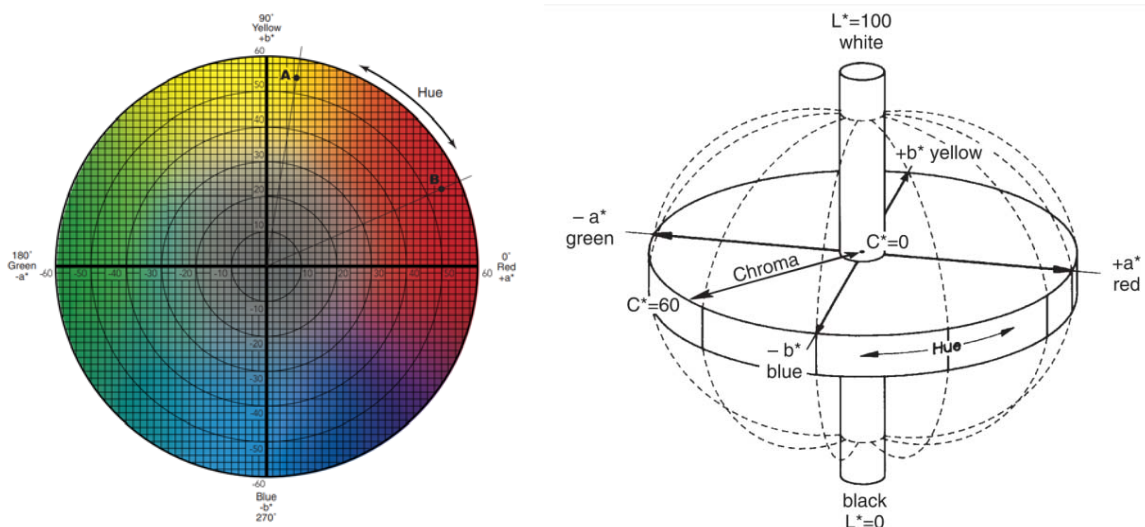


Рис. 12. Цветовое пространство CIE LCH

Как видно из рисунка 12, данное пространство, как и его срезы, имеет простую форму, что позволяет эффективно решить задачу формирования выборки. CIE LAB конвертируется в CIE LCH используя следующие формулы [14]:

$$\begin{aligned}
 L &= L \\
 C &= \sqrt{a^2 + b^2} \\
 H &= \begin{cases} \arctan\left(\frac{b}{a}\right), & \text{если } \arctan\left(\frac{b}{a}\right) \geq 0 \\ \arctan\left(\frac{b}{a}\right) + 360^\circ, & \text{если } \arctan\left(\frac{b}{a}\right) \leq 0 \end{cases}
 \end{aligned}
 \tag{13}$$

Использование пространства CIE LCH для построения пути обхода нежелательно, так как эллипсы МакАдама имеют неправильную форму (рис. 13), что ухудшает линейность восприятия цветового контрастирования.

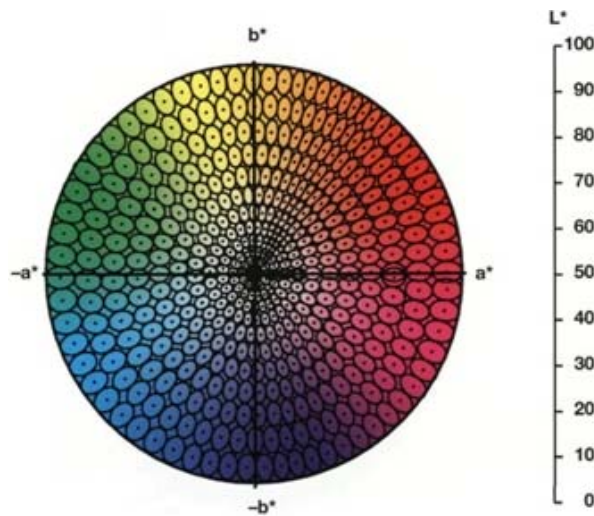


Рис. 13. Срез цветового пространства CIE LCH с эллипсами МакАдама

Однако, в этом пространстве можно формировать выборку цветов для оценки теплостойкости и задавать ключевые точки для пути обхода. После оценки теплостойкости для выборки цветов и оттенков выборку можно конвертировать в CIE LAB по формулам:

$$\begin{aligned}
 L &= L \\
 a &= C * \cos(H) \\
 b &= C * \sin(H)
 \end{aligned}
 \tag{14}$$

После чего можно приступать к поиску пути обхода цветового пространства.

Цветовое контрастирование

После выбора цветового пространства необходимо задаться путем его обхода. Для удобства описание и анализ будут проводиться для цветового пространства CIE 1931.

Рассмотрим пример построения пути обхода. Допустим, сигнал для обработки представляет из себя линейно нарастающий от минимальной яркости до максимальной клин, то есть градиционный клин яркости (рис. 14).

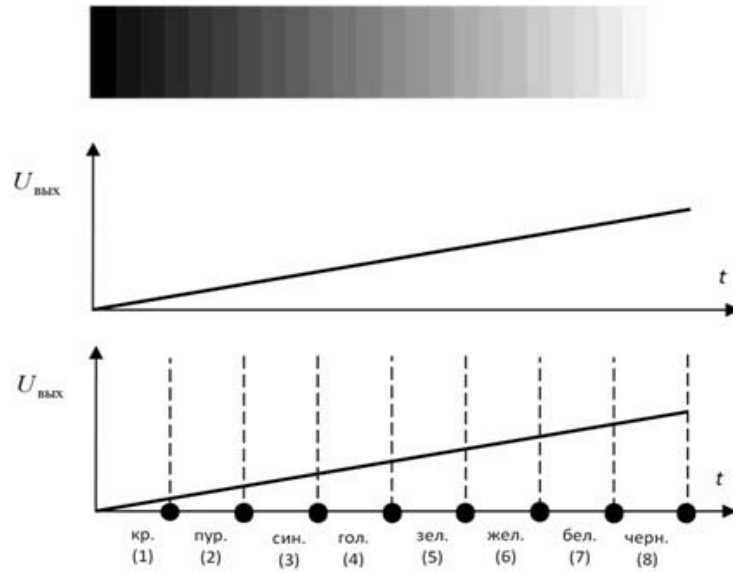


Рис. 14. Градационный клин яркости и его графическое представление

Разобьем его диапазон на 8 участков с присвоением каждому своей гаммы цветов на срезе цветового пространства CIE 1931. Это треугольный алгоритм с разнесенными в пространстве точками начала и конца пути обхода (рис. 15).

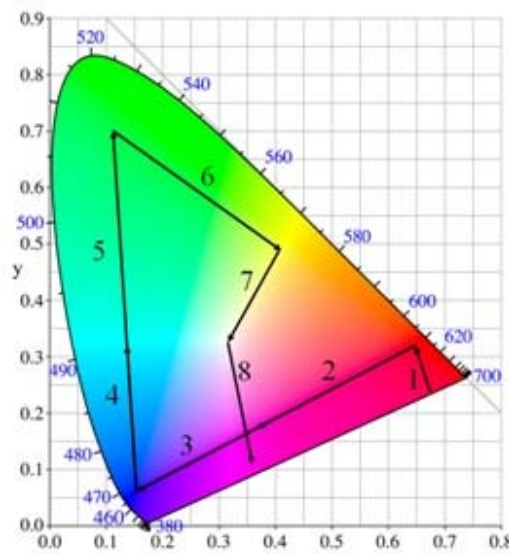


Рис. 15. Треугольный алгоритм с разнесенными в пространстве точками начала и конца пути обхода

Представим алгоритм в формульном виде. Универсальная формула для алгоритмов пути обхода следующая:

$$u_R + u_G + u_B = const \tag{15}$$

Однако значения сигналов u_R u_G u_B могут отличаться в зависимости от алгоритма. Например, для рассмотренного выше алгоритма сигналы u_R u_G u_B имеют вид, приведённый на рисунке 16:

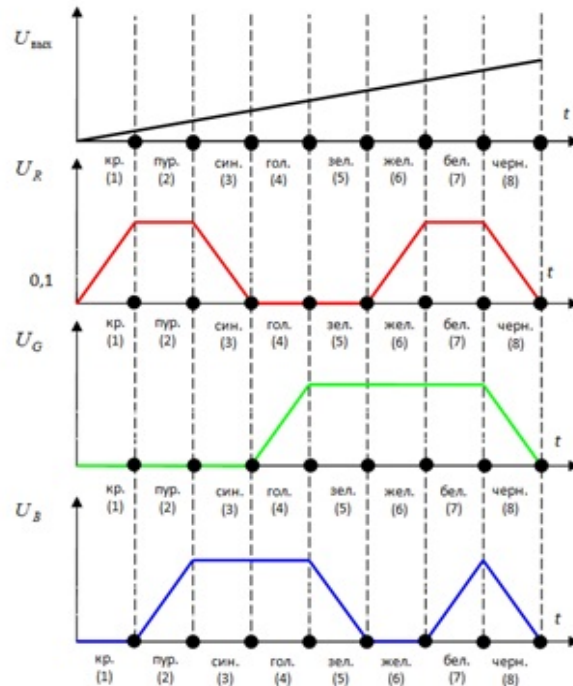


Рис. 16. Сигналы яркости и цветовых составляющих для треугольного алгоритма с разнесенными в пространстве точками начала и конца пути обхода

Существует множество алгоритмов построения пути обхода [15,16], каждый из которых обладает как достоинствами, так и недостатками. Например, стигмаобразный алгоритм обладает высоким быстродействием, а спирально треугольный - большим покрытием оттенков цветового пространства (рис. 17).

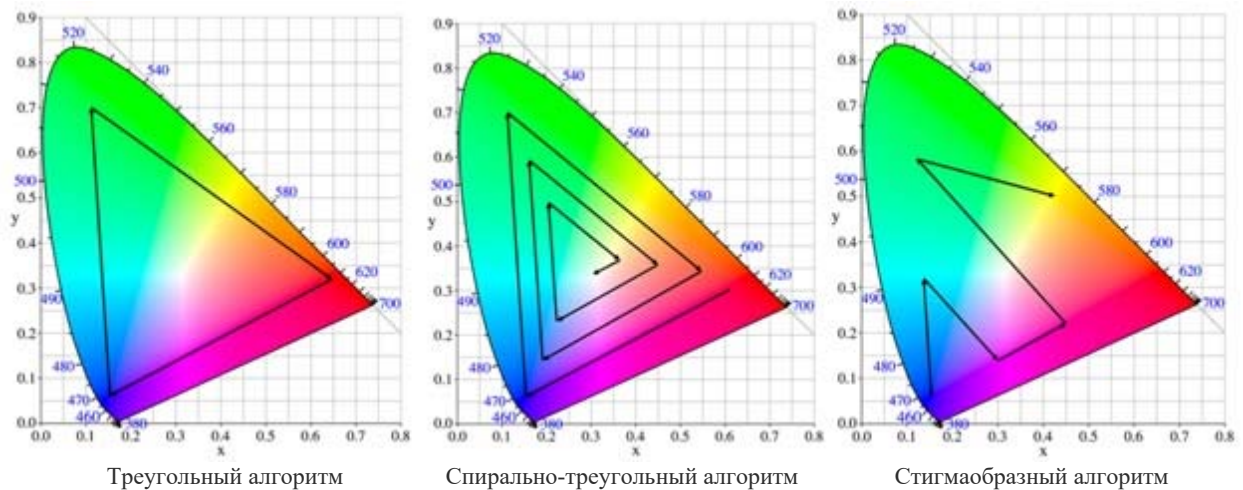


Рис. 17. Алгоритмы путей обхода

Однако, рассмотренные алгоритмы либо охватывают малое количество оттенков, либо интуитивно непонятны, что может негативно сказываться на интерпретации оператором результатов обработанного изображения. Например, спирально-треугольный алгоритм повторяет один и тот же тон для нескольких уровней яркости, что ухудшает интуитивное восприятие при интерпретации результатов.

Для построения интуитивно понятного алгоритма, покрывающего максимально возможное число оттенков, необходимо добавить в алгоритм учет параметра, который поможет человеческому восприятию. Так как в работе решается задача цветового контрастирования, можно выбрать в качестве такого параметра теплостуждность цвета. Использование параметра теплостуждности цвета для контрастирования изображения часто применяется, например, в тепловизионных системах. Равномерное использование областей цветового пространства и учет теплостуждности при построении пути обхода позволит создать интуитивно понятный и эффективный метод цветового контрастирования, подходящий для обработки изображений с большим динамическим диапазоном.

Заключение

В статье был проведен обзор современных систем рентгенографии и параметров рентгенографических изображений. Приведены и проанализированы цветовые пространства, использующиеся в современной колориметрии и некоторые алгоритмы путей обхода для цветового контрастирования. Показана и обоснована необходимость создания нового метода цветового контрастирования, основанного на учете теплоты цветовых оттенков.

Литература

1. Косарина Е.И., Крупнина О.А., Демидов А.А., Турбин Е.М. Цифровая радиография в неразрушающем контроле авиационной техники // *Авиационные материалы и технологии*. 2017. №5. С. 562-574.
2. Мазуров А.И., Потрахов Н.Н. О технологиях рентгеновских систем для контроля электронных компонентов // *Известия высших учебных заведений России. Радиоэлектроника*. 2019. №3. С. 113-120.
3. Чиликин А.Ю., Едомин Д.М., Корякин Д.В. Критерии выявляемости дефектов в сварных соединениях газопроводов методами неразрушающего контроля как способ повышения надежности сетей газораспределения и газопотребления // *Газовая промышленность*. 2018. №5. С. 84-90.
4. Buzug T. M. *Computed Tomography. From Photon Statistics to Modern Cone-Beam CT*. Berlin: Springer-Verlag Berlin Heidelberg, 2008.
5. Vaga R., Bryant K. Recent advances in x-ray technology // 2016 Pan Pacific Microelectronics Symposium (Pan Pacific). 2016.
6. Romano S., Brandão A., Gumpinger J., Gschweidl M., Beretta S. Qualification of AM parts: Extreme value statistics applied to tomographic measurements // *Mater. Des.* 131. 32-48. 2017.
7. B. Dutton W., Vesga J., Waller S. James, Seifi M. Additive Manufacturing Defect Formation and NDE Detectability // *ASTM Symposium on Structural Integrity of Additive Manufactured Parts*, Washington, DC, Nov. 6-8, 2018, ASTM STP 1620, M. Seifi and N. Shamsaei, Eds., ASTM International, West Conshohocken, PA, 2018.
8. Власюк И.В. Разработка модели зрительной системы человека для метода объективного контроля качества изображений в системах цифрового телевидения // *T-Comm: Телекоммуникации и транспорт*. 2009. №1.
9. Potashnikov A.M., Vlasuyk I.V., Ivanchev V.V., Balabanov A.V. The method of representing grayscale images in pseudo color using equal-contrast color space // *2020 Systems of Signals Generating and Processing in the Field of on Board Communications*. M.: Media Publisher, 2020.
10. Горьков А. О цветовых пространствах. URL: <https://habr.com/ru/post/181580/> (дата обращения: 10.08.2022).
11. Pasmantier, Nathaniel & Munakomi, Sunil. 2020. Physiology, Color Perception.
12. Ly, Bao & Dyer, Ethan & Feig, Jessica & Chien, Anna & Bino, Sandra. 2020. Research Techniques Made Simple: Cutaneous Colorimetry: A Reliable Technique for Objective Skin Color Measurement. *The Journal of investigative dermatology*. 140. 3-12.e1. 10.1016/j.jid.2019.11.003.
13. Durmus/ Alp. (2020). CIELAB color space boundaries under theoretical spectra and 99 test color samples. *Color Research & Application*. 45. 796-802. 10.1002/col.22521.
14. Kishimoto, Junko & Yamaguchi, Masahiro & Ohyama, Nagaaki. (2009). A Subjective Evaluation of High Chroma Color with Wide Color Gamut Display. 10.1117/12.805586.
15. Khaleel, Yazen & Muhammad Ali, Peshawa. A proposed method for colorizing grayscale images. 2013.
16. Semary, Noura. *Image Coloring Techniques and Applications*. 2011.

COLOR CONTRASTING IN X-RAY CONTROL

Vasily V. Ivanchev,
Postgraduate MTUCI, Moscow, Russia,
demidra18@gmail.com

Abstract

The article analyzes the current state of radiographic systems with the definition of the key parameters of the image obtained using these systems. An analysis of color spaces and color contrasting methods was also carried out, and a conclusion was made about the need to create a new color contrasting method that takes into account the warmth of color shades.

Keywords: radiography, color spaces, dynamic range processing, color contrasting.

МОДЕЛИРОВАНИЕ И РАСЧЁТ АКУСТИЧЕСКИХ ПАРАМЕТРОВ ТЕАТРАЛЬНОГО ЗАЛА СО СЛОЖНОЙ АРХИТЕКТУРНОЙ ФОРМОЙ

Чернышева Татьяна Васильевна,
доцент кафедры ТуЗВ МТУСИ, к.т.н., доцент, Москва, Россия,
krba2012@yandex.ru

Лукашова Екатерина Алексеевна,
магистрант МТУСИ, Москва, Россия,
iasha200@bk.ru

Попов Олег Борисович,
профессор кафедры ТуЗВ МТУСИ, к.т.н., профессор, Москва, Россия,
olegp45@yandex.ru

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

Аннотация:

Изложены результаты выбора акустического оформления и системы звукоусиления сложного театрального сценического пространства. Проведен анализ основных акустических характеристик и дополнительных критериев качества театрального зала. Проанализированы распределения уровней звукового поля, коэффициентов разборчивости речи и потери согласных звуков в речи по площади зрительного зала. Даны рекомендации по улучшению акустических характеристик рассматриваемого помещения.

Ключевые слова: уровни звукового поля, дополнительные критерии качества, системы звукоусиления, коэффициент разборчивости речи, коэффициент потери согласных.

Для любых помещений, предназначенных для массовых мероприятий с большим количеством участников необходимо, прежде всего, обеспечить необходимое качество звучания в помещении. Одним из требующих внимательного индивидуального изучения акустических характеристик является помещение театра, иногда являющегося помещением не типовой конструкции с довольно сложной геометрией сценического пространства и зрительного зала. [1,2].

Главной задачей звукорежиссеров в этом случае становится создание качественной акустической атмосферы в зале и донесение ее до слушателей. Зачастую, для ее выполнения используются системы звукоусиления, позволяющие получить равномерное распределение уровней звукового поля, что является одним из условий хорошего качества звучания [5].

В свою очередь, выбор системы звукоусиления предполагает необходимый расчет, который позволит определить количество источников звука и их расположение. Представляет интерес рассмотрение основных и дополнительных акустических характеристик для зала сложной формы. В качестве примера расчета, был выбран зал «Стравинский» Музыкального театра «Геликон-Опера», план которого представлен на рисунке 1.

Московский Музыкальный театр был организован в старинном особняке – помещении, не предназначенном для театральных представлений. Особую роль в рассматриваемом помещении играет, оставленное от бывшей постройки «крыльцо» – VIP-ложа для почетных гостей. Потолок «крыльца» имеет купольную форму, а сама крыша выполнена в виде пирамиды. Для его опоры служат четыре колонны сложной геометрической формы. Организация зрительских мест напоминает форму греческого амфитеатра, а между первыми рядами и сценой располагается оркестровая яма, находящаяся ниже поверхности пола на 4 м [8].

Обработка интерьера помещения была оставлена без изменения и состоит из кирпичной кладки. Выбор такого материала для помещений оперных театров не желателен из-за большой величины коэффициента поглощения, особенно на высоких частотах, что снижает степень диффузности звукового поля. Поэтому, для получения требуемых акустических характеристик в таких залах необходимо использовать дополнительные приемы [6]. В частности, в зале «Стравинский» была введена особая акустическая система в виде «геликонов» – сферических звукоотражающих элементов, подвешенных к потолку на разной высоте. В зависимости от типа представлений, высота подвеса «геликонов» регулируется.

Основная задача обработки интерьеров оперных театральных помещений заключается в выборе таких материалов, которые могли бы обеспечивать требуемые акустические характеристики. Зрительские кресла обычно выполняются из акустического дерева с бархатной обивкой [11].

В рассматриваемом зале для лучшего звучания низких частот выбраны двойные деревянные полы. Кроме того, для решений задач акустики и технического обеспечения сценических действий в зале были введены специальные платформы, отделанные дубовыми панелями. Выбранный материал позволит улучшить акустические свойства зала.

Архитектурный вид зала приведён на рисунке 2 [8].

Естественная конструкция бывшего здания не позволяет достичь требуемых характеристик.

В соответствии с рекомендациями [10,11] отношение длины зала к его средней ширине следует принимать в интервале от 1 до 2. При отношении, меньшем 1 (широкий зал малой длины), получается нежелательное запаздывание отражений от боковых стен, и вследствие направленности источников звука ухудшается слышимость на боковых местах [9,15]. В рассматриваемом зале из-за особенностей архитектуры данное требование не выполняется: отношение длины выбранного зала к его ширине составляет 0,66, что негативно влияет на качество создания звуковой панорамы.

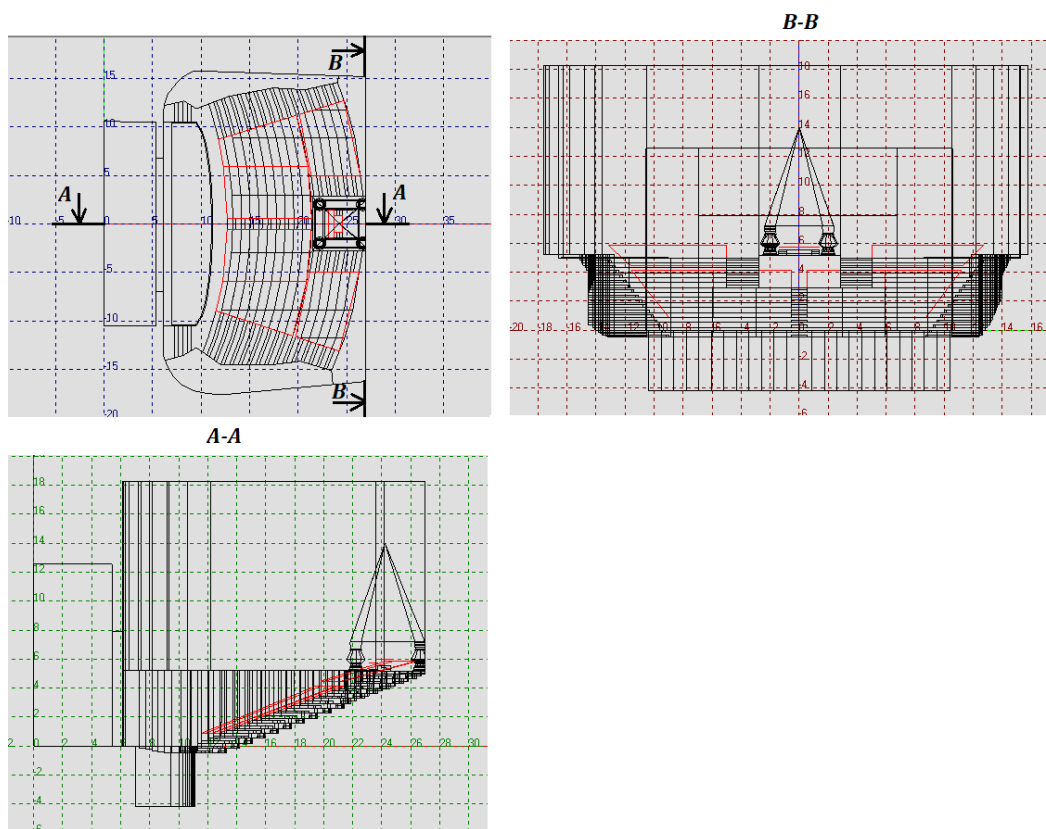


Рис. 1. План и разрезы музыкального зала «Стравинский» [8]

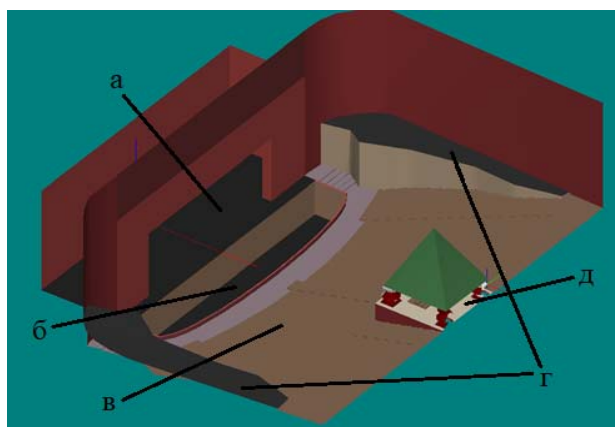


Рис. 2. Архитектурный вид зала:

а – сцена, б – оркестровая яма, в – зрительские места, г – боковая платформа, д – «крыльцо»

Для оценки акустических параметров, прежде всего, необходимо оценить основную акустическую характеристику – характеристику времени реверберации при объеме помещения зала 12532 м³.

С учётом этого весьма важен выбор системы звукоусиления и расположение источников звука. Опираясь на рекомендации и особенности зала, был выбран один из наиболее приемлемых вариантов места установки распределенной системы звукоусиления [13,14].

Система звукоусиления состоит из трёх источников звука, размещённых на каждой стороне платформы, на высоте в 3,5 м от уровня пола последних рядов и двух источников звука на задней стенке зала, размещённых на высоте 8 м от уровня пола последних рядов.

Известно, что увеличение времени реверберации уменьшает разборчивость речи, но способствует более слитному звучанию музыкальных произведений, поэтому необходимо обеспечить для исследуемого помещения такое время реверберации, при котором будет наибольшая разборчивость речи с лучшим музыкальным звучанием [12]. Кроме того, следует учитывать влияние прямого поля звука актеров на сцене на распределение суммарного поля в зрительном зале [7].

Результаты расчета частотной характеристики времени реверберации зала представлены на рисунке 3.

В зале театра важна передача не только голосов актеров на сцене, но и звучание музыкальных произведений. Для таких залов рекомендуемое оптимальное время реверберации должно составлять 1,6 с. При этом, на интервале измерений, допустимое отклонение от оптимальной величины не должно превышать 10% [3].

На рисунке 3 видно, что во всем диапазоне частот измеренное время реверберации звука в зале соответствует указанным рекомендациям с отклонением не более чем 9,5% от оптимального значения

Однако, обеспечение требуемого времени реверберации не полностью характеризует целостную картину акустики зала.

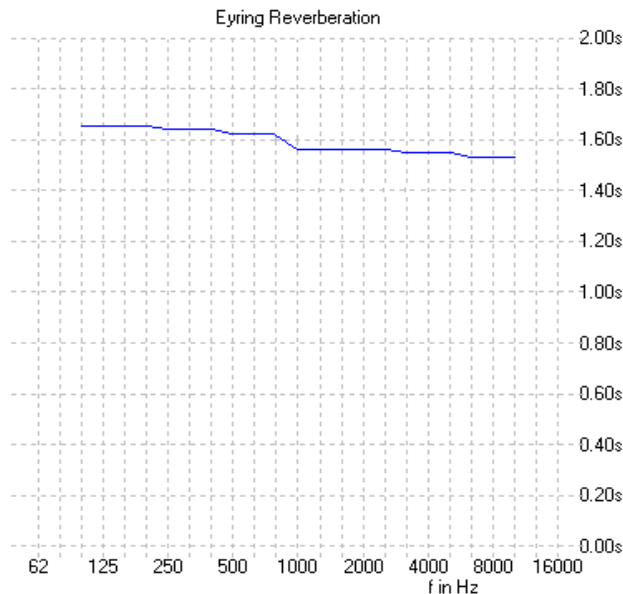


Рис. 3. Частотная характеристика времени реверберации

В сложном театральном сценическом пространстве для достижения качественных характеристик акустики зала, требуется также обеспечить необходимый уровень звукового поля при допустимых показателях его неравномерности [4].

На рисунке 4 приведены кривые, построенные по результатам расчетов частотных характеристик звукового поля для максимальных, средних и минимальных уровней звукоусиления в зале.

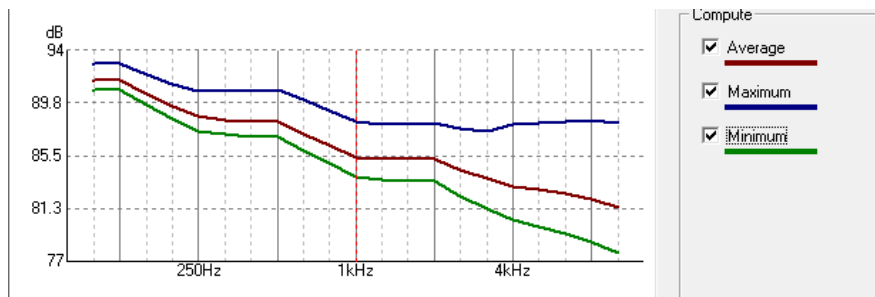


Рис. 4. Частотные характеристики уровней звукового поля в зале

Известно, что при воспроизведении музыкальных программ, необходимо обеспечить для комфортного восприятия звука на местах расположения слушателей уровень звукового поля в интервале 90÷94 дБ, а при прослушивании речи – 80÷86 дБ [2].

На рисунке 5 приведено распределение уровня звукового поля по площади зала с учётом его архитектуры.

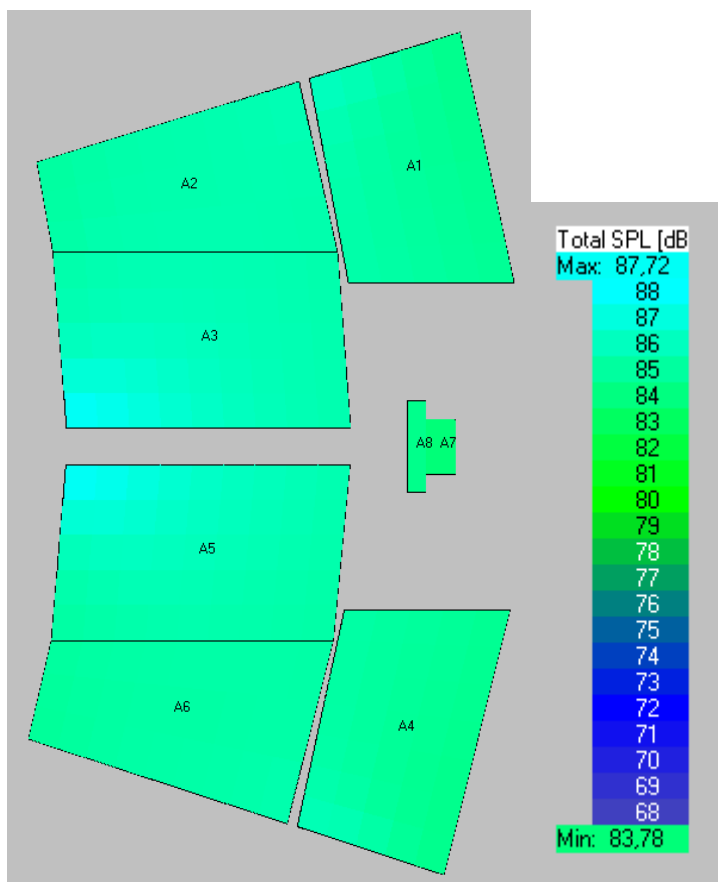


Рис. 5. Распределение уровня звукового поля по площади в зале

Из рисунка 5 видно, что уровень звукового поля меняется от 83,78 до 87,72 дБ.

Для систем озвучивания речи данный диапазон уровней укладывается в предельно-допустимые величины, однако для систем озвучивания музыкальных программ является не комфортным, так как ниже требуемых оптимальных величин 90÷94 дБ. [11]. Благодаря амфитеатральному расположению слушателей, уровни звукового поля распределены примерно равномерно по всем зрительским местам в зале. Вместе с тем отчётливо наблюдается заметное отклонение в самом «крыльце», несмотря на наличие симметрично расположенных на задней стенке зала двух источников звука. При этом неравномерность звукового поля в этой зоне составляет 3,94 (дБ). В то же время такая неравномерность укладывается в предельно-допустимое значение при воспроизведении музыки и допустима для воспроизведения речи [6].

Таким образом, с учетом, представленного на рисунке 5 распределения уровней звукового поля, не соответствующего оптимальным значениям 90÷94 дБ, выбранного количества источников звука оказалось недостаточно для обеспечения основных акустических характеристик на всей площади зала, что также влияет на дополнительные критерии качества акустики зала [15].

Как отмечалось выше, зал «Геликон-опера» имеет сложную, неправильную геометрическую форму, и поэтому, для получения более полной акустической картины зала в условиях достижения основных критериев качества, также необходимо исследовать дополнительные критерии. Для этого были выбраны параметры, характеризующие разборчивость речи в помещениях: индекс разборчивости речи и коэффициент потери согласных.

Разборчивость речи является интегральной характеристикой и зависит от многих факторов: громкости речи, объема зала, выбранных источников звука и от их расположения. Поэтому, индекс разборчивости весьма важен для выбора системы звукоусиления в зале, поскольку именно по нему будет оцениваться насколько слышны и различимы составные элементы языка [5].

На рисунке 6 приведено распределение коэффициента разборчивости речи по площади зала.

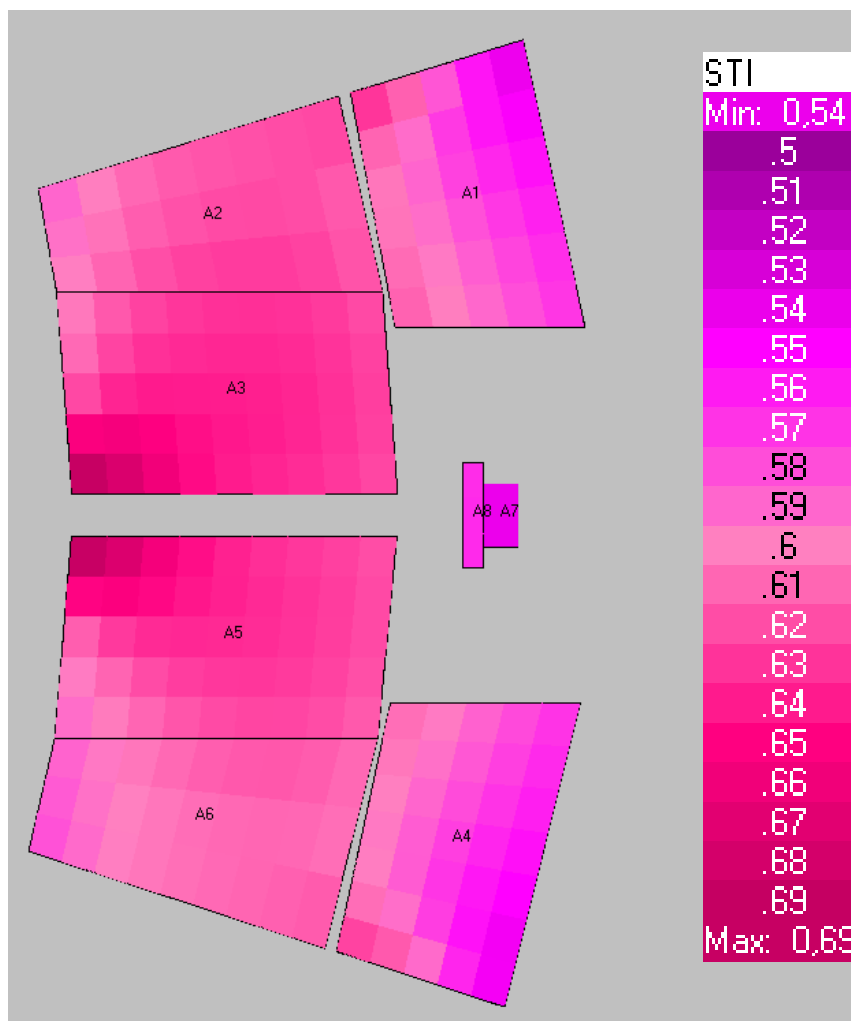


Рис. 6. Распределение коэффициента разборчивости речи в зале

Разборчивость уменьшается при увеличении расстояния между источником и слушателем, что особенно заметно наблюдается на рисунке 6 в «крыльце», где не были предусмотрены дополнительные источники звука.

В то же время, при выбранном расположении источников звука в зале, в целом достигается хороший результат коэффициента разборчивости речи для всех зрительских мест, за исключением зоны «крыльца». При этом, наилучший результат имеет место в центре зала – там коэффициент разборчивости достигает 0,69.

Как известно, разборчивость речи зависит, прежде всего, от правильного понимания согласных букв. Реверберация звука и шум искажают разборчивость речи, при этом потери согласных звуков в речи позволяет произвести оценку разборчивости речевой информации [5,14]. Допустимым значением коэффициента потери согласных звуков в речи без снижения качества её восприятия является значение, не превышающее 7%.

На рисунке 7 приведены данные распределения по площади зала коэффициента потерь согласных звуков в речевой информации.

Анализ коэффициента потерь согласных звуков речи по всей площади зала (рис. 7) показывает, что на местах размещения зрителей этот коэффициент не превышает 5-6%, однако с приближением к задним рядам коэффициент потерь возрастает до 9%, причём это фиксируется в тех же зонах зала, где из-за неравномерности звукового поля его уровень не достигался требуемых значений (зоны A1, A4, A7 и A8).

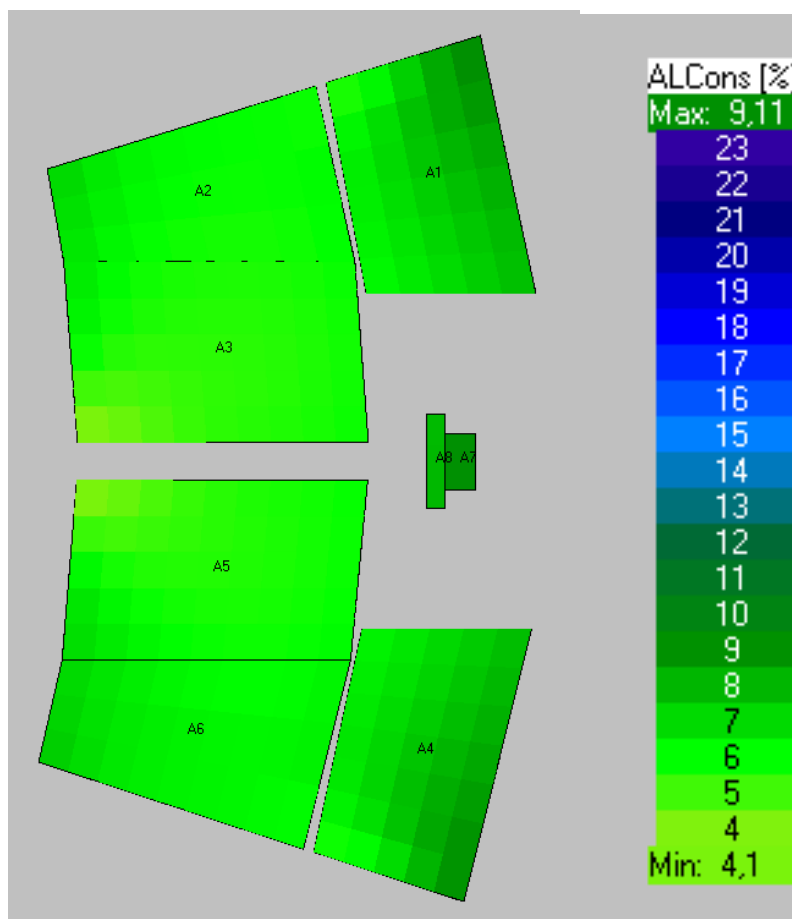


Рис. 7. Распределение коэффициента потери согласных по площади

Заключение

На основании построенной геометрической модели Большого зала «Стравинский», учитывающей его сложную архитектуру, дополнительные поверхности и покрытия, в зале обеспечивается нормативное значение времени реверберации во всём звуковом диапазоне частот.

Как показывают предварительные расчеты, ухудшение основных акустических параметров и дополнительных критериев качества звучания в зале фиксируется в зоне «крыльца» (A7, A8). В этой зоне необходимо разместить дополнительные звуковые источники.

Кроме того, анализ предварительных результатов моделирования показал, что количество источников звука, расположенных с каждой боковой стороны платформ (рис. 1), может ограничиваться двумя источниками. Целесообразно также опустить на 6-7% по высоте симметрично расположенные источники звука на задней стенке зала, а также добавить к ним ближе к углам зала по одному направленному в центр источнику звука. Это позволит достичь более равномерного распределения уровней звукового поля в задних рядах, а также улучшить основные акустические параметры и дополнительные критерии качества звука в зонах «крыльца».

Литература

1. Вахитов Ш.Я., Ковалгин Ю.А., Фадеев А.А., Шевьев Ю.П. Акустика: Учебник для вузов / Под ред. профессора Ю.А. Ковалгина. М.: Горячая линия – Телеком, 2009. 660 с.
2. Жабыко Е.И., Рублевская Н.И. Акустическое проектирование залов многоцелевого назначения: Учеб. Пособие. Владивосток: Изд-во ДВГТУ, 2008. 89 с.
3. Проектирование акустики зрительных залов. Методические указания: составитель Удлер Е.М.: Казань: КИСИ: 1980, 35 с.
4. Климухин А.А., Киселева Е.Г. Проектирование акустики зрительных залов: учебно-методические указания к курсовой расчетно-графической работе. М.: МАРХИ, 2012. 56 с.
5. Звук в театре [Электронный ресурс] URL: <https://lightsoundnews.ru/zvuk-v-teatre/>, (дата обращения: 23.10.2022).
6. Иордан В.Л. Акустическое проектирование концертных залов и театров. Перев. с англ. С.А. Хомутова, под ред. Л.И. Макриненко. М.: Стройиздат, 1986, 170 с.
7. Системы озвучения и звукоусиления [Электронный ресурс] URL: <https://uz.denemetr.com/docs/294/index-20898-1.html>, (дата обращения: 25.10.2022).

8. Театр «Геликон-опера». URL: <https://iskatel.com/places/teatr-gelikon-opera>, (дата обращения: 23.10.2022).
9. Ковригин С.Д., Крышовч С.И. Архитектурно-строительная акустика. М.: Высшая школа, 1986. С. 237-242.
10. Архитектура гражданских и промышленных зданий. Гражданские здания / под ред. А.В. Захарова. М.: Стройиздат, 1993. С. 178-181.
11. ТСН 31П317-99. Культурно - зрелищные сооружения. Нормы проектирования. М.: Госгражданстрой, 2000.
12. Рекомендации по проектированию концертных залов: Москомархитектура, 2004.
13. СНиП 2.08.02-89* Общественные здания и сооружения.
14. Кнудсен Верн О. Архитектурная акустика., пер. с англ.: изд.5: 2010. 520 с.
15. Оболенский Н.В. Архитектурная физика. Учебник: М.: Архитектура, 2005. 448 с.

MODELING AND CALCULATION OF ACOUSTIC PARAMETERS OF THEATER A HALL WITH A COMPLEX ARCHITECTURAL FORM

Tatyana V. Chernysheva,
Associate Professor of the Department of T&SB MTUCI,
Ph.D., Moscow, Russia,
krba2012@yandex.ru

Ekaterina A. Lukashova,
Graduate MTUCI, Moscow, Russia,
iasha200@bk.ru

Oleg B. Popov,
Professor of the Department of T&ZV MTUCI,
Ph.D., Professor, Moscow, Russia,
o.b.popov@mtuci.ru

Vladimir G. Orlov,
Chief specialist of the Department
of OofSRWofS of MTUCI, Ph.D., Moscow, Russia,
v.g.orlov@mtuci

Abstract

The results of the choice of acoustic design and sound amplification system for a complex theatrical stage space are presented. The analysis of the main acoustic characteristics and additional quality criteria of the theater hall is carried out. The distributions of sound field levels, index of speech intelligibility and percentage loss of consonants over the space of the hall are analyzed. Recommendations are given to improve the acoustic characteristics of the considered hall.

Keywords: *sound field levels, additional quality criteria, sound amplification systems, speech intelligibility index, percentage loss of consonants.*