Санкт-Петербургский государственный университет

Кафедра системного программирования

Группа 21.Б11-мм

Разработка технологий для реализации “цифрового почерка” абонента

Задорожный Александр Сергеевич

Отчёт по учебной практике
в форме «Производственное задание»

Научный руководитель:
д.ф.-м.н., проф. каф. СП А.Н. Терехов

Консультант:
главный конструктор ООО «ЛИС» В.О. Аксенов

Санкт-Петербург
2023

Оглавление

Введение

В современном информационном обществе, с увеличением влияния цифровых технологий в различных сферах жизни, вопросы безопасности и надежной идентификации пользователя в интернете приобретают все большую значимость. Развитие таких областей, как интернет вещей[1] и увеличение объемов онлайн-транзакций[2], делает необходимым постоянное усовершенствование методов защиты и верификации личных данных. Это в свою очередь, требует интеграции различных технологий, включая элементы биометрии и машинного обучения, для повышения надежности и точности идентификации. В этом контексте, концепция “цифрового почерка” абонента выступает как одно из перспективных направлений, предлагая новые подходы к обеспечению кибербезопасности.

Цифровой почерк представляет собой уникальный набор поведенческих характеристик пользователя, таких как стиль взаимодействия с пользовательским интерфейсом, образцы движения мыши и другие параметры, которые могут быть использованы для повышения безопасности в компаниях. Внедрение системы “свой-чужой” на основе анализа цифрового почерка позволяет выявлять необычное или подозрительное поведение даже после того, как устройство прошло техническую и физическую проверку. Т.е если поведение отклоняется от установленной нормы, система может автоматически инициировать дополнительные меры безопасности, включая временную блокировку доступа пользователя к корпоративным ресурсам. Такой подход обеспечивает дополнительный слой безопасности, позволяя предотвращать внутренние угрозы и атаки, основанные на украденных учетных данных [48, 49].

Переходя от общей концепции ”цифрового почерка” к более узкому фокусу на анализе сетевых данных, становится ясно, что сетевой трафик абонента несет в себе большое количество уникальных характеристик, которые могут быть ключевыми в идентификации конечного

[1] Statista: Internet of Things - Worldwide [Электронный ресурс]. (Дата обращения 10.11.2023)
[2] Statista: Digital Payments - Worldwide [Электронный ресурс]. (Дата обращения 10.11.2023)

пользователя. Именно анализ сетевого трафика позволяет глубоко понять индивидуальные особенности поведения человека в сети, что является особенно актуальным в условиях растущего количества сетевых взаимодействий и необходимости обеспечения высокой степени безопасности в цифровом пространстве [10, 14, 42, 43]. Это подводит к сути проекта, т.е создание комплексной системы идентификации личности, основанной на анализе цифрового почерка абонента, объединяющей поведенческие и сетевые характеристики для обеспечения высокоуровневой безопасности и точной верификации пользователя в цифровом мире.

Эта система разрабатывается с целью быть универсальной, чтобы она могла применяться в широком спектре компаний, охватывая подавляющее большинство сценариев использования. В данной работе будет представлена конкретная реализация этой системы для корпоративного мессенджера IVA[3], демонстрируя её эффективность и адаптивность в конкретном бизнес-контексте. Кроме того, будет описана общая схема интеграции разработанного инструмента с существующим продуктом компании Mobil-Group, который предназначен для анализа сетевого трафика. Это позволит понять, как новая система может усиливать и дополнять уже существующие механизмы анализа и безопасности, обеспечивая более глубокий и многоуровневый подход к защите данных и идентификации пользователя.

[3]Корпоративный мессенджер IVA [Электронный ресурс]. (Дата обращения 05.10.2023)

Дата сборки: 25 декабря 2023 г.

1. Постановка задачи

Целью работы является разработка и внедрение технологий для реализации "цифрового почерка" абонента, предназначенных для улучшения безопасности и идентификации пользователей в корпоративных сетях и приложениях, с помощью передовых методов машинного обучения и анализа поведенческих данных. Для её выполнения были поставлены следующие задачи.

1. Изучение существующих алгоритмов и техник распознавания поведенческих паттернов пользователей, включая анализ клавиатурного ввода, стилей взаимодействия с интерфейсом и прочих уникальных характеристик.

2. Разработка методов сбора и анализа сетевых данных, включая временные метки, частоту запросов, паттерны использования приложений и сервисов.

3. Разработка архитектуры системы, способной адаптироваться к различным корпоративным средам и приложениям. А также ее апробация на примере мессенджера IVA.

4. Обучение моделей для распознавания активности и пользователей. Сюда входит как подготовка обучающих данных, включая разметку и пред обработку для эффективного обучения моделей, так и настройка моделей для повышения точности и надёжности в идентификации индивидуального цифрового почерка.

5. Интеграция разработанного инструмента в существующие корпоративные системы организации Mobil-Group, предоставляющие инструменты анализа сетевого трафика и безопасности.

2. Обзор

В данном разделе представлен обширный анализ текущего состояния и перспектив развития технологий, связанных с идентификацией и верификацией личности в цифровой среде. Особое внимание уделено изучению различных видов цифрового почерка, их уникальных характеристик и приложений в различных сферах. Рассматриваются существующие аналоги и решения, уже применяемые в индустрии, что позволяет оценить текущее состояние технологий и выявить потенциальные возможности для дальнейшего развития. Кроме того, проанализированы методы классификации сетевого трафика, которые играют ключевую роль в определении и анализе цифрового почерка, особенно в контексте повышения безопасности и точности идентификации в условиях постоянно изменяющейся сетевой среды. И, наконец, освещаются используемые технологии, что дают комплексное представление о том, как различные инструменты взаимодействуют для создания эффективных систем идентификации личности.

2.1. Виды цифрового почерка

В первую очередь рассмотрим клавиатурный почерк пользователя‘[24]. Он основан на анализе уникальных поведенческих характеристик при вводе данных на клавиатуре, включая скорость набора, ритм, интервалы времени между нажатиями клавиш и время их удержания. Эти факторы обеспечивают уникальность клавиатурного почерка, делая его важным инструментом в задачах безопасности [50, 45]. Однако потенциальная нестабильность характеристик и возможность подделки при достаточно детальном изучении стиля набора текста требуют постоянной корректировки и обновления эталонных моделей для обеспечения точной идентификации.

Далее стоит обратить внимание на динамику работы с манипулятором “мышь”[47]. Анализируя скорость движения курсора, паттерны нажатий и общее взаимодействие с мышью, можно выявить уникальные особенности, которые служат ценным инструментом для идентифика-

ции личности. Этот вид цифрового почерка ценится за его способность предоставлять детальную информацию о пользователе, что делает его подделку сложной задачей [46]. Однако, несмотря на эти преимущества, данный метод имеет и недостатки. Влияние внешних факторов, таких как физическое состояние пользователя [34] или использование различных устройств ввода, может влиять на точность и стабильность динамики работы с мышью, а следовательно привести к ошибкам в идентификации. Кроме того, этот метод требует специального программного обеспечения для сбора и анализа данных, что может ограничивать его применение в некоторых средах.

Идентификация пользователей смартфонов на основе их взаимодействия с устройством [19, 27, 20] является передовым методом в области цифровой идентификации. Она включает анализ данных акселерометра, геолокации, GPS, а также наблюдение за тем, как пользователи кладут свой телефон на стол или держат его в руках. Эти параметры предоставляют обширный набор поведенческих данных, способствуя точной идентификации, затрудняя процесс имитации уникальных характеристик пользователя. Однако, существуют недостатки данного метода. Изменения в поведении пользователя, вызванные, например, травмами рук или изменением привычек использования [34, 17], могут снижать точность идентификации. Кроме того, вопросы приватности и сбора данных могут вызывать опасения среди пользователей, требуя баланса между точностью идентификации и защитой личных данных [18].

Биометрия, включающая идентификацию по отпечаткам пальцев, распознавание лица и другие физиологические характеристики, считается одной из самых надежных форм идентификации [40, 4]. Она широко используется в различных сферах, от безопасности доступа до транзакционных систем. Однако, помимо риска компрометации биометрических данных, существуют и другие недостатки. Ключевым из них является необходимость постоянного подтверждения личности, особенно в динамичных средах, где требуется частый доступ к защищенным ресурсам. Это может создавать неудобства для пользователей и требовать наличия непрерывного доступа к биометрическим устройствам.

Кроме того, проблемы с функционированием или доступностью этих устройств могут привести к сложностям в обеспечении стабильного и надежного доступа к защищенным системам или услугам.

Последний вид “цифрового почерка”, основывается на анализе сетевого трафика [32]. Данный способ является ключевым элементом в современных методах идентификации. Он включает изучение паттернов использования интернета, предпочтений в веб-поиске и общих поведенческих особенностей пользователя [37]. В современном мире, где интернет стал неотъемлемой частью жизни большинства людей и компаний [8], этот подход приобретает особую актуальность. Он позволяет точно идентифицировать “своих” и “чужих” в цифровом пространстве, обеспечивая высокий уровень безопасности. Одним из главных преимуществ анализа сетевого трафика является его способность непрерывно собирать и обновлять данные, оставаясь незаметным для пользователя [16]. Это обеспечивает постоянное обновление информации о поведенческих паттернах, увеличивая точность идентификации.

Таким образом, в условиях глобализации и всё более тесной интеграции бизнеса и личной жизни с интернетом, анализ сетевого трафика становится одним из самых эффективных способов обеспечения безопасности и контроля доступа. Это делает его важным инструментом в арсенале современных методов кибербезопасности.

2.2. Существующие аналоги

В первую очередь стоит обратить внимание на системы анализа сетевого трафика (Network Traffic Analysis, NTA) [21], которые предназначены для перехвата потоков данных и обнаружения признаков сложных, чаще всего целевых атак (APT). С их помощью можно проводить ретроспективное изучение сетевых событий, обнаруживать и расследовать операции злоумышленников в информационной инфраструктуре предприятия, а также эффективно реагировать на соответствующие происшествия [1]. Решения класса NTA отслеживают трафик и коммуникации внутри корпоративной сети и применяют различные техники

(поведенческий анализ, машинное обучение) для того, чтобы быстро обнаружить, проанализировать и обработать угрозы, которые в противном случае оставались бы скрытыми. NTA может применяться в сетях любого масштаба, равно как и любой архитектуры: локальной, облачной, гибридной.

В 2019 году компания Gartner проводила исследование решений класса Network Traffic Analysis[4]. Результаты сосредоточены в отчёте «Market Guide for Network Traffic Analysis», согласно которому на сегодняшний день на мировом рынке выбор поставщиков на мировом рынке NTA-решений весьма велик. Однако для решения непосредственно нашей задачи стоит рассматривать представителей рынка, которые специализируются на анализе трафика с использованием таких технологий для обнаружения угроз внутри сети, как машинное обучение, правила детектирования, индикаторы компрометации, глубокая аналитика. Среди всех инструментов можно выделить следующих игроков: Awake Security Platform [3], Bricata [6], Cisco Stealthwatch [9], Corelight Sensor [11], Corvil Security Analytics [12], и др. Хотя изучив данные инструменты более детально, стоит отметить, что все эти решения эффективны в поиске подозрительных или вредоносных данных в сети, однако они оказываются бессильными, когда речь идёт об отслеживании подозрительной активности со стороны сотрудника компании, который собирает или наблюдает за потоками данных, не имея к ним доступа. Это подчеркивает необходимость разработки решений, способных распознавать и предотвращать такие виды инсайдерских угроз. Аналогичная ситуация наблюдается и на российском рынке: PT Network Attack Discovery (Positive Technologies) [29], Threat Detection System (Group-IB) [35], Kaspersky Anti Targeted Attack («Лаборатория Касперского») [23], Гарда Монитор («Гарда Технологии») [44].

После тщательного рассмотрения существующих индустриальных решений и не находя в них подходящих вариантов, стоит перейти к изучению академических исследований в области науки, которые могут предложить новые перспективы и подходы к поставленной зада-

[4]Market Guide for Network Traffic Analysis [Электронный ресурс]. (Дата обращения 18.11.2023)

че. На сегодняшний день проведено множество исследований по теме классификации сетевого трафика [22, 28, 7, 31]. Однако большинство из них сосредоточились на классификации семейства протоколов, также известного как характеристика трафика (например, потоковая передача, чат, P2P и т.д.) либо на идентификации отдельного приложения, которая известна как идентификация приложения (например, Spotify, Hangouts, BitTorrent и т.д.) [25, 15].

Таким образом, на текущий момент не было обнаружено работ, направленных непосредственно на составление цифрового почерка через сетевой трафик с целью обеспечения кибербезопасности и создания системы “свой-чужой”. А значит остаётся значительный потенциал для развития в этом направлении. Если сосредоточить усилия не на классификации протоколов или идентификации приложений, а на анализе действий, которые пользователи выполняют в этих приложениях, можно достичь более глубокого понимания поведения пользователя. Это откроет возможности для создания уникального цифрового почерка каждого человека, взаимодействующего с программой, что будет способствовать повышению уровня кибербезопасности и эффективности системы различения “своих” и “чужих”.

2.3. Методы классификации сетевого трафика

В этом разделе приводится обзор наиболее важных методов классификации сетевого трафика. В частности, данный подход можно разделить на три основные категории следующим образом: (i) основанные на портах, (ii) проверка полезной нагрузки и (iii) статистическое и машинное обучение. Ниже приведен краткий обзор наиболее важных и недавних исследований, касающихся каждого из вышеупомянутых подходов.

Подход, основанный на портах: Классификация трафика по номеру порта является старейшим и наиболее известным методом решения этой задачи [14]. Классификаторы на основе портов используют информацию в заголовках TCP/UDP пакетов для извлечения номера порта, который, как предполагается, связан с конкретным приложени-

ем или действием [30]. Классификация на основе портов, как известно, является одним из самых простых и быстрых методов идентификации сетевого трафика. Однако повсеместное использование обфускации портов, преобразования сетевых адресов (NAT), переадресации портов и другие техники значительно снизило точность этого подхода. Согласно [28, 26], только от 30% до 70% текущего интернет-трафика может быть классифицировано с использованием методов классификации на основе портов. Потому что по вышеупомянутым причинам для классификации современного сетевого трафика необходимы более сложные методы классификации трафика.

Большинство **методов проверки полезной нагрузки, также известных как глубокая проверка пакетов (DPI)**, используют предопределенные шаблоны, такие как регулярные выражения, в качестве подписей для каждого протокола [13, 33]. Производные шаблоны затем используются для отличия протоколов друг от друга. Необходимость обновления шаблонов всякий раз, когда выпускается новый протокол, и проблемы конфиденциальности пользователей являются одними из наиболее важных недостатков этого подхода. И хотя в работе [5] была предложена новая система DPI, которая может проверять зашифрованную полезную нагрузку без расшифровки, таким образом решив проблему конфиденциальности пользователей проблема, но он может обрабатывать только защищенный HTTP (HTTPS) трафик.

Статистический подход и подход к машинному обучению: Статистический подход включает в себя следующие два основных метода.

1. **Простые и сложные статистические методы:** Эти методы основаны на предвзятом предположении, что базовый трафик для каждого приложения обладает некоторыми статистическими характеристиками, которые практически уникальны для каждого приложения. Каждый статистический метод использует свои собственные функции и статистику. В статье [38] предложили идентифицировать отпечатки протоколов, основываясь на функции плотности вероятности (PDF) времени между поступлениями па-

кетов и нормализованных пороговых значениях. Они достигли точности до 91% для группы протоколов, таких как HTTP, Post Office Protocol 3 (POP3) и Simple Mail Transfer Protocol (SMTP). В аналогичной работе [41] рассматривался формат PDF и размер пакета. Их схема позволила идентифицировать более широкий спектр протоколов, включая протокол передачи файлов (FTP), протокол доступа к интернет-сообщениям (IMAP)., SSH и TELNET с точностью до 87%.

2. **Методы, основанные на машинном обучении:** Для классификации трафика было опубликовано огромное количество подходов к машинному обучению. Например, в [2] предложили байесовскую нейронную сеть, которая была обучена классифицировать наиболее известные протоколы P2P, включая Kazaa, BitTorrent, GnuTella, и достигли точности 99%. Были предложены подходы с применением искусственных нейронных сетей (ANN) для идентификации трафика [39, 36]. Более того, в [36] было показано, что подход ANN может превзойти наивные байесовские методы. Наиболее важной статьей, опубликованной с использованием набора данных "VPN-nonVPN dataset (ISCXVPN2016)", является работа [7]. В ней для обучения были использованы параметры связанные со временем, такие как продолжительность потока, количество байтов потока в секунду, время прямого и обратного взаимного поступления и т.д. Для распознавания сетевого трафика использовали метод k-ближайших соседей (k-NN) и алгоритмы деревьев принятия решений. Они достигли примерно 92% точности, характеризуя шесть основных классов трафика, включая просмотр веб-страниц, электронную почту, чат, потоковую передачу, передачу файлов и VoIP.

Исходя из обширного обзора различных подходов и методов, представленных выше, было принято решение ориентироваться на метод, основанный на алгоритмах с деревьями принятия решений, подобно описанному в исследовании [7]. Этот метод показал хорошую точность

и высокую эффективность в классификации сетевого трафика. Однако, в рассматриваемой задаче фокус будет смещен, как отмечалось в подразделе 2.2, с определения конкретных приложений на идентификацию активности из ранее определенной группы или же конкретного пользователя, соответствующего определенному паттерну.

2.4. Используемые технологии

1. **Python==3.10.10** — интерпретируемый, интерактивный, объектно-ориентированный язык программирования. Он включает в себя модули, динамические типы данных очень высокого уровня и классы.

2. **Scikit-learn==1.3.0** — открытая библиотека машинного обучения для языка программирования Python. Она предоставляет простой и эффективный инструментарий для решения разнообразных задач машинного обучения, включая классификацию, регрессию, кластеризацию, обучение с учителем и без учителя, а также предварительную обработку данных.

3. **Numpy==1.24.3, Pandas==2.0.3** — это ключевые инструменты в экосистеме Python для работы с данными и выполнения вычислений. Они предоставляют мощные функциональности и структуры данных, упрощая анализ и манипуляции с данными.

4. **FastAPI==0.104.0** — это современный, быстрый (высокопроизводительный) веб-фреймворк для создания API с использованием современных практик разработки, который полностью соблюдает стандарты для документации и взаимодействия с API.

3. Архитектура системы

В данном разделе представлено подробное описание компонентов и функциональности системы, разработанной для верификации пользователей на основе анализа их поведения в сети. Этот раздел является ключевым для понимания того, как различные элементы системы взаимодействуют друг с другом для обеспечения надежной идентификации пользователя. Первая часть раздела посвящена распознаванию активности пользователя. Здесь описывается как система анализирует сетевой трафик для выявления конкретных действий пользователя в реальном времени. Во второй части рассматривается распознавание паттерна поведения пользователя. Этот процесс включает анализ долгосрочных поведенческих моделей для создания профиля цифрового почерка пользователя. Система учитывает различные аспекты поведения, такие как частота, время активности и предпочтения в использовании определенных сервисов. Третья часть описывает общую архитектуру всей системы. Она представляет собой комплексное решение, включающее сбор данных, их обработку (обработчики событий) и анализ, а также механизмы реагирования на обнаруженные угрозы.

3.1. Обработка начальных данных

Перед разработкой любой системы с применением искусственного интеллекта критически важно уделить первостепенное внимание обработке и подготовке исходных данных. При работе с данной задачей обычно данные представлены в формате pcap (Packet Capture). Этот формат является широко применяемым стандартом в области сетевых технологий и используется для записи сетевых пакетов. Они могут содержать обширную информацию о сетевых взаимодействиях, включая, но не ограничиваясь, передачей данных между различными IP-адресами пользователей. Поэтому важным этапом является подготовка сырых данных к дальнейшему анализу. В первую очередь необходимо вручную подготовить и структурировать имеющуюся информацию в формат CSV, содержащий такие поля как IP-адрес, имя файла и тип

активности, который был использован во время записи pcap (табл. 1).

IP ADDRESS	FILENAME	ACTIVITY
192.168.202.1	FILE1.PCAP	ACT1
192.168.202.2	FILE1.PCAP	ACT2
192.168.202.3	FILE2.PCAP	ACT5
192.168.202.4	FILE3.PCAP	ACT1
192.168.202.5	FILE3.PCAP	ACT1

Таблица 1: Пример файла csv, после этапа сбора данных от приложения

Далее следует процесс преобразования входных данных, или же Feature Engineering, где выделяются ключевые характеристики, которые будут использоваться для распознавания активности. Этот этап критичен, так как качество и релевантность выделенных характеристик напрямую влияют на способность системы корректно распознавать различные типы активности. В предложенном подходе данный шаг выполняется в три шага:

1. Извлечение 90 признаков на основе tshark статистики с параметрами -z conv,ip, -z conv,tcp, -z conv,udp, -z plen,tree.

2. Удаление всех столбцов, которые никаким образом не способствуют распознаванию активности, а именно либо являются None(Null), либо содержат только 1 значение.

3. На основе корреляционного анализа происходит отсеивание менее значимых признаков, таким образом остаются только те свойства, коэффициент корреляции которых больше наперед заданного threshold в зависимости от сложности и кол-ва данных.

3.2. Распознавание активности пользователя

Распознавание активности пользователя представляет собой ключевой процесс в рамках разработки передовых технологий, нацеленных на реализацию концепции “цифрового почерка” абонента. Начинается

он как было описано в предыдущем разделе 3.1 с обработки исходных данных.

Далее будет полезно обсудить вопрос о группировке данных по активностям за один временной промежуток. Это значительно облегчает понимание характера взаимодействия и упрощает дальнейшую классификацию каждого фрагмента. А необходимо это для того, чтобы иметь возможность распознать все активности происходящие в течение времени, покрываемого данными, а не рассматривать весь входной файл в качестве одного действия. Таким образом, разработана методика разделения pcap файла на сгруппированные временные отрезки, основываясь на пакетах в сетевом трафике:

1. Выделение начальных признаков. Из исходного pcap выделяется четыре основных признака для каждого пакета: frame.len, frame.time_epoch, frame.time_delta, ip.proto

2. Расширение признакового пространства. На основе этих четырех признаков используя различные математические операции получаем 16 более информативных.

3. Выбор алгоритма кластеризации и дальнейшее применение на каждом из пакетов. В зависимости от задачи, могут применяться различные методы кластеризации(DBSCAN, KMeans, AgglomerativeClustering, MeanShift, OPTICS), которые можно настраивать для достижения оптимальной точности и эффективности.

4. Ликвидация выбросов в кластерах. Данный шаг повышает точность анализа, убирая ошибки, которые могли возникнуть из-за недостатков в алгоритмах кластеризации или из-за шума в данных.

5. Формирование временных интервалов. Необходимо найти все последовательности пакетов, имеющие один и тот же кластер, и записать все временные метки, в которых происходит смена значения.

В итоге благодаря алгоритму кластеризации, система может идентифицировать неявные структуры в данных и классифицировать меняющееся кол-во активности без явных меток классов, что может быть полезно в задачах, где нет возможности однозначно определить размерность выходного пространства.

Завершающий шаг в данной цепи заключается в обучении модели и её использовании для распознавания активности абонента. Как было отмечено в разделе 2.3 наиболее эффективным решением для данной задачи является использование модели случайного леса(RandomForest), который обучается на основе ранее подготовленных данных. Процесс обучения также включает в себя автоматическую настройку параметров, дабы максимально повысить точность и уменьшить процент ошибок при распознавании.

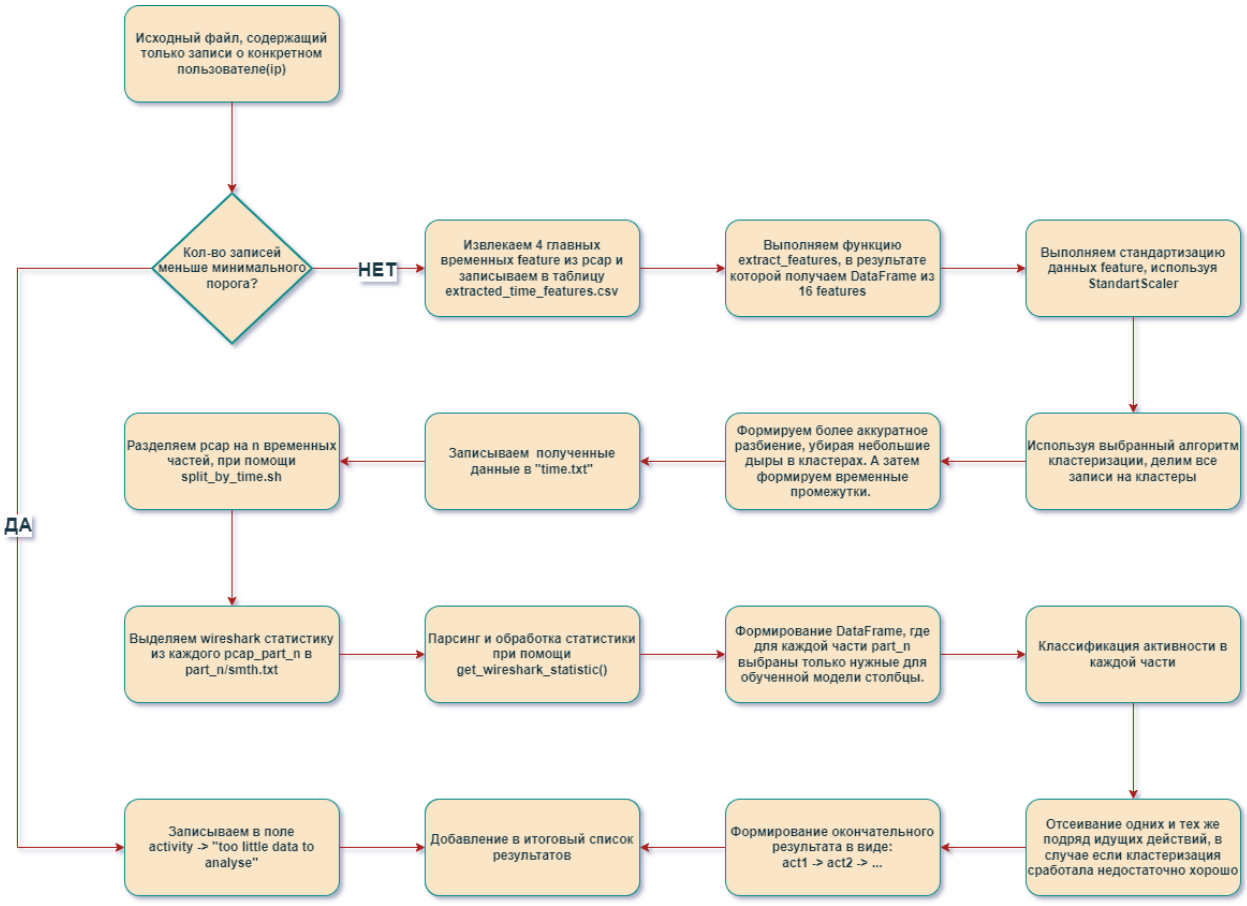

Рис. 1: Схема распознавания активности пользователя

В целом, такая последовательность шагов создает комплексный про-

цесс, который позволяет глубоко анализировать и понимать активности пользователя на основе сетевого трафика. Это сочетание предварительной обработки, анализа и пост обработки данных гарантирует, что результаты являются как точными, так и полезными для понимания поведения пользователя. Более подробная схема представлена на рис. 1.

3.3. Распознавание паттерна поведения пользователя

Распознавание паттерна поведения пользователя — это не менее сложный процесс, требующий анализа больших объемов данных и точной идентификации индивидуальных особенностей. Входные данные для этой задачи аналогично разделу 3.2 представляют собой сетевые пакеты, записанные в формате pcap, которые уже отфильтрованы таким образом, чтобы содержать информацию, относящуюся только к конкретному пользователю.

Перед тем как переходить к дальнейшим шагам реализации важно отметить ключевое отличие распознавания паттерна поведения от активности. Для классификации последовательности действий наилучшим решением является использование только последней записи, так как это позволит системе быстрее и точнее реагировать на текущие действия пользователя. В то же время для классификации пользователя необходимо сохранять гораздо больший объем данных, дабы обеспечить более точную и всестороннюю классификацию, учитывая историю действий пользователя.

Исходя из сказанного выше вытекает следующая проблема: необходимость обработки огромного объема данных, записанных в исходном pcap файле, без учета фильтрации. Для точного идентификации пользователя требуются обширные данные, которые могут охватывать длительные периоды времени и содержать миллионы пакетов, что затрудняет работу программы в реальном времени из-за ограничений производительности и объема загружаемых в оперативную память данных.

Это означает, что необходимо найти способ эффективно обрабатывать большие объемы данных, сохраняя при этом точность идентификации.

Решением этой проблемы является создание специализированного хранилища, в котором информация о каждом конкретном пользователе будет сохраняться в отдельном файле. Такой подход позволит системе постепенно дополнять и уточнять профиль каждого пользователя по мере поступления новых данных. Кроме того, в описываемой системе реализован механизм, который контролирует размер сохраняемого файла, поддерживая временной инвариант(например, 300 секунд). Это позволяет инструменту адаптироваться к ограничениям по памяти и производительности, сохраняя при этом высокую точность идентификации за счет до уточнения паттерна поведения. Подробная схема данного метода представлена ниже(рис. 2).

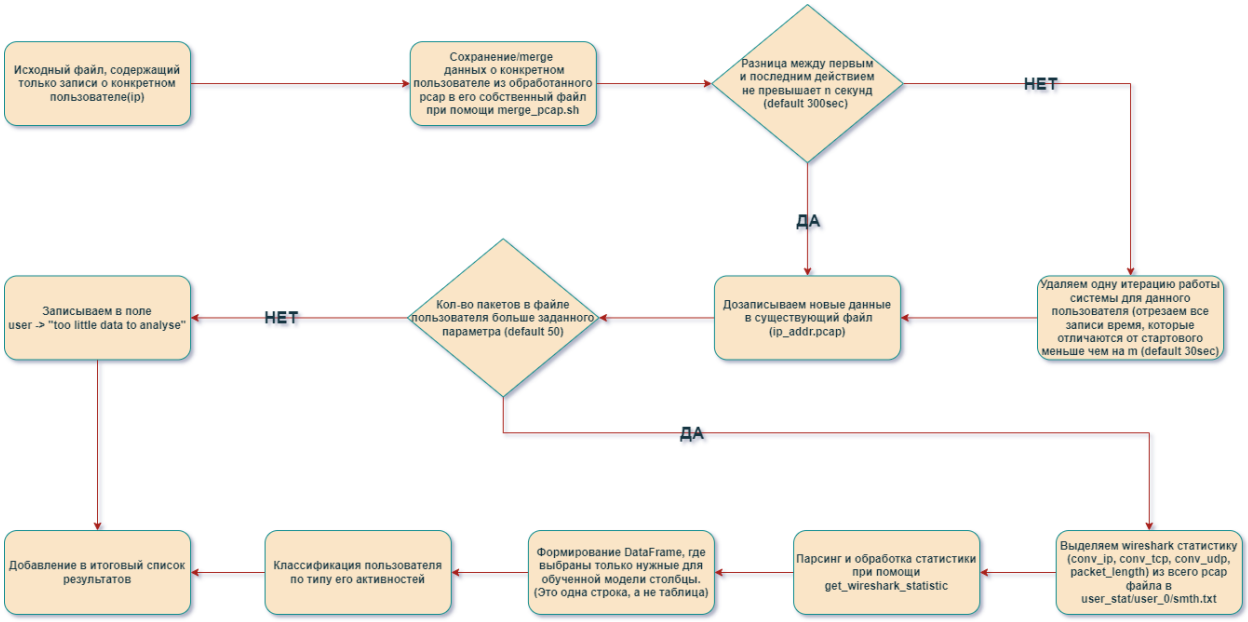

Рис. 2: Схема распознавания паттерна пользователя

Дальнейший этап заключается в обработке полученного объединенного pcap, аналогично распознаванию действий, совершаемых пользователем, как было описано в разделе 3.1. И наконец, финальный шаг в имплементации инструмента для классификации паттерна поведения абонента состоит непосредственно в обучении модели случайного леса (RandomForest). Однако в отличие от предыдущего этапа, целью явля-

ется не определение типа активности, а распознавание самого паттерна.

3.4. Целостная система

В данном разделе подробно рассматривается как общая архитектура, так и основные функциональные аспекты разработанной комплексной системы, предназначенной для реализации “цифрового почерка” сетевого трафика абонента. Ниже представлена детальная схема (рис. 8), а после описание ключевых этапов решения.

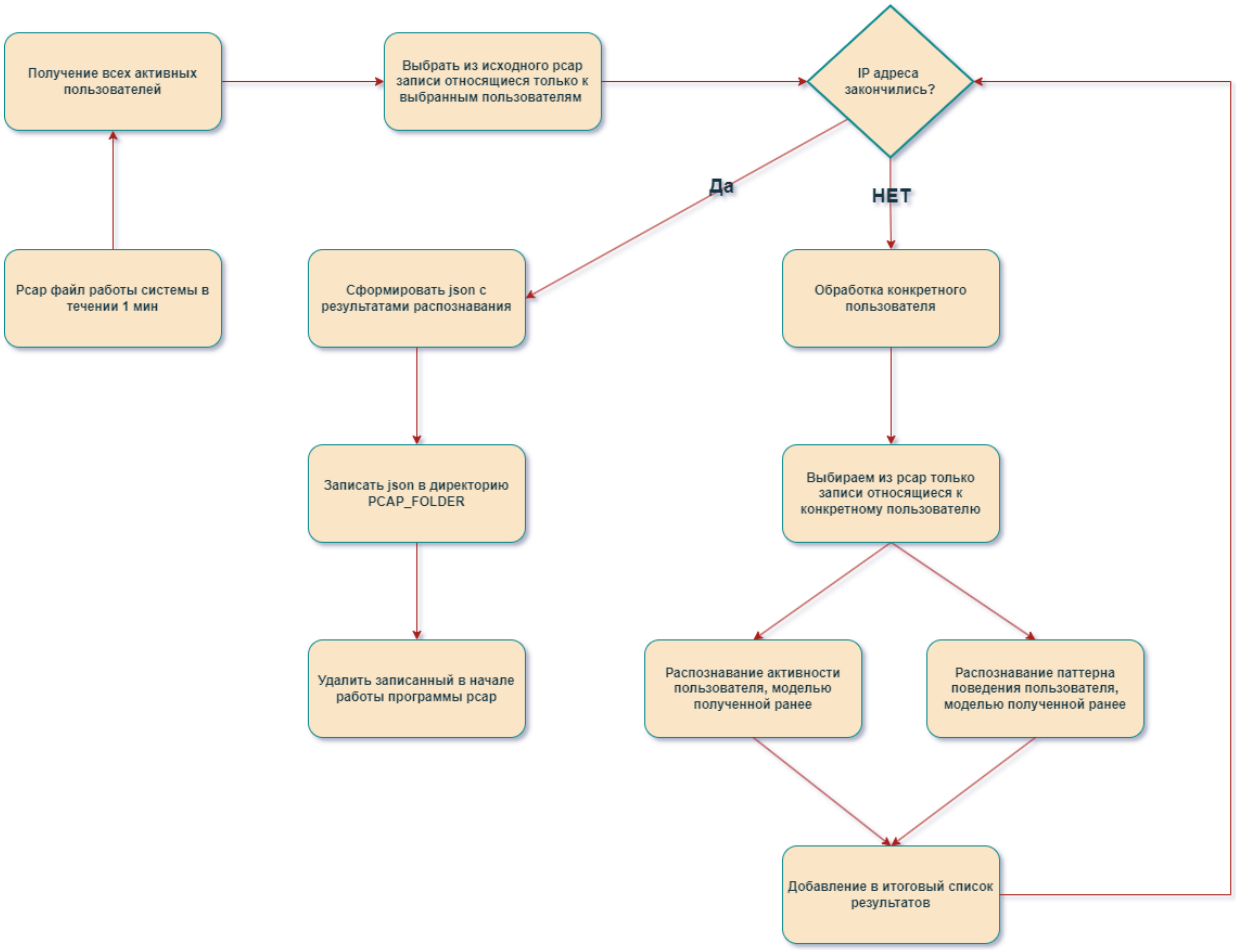

Рис. 3: Общая схема работы разрабатываемой системы

Процесс начинается с получения одного большого pcap файла, который содержит данные о любой активности в приложении за некоторый период. После чего необходимо отобрать только те пакеты, которые относятся к пользователям, взаимодействующим с вашим приложением

в настоящий момент. Это позволяет сфокусировать анализ на интересующих абонентов и уменьшить объем данных для обработки.

После этого производится детальный анализ информации, связанной с каждым пользователем. А именно распознавание активности и паттерна поведения абонента, обученными на предыдущих шагах моделями. В результате позволяя не только определять, что делает пользователь в данный момент, но и сравнивать полученный паттерн с его обычным поведением. Табл. 2, расположенная ниже, является примером того, как может выглядеть результат работы реализованного инструмента.

В итоге, все полученные данные сохраняются в формате json, что позволяет использовать их в дальнейшем в различных приложениях. Так, например, для эффективной интеграции с другими web-инструментами анализа трафика предлагается создать специализированный сервер, к которому будут обращаться анализаторы. На этом сервере будут выполняться все необходимые вычисления, а результаты будут отправляться обратно. Это позволит легко интегрировать систему распознавания с существующей инфраструктурой без значительных изменений в работе заказчика.

IP ADDRESS	ACTIVITIES	USER
192.168.202.1	TOO LITTLE DATA TO ANALYZE	TOO LITTLE DATA TO ANALYZE
192.168.202.2	ACT1 -> ACT2	1
192.168.202.3	ACT2	2
192.168.202.4	ACT2	2
192.168.202.5	ACT1 -> ACT4 -> ACT1	3
192.168.202.6	TOO LITTLE DATA TO ANALYZE	2
192.168.202.7	ACT1 -> ACT3 -> ACT1	3
192.168.202.8	ACT2 -> ACT1	2
192.168.202.9	ACT1	5
192.168.202.10	ACT2	2

Таблица 2: Итоговый результат работы реализованной системы

Таким образом, разработанная система представляет собой мощный инструмент для анализа сетевого трафика, способный обеспечить детальное понимание сетевого поведения абонентов и их идентификацию на основе уникальных характеристик. Однако необходимо подчеркнуть,

что эффективная реализация и интеграция этой системы требует тщательного планирования и учета всех аспектов взаимодействия с другими компонентами вашей инфраструктуры кибербезопасности.

4. Реализация на примере IVA

Корпоративный мессенджер IVA - это инструмент, предназначенный для облегчения общения внутри организаций. Обычно он позволяет пользователям обмениваться сообщениями, фотографиями, видео- и аудиофайлами, а также совершать звонки в чатах. Это приложение разработано специально для улучшения совместной работы и производительности в корпоративной среде, предоставляя безопасный и эффективный способ общения между членами команды.

Использование системы “свой-чужой” на основе “цифрового почерка” при работе с IVA значительно скажется на безопасности и сохранности данных компании. Анализируя сетевой трафик и отслеживая почерк абонентов, организации могут гарантировать конфиденциальность, предотвращая несанкционированный доступ и поддерживая целостность корпоративных обсуждений и обмена данными не только на уровне кибератак из вне, но и постоянно следить за действиями уже проверенных членов компании, чтобы предотвратить доступ злоумышленников через их аккаунты.

4.1. Набор данных

Данные о активностях мессенджера IVA включают информацию, собранную с устройств пользователей во время использования приложения, а также во время видео-конференций. Подробное описание полученных данных представлено в табл. 3. Используемые же для обучения данные в формате pcap были собраны с использованием инструмента tcpdump во время активных сессий непосредственно на сервере IVA, развернутом в частной сети Mobil-Group. Общее кол-во собранных данных равняется примерно 250 действиям, что является достаточным для анализа и дальнейшего обучения модели. В дальнейшем все собранные данные классифицируются и хранятся в защищенной среде для последующего использования.

activity type	description	system name
Отправка сообщений	отправка текстовых сообщений через приложение	message
Звонок	Пакеты, связанные с аудио-звонком	call
Видео-звонок	активность видеосвязи с передачей видеоданных	vcall
Передача файлов	передача файлов между пользователями приложения	file
Видеоконференция без активности	сценарии, когда в приложении отсутствовала активность	dvks
Активная видеоконференция	камера, поднятая рука, голосование	avks

Таблица 3: Собранные данные о видах активности пользователей в приложении IVA

Данные о паттернах поведения пользователей в IVA, аналогично активностям, были собраны с устройств пользователей во время использования приложения, а также во время видео-конференций за период в 10 минут. Подробное описание полученных данных представлено в табл. 4. Общее кол-во собранных данных равняется примерно 200 паттернам, разделенных на 10 категорий. Стоит отметить, что так как на данном этапе был выбран достаточно небольшой промежуток в 10 минут, то 200 паттернов вполне хватит для обучения классификатора, однако для интервала превышающего 60 минут потребуется в разы больше данных.

user id	pattern
1	5 messages -> 2 call 30 sec -> 5 messages
2	4 call 30sec -> 3 messages -> 1 dvks 30sec
3	2 call 20sec -> 2 vcall 20sec -> 2 file -> 1 avks 30sec
4	2 avks 25sec -> 2 messages -> 1 file -> 1 call 45sec
5	1 messages -> 1 call 10sec -> 1 vcall 10sec -> 1 avks 30sec -> 1 file
6	10 call 10sec
7	2 vcall 20sec -> 5 messages -> 1 call 15sec -> 2 messages
8	2 files -> 1 messages -> 3 files
9	3 call 15-20sec -> 3 messages -> 1avks 10sec -> 1 dvks 15sec
10	4 dvks 20sec

Таблица 4: Собранные данные о паттернах поведения пользователей

4.2. Обучение моделей

Согласно схеме описанной в разделе 3.2 был проведен анализ данных, включая построение множества графиков, отражающих картину активности в приложении. Из 90 признаков, извлеченных из pcap

файлов, было отобрано 38 наиболее значимых для дальнейшей работы(рис. 4).

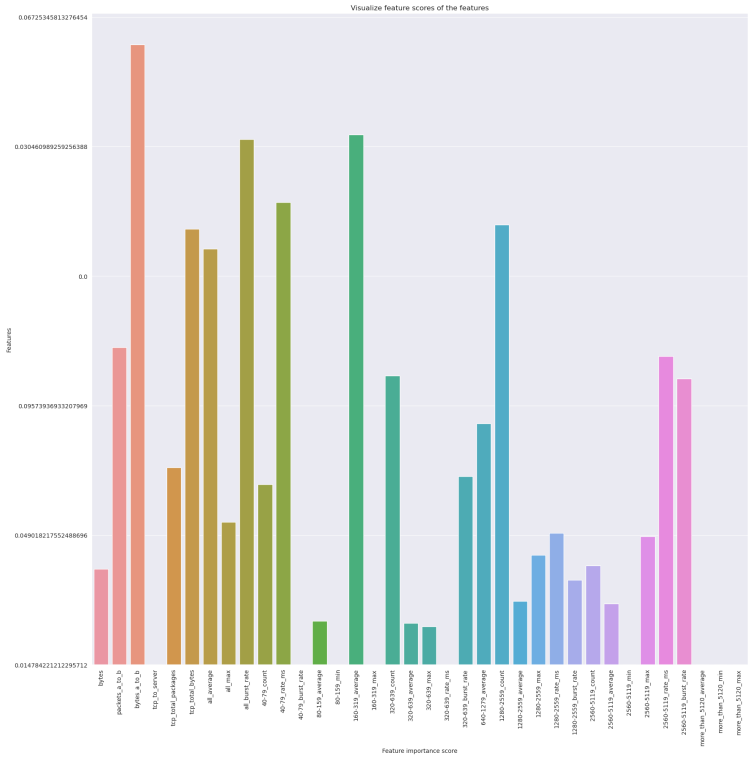

Рис. 4: Наиболее полезные feature при классификации активности

Согласно разделу 3.2 обученный классификатор основывается на модели RandomForest. В результате, после обучения, алгоритм обеспечивает стопроцентную точность (рис. 5) классификации в задаче, где присутствуют 6 типов активности. Также стоит отметить, что обученная модель способна эффективно обрабатывать большие объемы данных, что положительно сказывается на ее использовании в дальнейших частях системы.

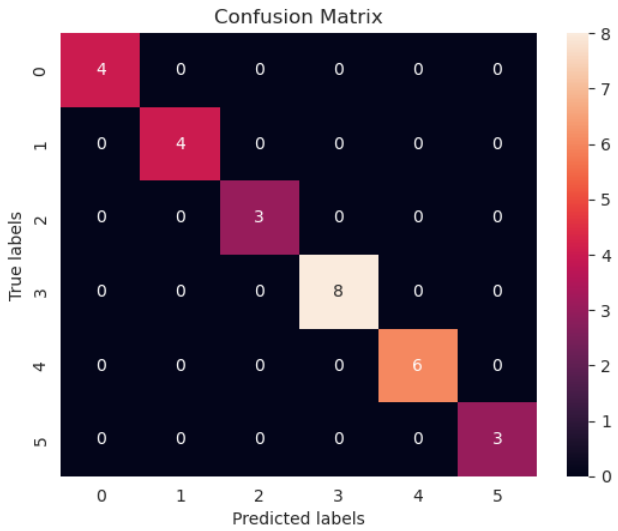

Рис. 5: Результаты обучения классификатора активности

Аналогично распознаванию активности и схеме, описанной в разделе 3.3 был проведен анализ данных, где из 90 признаков, извлеченных из pcap файлов, было отобрано 18 наиболее значимых для дальнейшей работы(рис. 6).

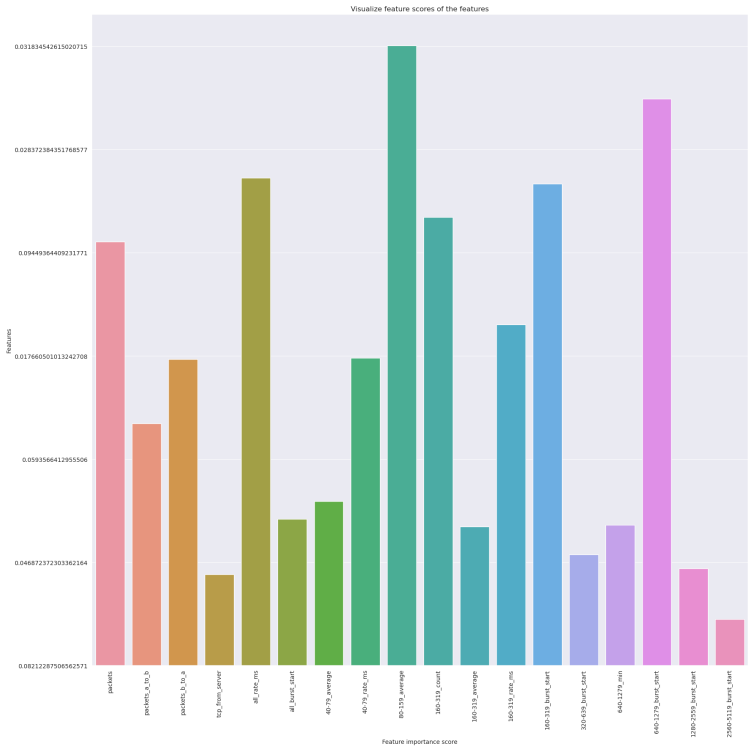

Рис. 6: Наиболее полезные feature при классификации пользователя

В результате чего обученная модель обеспечивает 94% точность (рис. 7) классификации.

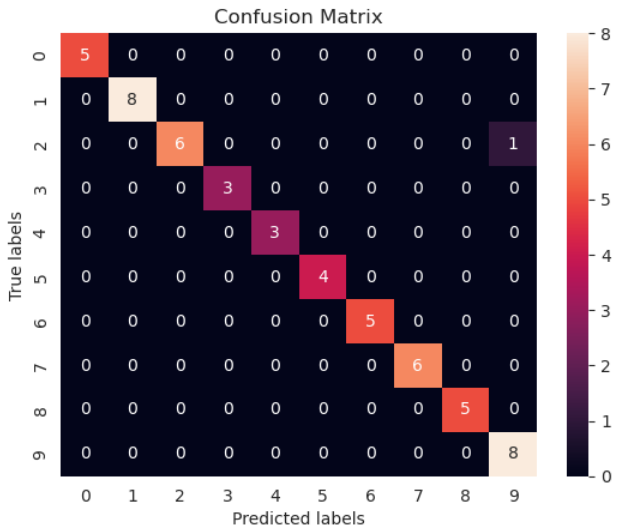

Рис. 7: Результаты обучения классификатора пользовательских паттернов

4.3. Внедрение реализованных моделей в существующие инструменты

Система “свой-чужой” была интегрирована в существующую инфраструктуру Mobil-Group для анализа сетевой активности. Для этого был написан промежуточный сервер на основе FastAPI и websockets, который обеспечивает связь между моделью и приложением. Как показано на схеме ниже(рис. 8) данный этап состоит из нескольких ключевых шагов.

1. **Запуск серверной части.** Это основополагающий шаг, который инициирует работу всей системы, запуская необходимые серверные процессы, обрабатывающие запросы. При помощи служб iva_client.service и iva_server.service запускаются все необходимые фоновые процессы для дальнейшей работы клиентской и серверной части.

2. **Запись серверного трафика.** Все данные о сетевом трафике в течении заранее отведенного временного интервала записываются в pcap файл и передаются дальше для последующего анализа и отладки. При этом систему распознавания пользователей

регулярно проверяет наличие новых файлов для обработки, что позволяет системе оперативно реагировать на поступление новых данных.

3. **Файл с результатами работы системы в формате JSON.** После обработки данных о действиях пользователей результаты работы системы сохраняются в формате JSON.

4. **Инструмент для анализа сети от Mobil-Group.** Далее все полученные данные передаются в специализированный инструмент от Mobil-Group для анализа сетевого трафика, который ранее был подключен к реализованному серверу. При этом постоянное соединение между клиентом и сервером через WebSockets позволяет в режиме реального времени обмениваться всеми необходимыми данными.

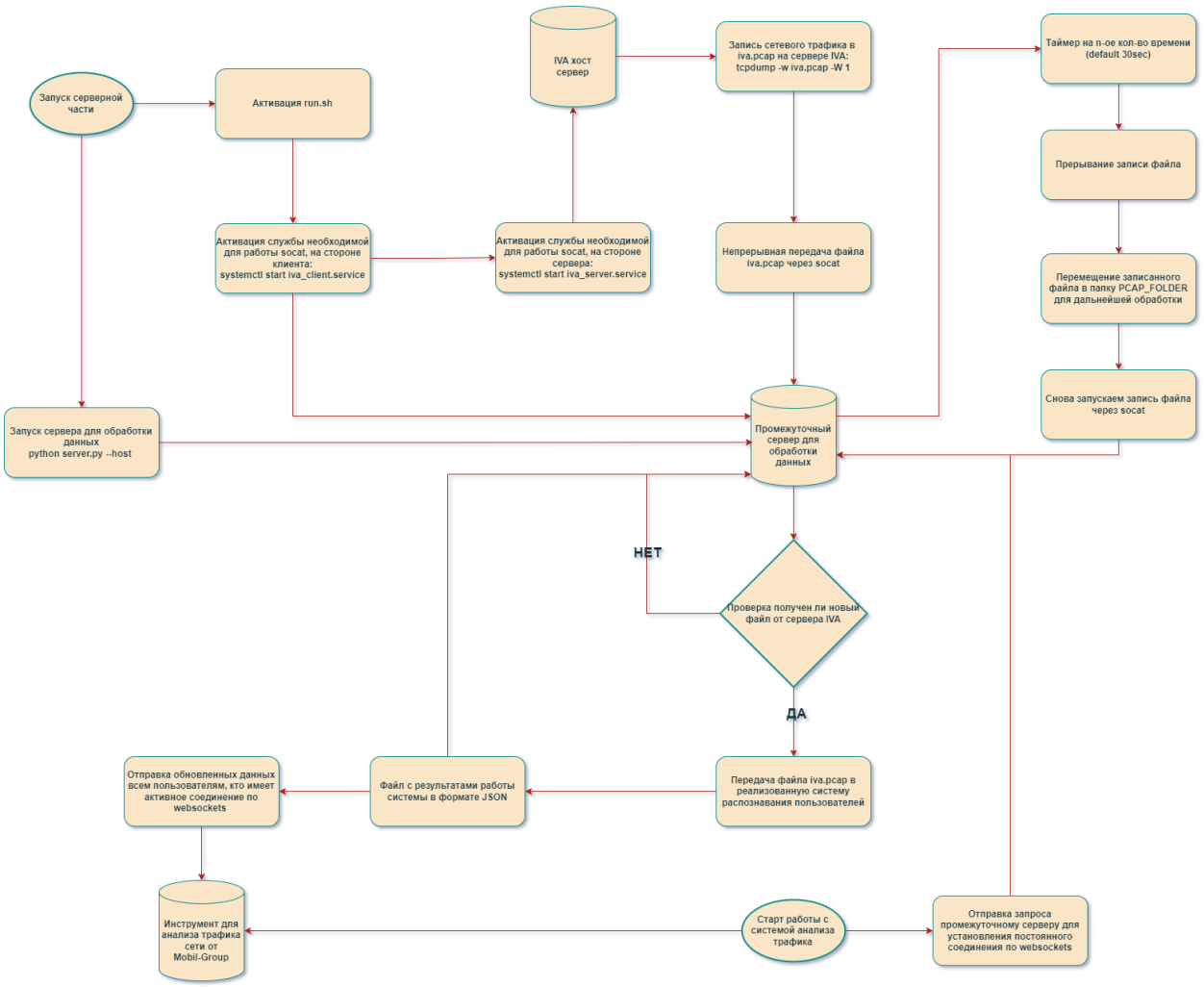

Рис. 8: Общая схема работы разрабатываемой системы

4.4. Пример работы

В реальной жизни система работает следующим образом: когда пользователь начинает сессию в IVA, его поведение в сети анализируется системой “свой-чужой”. Система сравнивает полученные данные с известными паттернами и определяет, соответствует ли активность поведению известного пользователя. Пример того как отображается работа, построенной системы, в инструменте для анализа сетевого трафика на реальных данных(рис. 9).

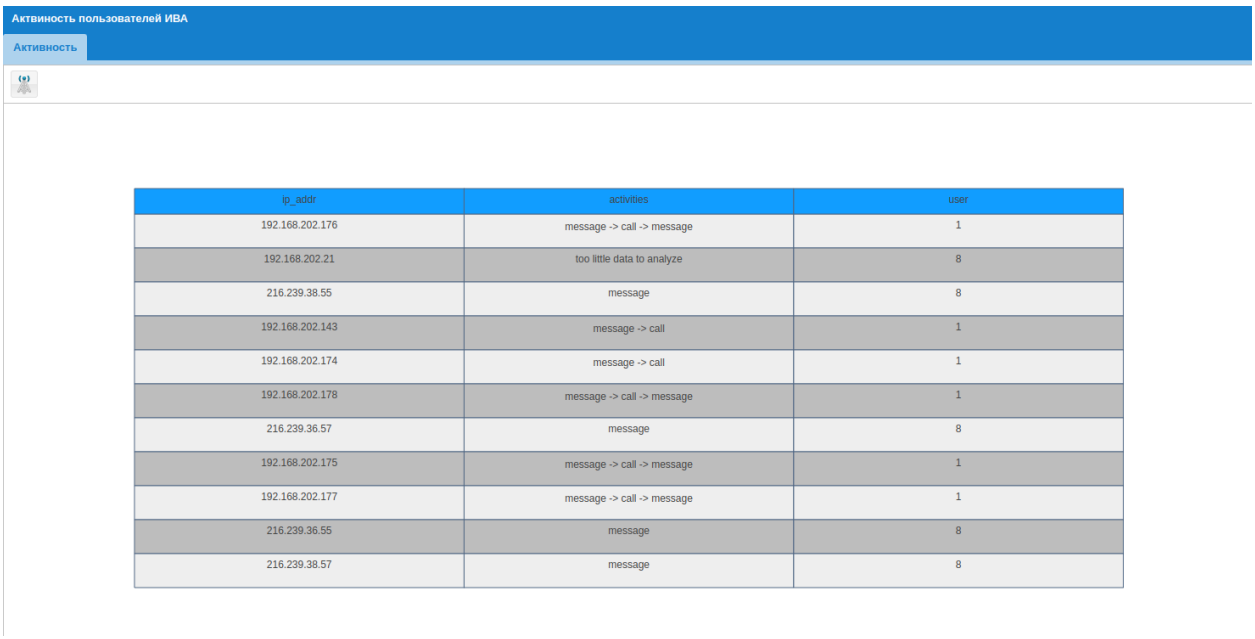

Рис. 9: Интерфейс пользователя при работе с анализатором сетевого трафика от Mobil-Group

Если система обнаруживает отклонения или подозрительную активность, она может предпринять соответствующие действия, например, запросить дополнительную аутентификацию или уведомить администратора системы. Это обеспечивает дополнительный уровень защиты и помогает предотвратить несанкционированный доступ к конфиденциальной информации.

Заключение

Целью данной работы было создание и апробация комплексной системы для формирования “цифровой подписи” пользователя, направленной на повышение уровня безопасности. Поставленная цель была достигнута не полностью. Для реализации проекта были решены следующие задачи:

1. Проанализированы и исследованы передовые методы и подходы к распознаванию поведенческих особенностей пользователей, включая изучение их взаимодействия с устройствами ввода и интерфейсами, а также другие индивидуальные особенности.

2. Разработаны комплексные процедуры для сбора и обработки сетевой информации, охватывающие анализ временных параметров, частоты пользовательских запросов, а также специфических моделей взаимодействия с приложениями.

3. Сконструирована и опробована уникальная архитектура системы на основе машинного обучения, настроенной для гибкой интеграции в разнообразные корпоративные среды, в том числе на практике с использованием мессенджера IVA.

4. Выполнено тщательное обучение моделей для идентификации пользовательской активности, включая всестороннюю подготовку и обработку данных для обучения, а также fine-tuning моделей для достижения высокой точности и надежности при распознавании уникальных цифровых отпечатков.

5. Успешно интегрирован разработанный инструмент в инфраструктуру корпорации Mobil-Group, обеспечивая тем самым улучшенный контроль и анализ сетевой активности.

Исходный код проекта является проприетарным и не подлежит разглашению.

Направления будущей работы. На текущем этапе был создан лишь первичный макет системы, который служит начальной точкой в разработке. Этот макет является концептуальной основой, демонстрирующей основные возможности и подходы. В дальнейшем, необходимо разработать дополнительные макеты для различных приложений, что позволит понять спектр применения и гибкость полученного инструмента в разнообразных сценариях использования. После успешного создания и адаптации нескольких макетов, следующим шагом станет разработка производственного образца, который будет перерастать в полноценную развитую систему.

Ключевым этапом в дальнейшей работе также является более обширное тестирование разработанного инструмента. Это не только поможет выявить и устранить возможные недочеты, но и обеспечит подтверждение эффективности системы в различных условиях и сценариях применения. Такое тестирование будет необходимо для обеспечения надежности, масштабируемости и безопасности системы перед ее внедрением в реальную рабочую среду.

Список литературы

[1] Abbasi Mahmoud, Shahraki Amin, Taherkordi Amir. Deep learning for network traffic monitoring and analysis (NTMA): A survey // Computer Communications. — 2021. — Vol. 170. — P. 19–41.

[2] Auld Tom, Moore Andrew W, Gull Stephen F. Bayesian neural networks for internet traffic classification // IEEE Transactions on neural networks. — 2007. — Vol. 18, no. 1. — P. 223–239.

[3] Awake Security Platform — автономный аналитик сетевых данных и платформа для обеспечения безопасности на основе искусственного интеллекта. [Электронный ресурс] (дата обращения: 07.11.2023) URL: https://www.arista.com/en/solutions/security.

[4] Biometric user identification with dynamic footprint / Jaeseok Yun, Gregory Abowd, Woontack Woo, Jeha Ryu // 2007 Second International Conference on Bio-Inspired Computing: Theories and Applications / IEEE. — 2007. — P. 225–230.

[5] Blindbox: Deep packet inspection over encrypted traffic / Justine Sherry, Chang Lan, Raluca Ada Popa, Sylvia Ratnasamy // Proceedings of the 2015 ACM conference on special interest group on data communication. — 2015. — P. 213–226.

[6] Bricata — автономный аналитик сетевых данных и платформа для обеспечения безопасности на основе искусственного интеллекта. [Электронный ресурс] (дата обращения: 07.11.2023) URL: https://bricata.com.

[7] Characterization of encrypted and vpn traffic using time-related / Gerard Draper-Gil, Arash Habibi Lashkari, Mohammad Saiful Islam Mamun, Ali A Ghorbani // Proceedings of the 2nd international conference on information systems security and privacy (ICISSP). — 2016. — P. 407–414.

[8] Chowdhury Rajarshi Roy, Idris Azam Che, Abas Pg Emeroylariffion. Internet of things: Digital footprints carry a device identity // AIP Conference Proceedings / AIP Publishing. — Vol. 2643. — 2023.

[9] Cisco Stealthwatch — автономный аналитик сетевых данных и платформа для обеспечения безопасности на основе искусственного интеллекта. [Электронный ресурс] (дата обращения: 07.11.2023) URL: https://www.cisco.com/site/us/en/products/security/security-analytics/secure-network-analytics/index.html.

[10] Comparison of machine-learning algorithms for classification of VPN network traffic flow using time-related features / Sikha Bagui, Xingang Fang, Ezhil Kalaimannan et al. // Journal of Cyber Security Technology. — 2017. — Vol. 1, no. 2. — P. 108–126.

[11] Corelight Sensor — автономный аналитик сетевых данных и платформа для обеспечения безопасности на основе искусственного интеллекта. [Электронный ресурс] (дата обращения: 07.11.2023) URL: https://corelight.com/products.

[12] Corvil Security Analytics — автономный аналитик сетевых данных и платформа для обеспечения безопасности на основе искусственного интеллекта. [Электронный ресурс] (дата обращения: 07.11.2023) URL: https://www.pico.net/corvil-analytics/.

[13] Cute: Traffic classification using terms / Soheil Hassas Yeganeh, Milad Eftekhar, Yashar Ganjali et al. // 2012 21st International Conference on Computer Communications and Networks (ICCCN) / IEEE. — 2012. — P. 1–9.

[14] Dainotti Alberto, Pescape Antonio, Claffy Kimberly C. Issues and future directions in traffic classification // IEEE network. — 2012. — Vol. 26, no. 1. — P. 35–40.

[15] Deep packet: A novel approach for encrypted traffic classification using deep learning / Mohammad Lotfollahi, Mahdi Jafari Siavoshani,

Ramin Shirali Hossein Zade, Mohammdsadegh Saberian // Soft Computing. — 2020. — Vol. 24, no. 3. — P. 1999–2012.

[16] Digital footprints: Using WiFi probe and locational data to analyze human mobility trajectories in cities / Martin W Traunmueller, Nicholas Johnson, Awais Malik, Constantine E Kontokosta // Computers, Environment and Urban Systems. — 2018. — Vol. 72. — P. 4–12.

[17] Digital footprints and changing networks during online identity transitions / Oliver L Haimson, Jed R Brubaker, Lynn Dombrowski, Gillian R Hayes // Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. — 2016. — P. 2895–2907.

[18] Henry Colin, Gohdes Anita, Dorff Cassy. Digital footprints and data-security risks for political scientists // PS: Political Science & Politics. — 2022. — Vol. 55, no. 4. — P. 804–808.

[19] Identifying smartphone users based on how they interact with their phones / Mohammed A Alqarni, Sajjad Hussain Chauhdary, Maryam Naseer Malik et al. // Human-centric Computing and Information Sciences. — 2020. — Vol. 10. — P. 1–14.

[20] Identifying smartphone users based on their activity patterns via mobile sensing / Muhammad Ehatisham-ul Haq, Muhammad Awais Azam, Usman Naeem et al. // Procedia computer science. — 2017. — Vol. 113. — P. 202–209.

[21] Joshi Manish, Hadi Theyazn Hassn. A review of network traffic analysis and prediction techniques // arXiv preprint arXiv:1507.05722. — 2015.

[22] Karagiannis Thomas, Papagiannaki Konstantina, Faloutsos Michalis. BLINC: multilevel traffic classification in the dark // Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications. — 2005. — P. 229–240.

[23] Kaspersky Anti Targeted Attack — автономный аналитик сетевых данных и платформа для обеспечения безопасности на основе искусственного интеллекта. [Электронный ресурс] (дата обращения: 07.11.2023) URL: https://www.kaspersky.ru/enterprise-security/anti-targeted-attack-platform.

[24] Kellas-Dicks Mechthild R, Stark Yvonne J. Keystroke dynamics authentication techniques. — 2012. — 11. — US Patent 8,332,932.

[25] Khalife Jawad, Hajjar Amjad, Diaz-Verdejo Jesus. A multilevel taxonomy and requirements for an optimal traffic-classification model // International Journal of Network Management. — 2014. — Vol. 24, no. 2. — P. 101–120.

[26] Madhukar Alok, Williamson Carey. A longitudinal study of P2P traffic classification // 14th IEEE international symposium on modeling, analysis, and simulation / IEEE. — 2006. — P. 179–188.

[27] Mahfouz Ahmed, Mahmoud Tarek M, Eldin Ahmed Sharaf. A survey on behavioral biometric authentication on smartphones // Journal of information security and applications. — 2017. — Vol. 37. — P. 28–37.

[28] Moore Andrew W, Papagiannaki Konstantina. Toward the accurate identification of network applications // International workshop on passive and active network measurement / Springer. — 2005. — P. 41–54.

[29] PT Network Attack Discovery — автономный аналитик сетевых данных и платформа для обеспечения безопасности на основе искусственного интеллекта. [Электронный ресурс] (дата обращения: 07.11.2023) URL: https://www.ptsecurity.com/ru-ru/products/network-attack-discovery/.

[30] Packet classification algorithms: From theory to practice / Yaxuan Qi, Lianghong Xu, Baohua Yang et al. // IEEE INFOCOM 2009 / IEEE. — 2009. — P. 648–656.

[31] Personalized federated learning framework for network traffic anomaly detection / Jiaming Pei, Kaiyang Zhong, Mian Ahmad Jan, Jinhai Li // Computer Networks. — 2022. — Vol. 209. — P. 108906.

[32] Robust smartphone app identification via encrypted network traffic analysis / Vincent F Taylor, Riccardo Spolaor, Mauro Conti, Ivan Martinovic // IEEE Transactions on Information Forensics and Security. — 2017. — Vol. 13, no. 1. — P. 63–78.

[33] Sen Subhabrata, Spatscheck Oliver, Wang Dongmei. Accurate, scalable in-network identification of p2p traffic using application signatures // Proceedings of the 13th international conference on World Wide Web. — 2004. — P. 512–521.

[34] Studying the impact of mood on identifying smartphone users / Khadija Zanna, Sayde King, Tempestt Neal, Shaun Canavan // arXiv preprint arXiv:1906.11960. — 2019.

[35] Threat Detection System — автономный аналитик сетевых данных и платформа для обеспечения безопасности на основе искусственного интеллекта. [Электронный ресурс] (дата обращения: 07.11.2023) URL: https://www.facct.ru/products/threat-intelligence/.

[36] Ting Hu, Yong Wang, Xiaoling Tao. Network traffic classification based on kernel self-organizing maps // 2010 International Conference on Intelligent Computing and Integrated Systems / IEEE. — 2010. — P. 310–314.

[37] Tourists' digital footprint: the spatial patterns and development models of rural tourism flows network in Guilin, China / Xiaohua Qin, XingMing Li, Wei Chen et al. // Asia Pacific Journal of Tourism Research. — 2022. — Vol. 27, no. 12. — P. 1336–1354.

[38] Traffic classification through simple statistical fingerprinting / Manuel Crotti, Maurizio Dusi, Francesco Gringoli, Luca Salgarelli // ACM SIGCOMM Computer Communication Review. — 2007. — Vol. 37, no. 1. — P. 5–16.

[39] Traffic classification using probabilistic neural networks / Runyuan Sun, Bo Yang, Lizhi Peng et al. // 2010 Sixth International Conference on Natural Computation / IEEE. — Vol. 4. — 2010. — P. 1914–1919.

[40] Uhl Andreas, Wild Peter. Footprint-based biometric verification // Journal of Electronic Imaging. — 2008. — Vol. 17, no. 1. — P. 011016–011016.

[41] Wang Xiaoming, Parish David J. Optimised multi-stage tcp traffic classifier based on packet size distributions // 2010 Third International Conference on Communication Theory, Reliability, and Quality of Service / IEEE. — 2010. — P. 98–103.

[42] A survey of methods for encrypted traffic classification and analysis / Petr Velan, Milan Čermák, Pavel Čeleda, Martin Drašar // International Journal of Network Management. — 2015. — Vol. 25, no. 5. — P. 355–374.

[43] A survey of payload-based traffic classification approaches / Michael Finsterbusch, Chris Richter, Eduardo Rocha et al. // IEEE Communications Surveys & Tutorials. — 2013. — Vol. 16, no. 2. — P. 1135–1156.

[44] Гарда Монитор — автономный аналитик сетевых данных и платформа для обеспечения безопасности на основе искусственного интеллекта. [Электронный ресурс] (дата обращения: 07.11.2023) URL: https://gardatech.ru/produkty/monitor/.

[45] Диденко С. М., Шапцев В. А. Методика отображения информационного почерка пользователя // Вестник кибернетики. — 2005. — no. 4. — P. 74–78.

[46] Мазниченко Н. І. БИОМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ ПОЛЬЗОВАТЕЛЕЙ НА ОСНОВЕ ПОВЕДЕНЧЕСКИХ ХАРАКТЕРИСТИК: ОБЗОР И АНАЛИЗ // ГРААЛЬ НАУКИ. — 2021. — no. 8. — P. 183–192.

[47] Рублев Д. П., Федоров В. М. Идентификация пользователя по динамическим характеристикам работы с манипулятором “мышь” с использованием нейронных сетей // Известия Южного федерального университета. Технические науки. — 2017. — no. 5 (190). — P. 67–74.

[48] Трокоз Д.А. Диссертация: Методы машинного обучения для динамической биометрической верификации операторов беспилотных летательных аппаратов // Место защиты: ФГБОУ ВО «Пензенский государственный университет». — 2021. — P. 0–295.

[49] Трокоз Д.А. Методика прямой динамической биометрической верификации // Вестник Ростовского государственного университета путей сообщения. — 2021. — no. 1. — P. 70–79.

[50] Ямченко Ю.В. Методы решения задач аутентификации и идентификации пользователя на основе анализа клавиатурного почерка // Вестник Московского государственного технического университета им. НЭ Баумана. Серия «Приборостроение». — 2020. — no. 1 (130). — P. 124–139.