УДК 004.056

РАЗРАБОТКА СКАНЕРА СКРЫТЫХ КАНАЛОВ В ПРОТОКОЛЕ ПЕРЕДАЧИ ГИПЕРТЕКСТА СИСТЕМ БАНК-КЛИЕНТ ПРИ ПОМОЩИ КЛАССИФИКАТОРОВ МАШИННОГО ОБУЧЕНИЯ

А.Ю. Симачев
Студент группы М22-508 НИЯУ МИФИ, simachev.anton@mail.ru

Аннотация: Представлены результаты разработки сканера для выявления скрытых каналов в протоколе передачи гипертекста систем банк-клиент с использованием классификаторов машинного обучения. В работе рассматриваются архитектура систем банк-клиент, анализируются актуальные сетевые протоколы и выделяются основные признаки наличия скрытых каналов. Также представлены методы машинного обучения, применяемые для классификации трафика. Проведен анализ результативности разработанного сканера, подтверждающий его эффективность в условиях реального сетевого трафика.

Ключевые слова: скрытые каналы, протокол HTTP/3, системы банк-клиент, машинное обучение, классификаторы, информационная безопасность.

Введение

С увеличением числа систем банк-клиент возросла необходимость в обеспечении их безопасности. Одной из основных угроз является наличие скрытых каналов, которые позволяют злоумышленникам передавать данные незаметно для стандартных систем защиты.

Системы банк-клиент стали неотъемлемой частью современного банковского дела. Они предоставляют удобный и оперативный доступ к банковским услугам через интернет, что существенно упрощает работу клиентов с банком. Однако с ростом их популярности возрастает и количество угроз безопасности. Безопасность систем банк-клиент является критически важной задачей, поскольку они обрабатывают чувствительные данные, такие как личные данные клиентов и финансовую информацию. Уязвимости в этих системах могут привести к утечке данных, финансовым потерям и подрыву доверия клиентов. Одной из главных угроз безопасности являются сквозные атаки, при которых злоумышленники проникают в систему и передают данные через скрытые каналы, оставаясь незамеченными для стандартных средств защиты. В контексте систем банк-клиент, это может означать передачу данных, таких как пароли и личная информация, через законные на первый взгляд HTTP/3-запросы. HTTP/3 является последней версией протокола передачи гипертекста, которая использует протокол QUIC для улучшения скорости и безопасности передачи данных. Он предоставляет ряд преимуществ, таких как улучшенная производи-

тельность и уменьшенная задержка, что делает его привлекательным для использования в системах банк-клиент. Несмотря на свои преимущества, HTTP/3 также имеет уязвимости, которые могут быть использованы злоумышленниками для создания скрытых каналов передачи данных. Это связано с его сложной структурой и новизной, что делает его менее изученным по сравнению с предыдущими версиями протокола.

Для эффективного обнаружения скрытых каналов в HTTP/3 протоколе необходимо использовать современные методы анализа данных, такие как машинное обучение. Классификаторы машинного обучения могут быть обучены на выявление аномалий и необычных паттернов в сетевом трафике. На первом этапе необходимо собрать большой объем данных сетевого трафика, включающего как легитимные запросы, так и запросы с использованием скрытых каналов. Затем данные анализируются для выявления характеристик, которые могут указывать на наличие скрытых каналов. На основе полученных характеристик создается и обучается модель машинного обучения, способная классифицировать запросы. Разработанная модель проходит тестирование на различных наборах данных для проверки ее точности и надежности. Использование машинного обучения для выявления скрытых каналов в протоколе HTTP/3 позволяет значительно повысить точность и скорость обнаружения потенциальных угроз, что способствует улучшению общей безопасности систем банк-клиент.

Цель данной работы – создать сканер для выявления скрытых каналов в протоколе передачи гипертекста (HTTP/3) систем банк-клиент с использованием классификаторов машинного обучения.

Архитектура систем банк-клиент

Пользователь взаимодействует с системой банк-клиент с помощью веб-браузера, который является универсальным инструментом доступа к интернет-ресурсам. Веб-браузер позволяет пользователю вводить URL-адреса, переходить по ссылкам и загружать веб-страницы, предоставляющие интерфейс для работы с банковскими услугами. При использовании системы банк-клиент пользователь отправляет различные типы запросов на сервер банка, в том числе запросы на просмотр баланса счета и выполнение переводов.

Актуальные сетевые протоколы систем банк-клиент

Системы банк-клиент полагаются на различные сетевые протоколы для передачи данных и обеспечения надежного и безопасного взаимодействия между пользователями и банковскими серверами. Эти протоколы выполняют

различные функции, начиная от разрешения доменных имен до маршрутизации пакетов данных через интернет. Основные из них – это HTTP/3, DNS, BGP, OSPF и RIP. HTTP/3 является новейшей версией протокола передачи гипертекста, предназначенной для ускорения и повышения безопасности интернет-коммуникаций [1]. Он использует протокол QUIC, который работает поверх UDP (User Datagram Protocol). Благодаря использованию QUIC, HTTP/3 значительно уменьшает задержки при установке соединений и передаче данных. Протокол включает в себя встроенное шифрование, что делает коммуникации более защищенными по сравнению с предыдущими версиями HTTP. Также он обеспечивает более стабильное соединение, особенно в условиях нестабильных сетей, за счет встроенных механизмов восстановления утраченных пакетов. DNS (Domain Name System) – это протокол, который используется для преобразования доменных имен в IP-адреса. Он играет ключевую роль в интернете, обеспечивая пользователям возможность обращаться к веб-сайтам и другим ресурсам по удобным для восприятия именам, таким как example.com. Когда пользователь вводит URL адрес банка в браузере, DNS преобразует этот адрес в соответствующий IP-адрес сервера банка. Системы банк-клиент могут использовать расширения DNS, такие как DNSSEC, для обеспечения целостности и аутентичности DNS-запросов. BGP (Border Gateway Protocol) – это протокол динамической маршрутизации, который используется для обмена маршрутизируемой информацией между автономными системами в интернете. BGP помогает находить наиболее эффективные маршруты для передачи данных, что улучшает скорость и надежность соединений. Протокол также позволяет быстро перенастраивать маршруты в случае сбоев, что минимизирует перерывы в обслуживании. OSPF (Open Shortest Path First) и RIP (Routing Information Protocol) – это протоколы динамической маршрутизации, используемые для обмена маршрутной информацией внутри одной автономной системы. OSPF использует алгоритм на основе состояния канала для определения наикратчайшего пути и имеет высокую скорость конвергенции, быстро адаптируясь к изменениям в сети.

Скрытые каналы HTTP/3 и признаки их наличия

Для выявления скрытых каналов необходимо анализировать признаки, которые могут свидетельствовать о передаче данных через скрытые каналы. Основные признаки включают необычную частоту или последовательность определённых типов кадров, непропорционально большое количество потоков или фреймов, а также нестандартные значения параметров в кадрах SETTINGS. Необычная частота или последовательность определённых типов кадров может

указывать на использование скрытых каналов, так как злоумышленники могут изменять типы кадров или их последовательность для маскировки данных. Например, последовательности кадров, которые не соответствуют обычному паттерну трафика, могут быть признаком передачи данных через скрытые каналы. Непропорционально большое количество потоков или фреймов также может свидетельствовать о наличии скрытых каналов. Злоумышленники могут создавать большое количество дополнительных потоков или фреймов для разделения и передачи данных небольшими частями, что затрудняет их обнаружение стандартными методами анализа трафика. Анализ количества и распределения потоков и фреймов может помочь выявить такие аномалии. Нестандартные значения параметров в кадрах SETTINGS являются ещё одним признаком, который может указывать на скрытые каналы. Кадры SETTINGS используются для настройки параметров соединения между клиентом и сервером [2]. Злоумышленники могут изменять стандартные значения параметров в этих кадрах для передачи дополнительных данных. Анализ этих значений и выявление отклонений от нормальных параметров может помочь в обнаружении скрытых каналов. Для анализа данных и выявления этих признаков можно использовать методы машинного обучения. Классификаторы машинного обучения, обученные на наборах данных, содержащих примеры как легитимного трафика, так и трафика с использованием скрытых каналов, могут эффективно различать нормальные и аномальные паттерны. Обучение моделей включает сбор и предварительную обработку данных, выбор признаков, обучение и валидацию моделей. Использование машинного обучения позволяет автоматизировать процесс анализа и повысить точность обнаружения скрытых каналов. Выявление скрытых каналов в протоколах передачи данных, таких как HTTP/3, требует анализа различных признаков аномального поведения трафика. Основные признаки включают необычную частоту или последовательность кадров, непропорционально большое количество потоков или фреймов, и нестандартные значения параметров в кадрах SETTINGS. Применение методов машинного обучения позволяет автоматизировать и улучшить процесс обнаружения скрытых каналов, что способствует повышению безопасности систем банк-клиент.

Классификаторы машинного обучения

Для классификации трафика и выявления скрытых каналов используются различные методы машинного обучения. В данной работе рассматриваются следующие классификаторы: наивный байесовский классификатор, классификатор LSTM, Random Forest, Feedforward Neural Network и Support Vector Machine (SVM) [3].

Наивный байесовский классификатор – это простой и эффективный метод для начального анализа данных. Он основывается на применении теоремы Байеса с предположением о независимости признаков. Этот классификатор быстро обучается и работает, что делает его хорошим выбором для предварительного анализа больших объемов данных.

Классификатор LSTM (Long Short-Term Memory) представляет собой рекуррентную нейронную сеть, способную учитывать временные зависимости в данных. LSTM используется для анализа последовательностей данных и может эффективно выявлять скрытые паттерны во временных рядах, что делает его подходящим для задач анализа сетевого трафика, где временные зависимости играют важную роль.

Random Forest – это ансамблевый метод, основанный на использовании множества деревьев решений. Каждый отдельный классификатор (дерево) обучается на случайном подмножестве данных, а финальное решение принимается на основе голосования всех деревьев. Этот метод отличается высокой точностью и устойчивостью к переобучению, что делает его эффективным инструментом для анализа сложных данных.

Feedforward Neural Network представляет собой многослойную нейронную сеть, способную выявлять сложные зависимости в данных. Такие сети состоят из нескольких слоев нейронов, где каждый слой обучается выделять определенные признаки из данных. Feedforward Neural Network может обучаться на больших объемах данных и выявлять сложные взаимосвязи, что делает её мощным инструментом для классификации трафика.

Support Vector Machine (SVM) – это метод, использующий гиперплоскости для разделения классов данных. SVM находит оптимальную гиперплоскость, которая максимально разделяет классы в многомерном пространстве признаков. Этот метод эффективен для задач классификации, где необходимо найти четкое разделение между классами, и он хорошо работает с линейно и нелинейно разделимыми данными.

Применение этих методов в совокупности позволяет эффективно классифицировать трафик и выявлять скрытые каналы. Наивный байесовский классификатор обеспечивает быструю и простую начальную оценку данных. Классификатор LSTM учитывает временные зависимости и выявляет паттерны в последовательностях данных. Random Forest объединяет результаты множества деревьев решений, обеспечивая высокую точность и устойчивость. Feedforward Neural Network выявляет сложные зависимости и адаптируется к особенностям данных. SVM использует гиперплоскости для точного разделения классов, что позволяет четко классифицировать трафик [4]. Все эти методы

вместе позволяют создавать комплексные системы анализа, способные обнаруживать скрытые каналы в сетевом трафике с высокой точностью.

Построение модулей сканера скрытых каналов

Сканер скрытых каналов разработан с учётом функциональных и программных требований, а также ресурсов, необходимых для его работы. Архитектура сканера включает следующие модули: модуль захвата, модуль анализа, модуль классификации и модуль представления [5].

Модуль захвата отвечает за сбор сетевого трафика. Этот модуль постоянно мониторит сеть, перехватывая и записывая весь входящий и исходящий трафик. Для этого могут использоваться различные инструменты и технологии, такие как пакеты захвата на уровне сетевого интерфейса или использование специализированных устройств для мониторинга трафика. Основная цель модуля захвата – обеспечить полное и точное представление о сетевом трафике для последующего анализа [6].

Модуль анализа выполняет предварительную обработку данных и выявляет подозрительные признаки. Этот модуль очищает и нормализует захваченные данные, устраняя шум и некорректные записи. После этого он анализирует данные на наличие аномалий и признаков скрытых каналов, таких как необычная частота или последовательность кадров, непропорционально большое количество потоков или фреймов, а также нестандартные значения параметров в кадрах SETTINGS. Выявленные подозрительные признаки передаются в следующий модуль для более детального анализа.

Модуль классификации использует методы машинного обучения для определения наличия скрытых каналов. Этот модуль применяет различные классификаторы, такие как наивный байесовский классификатор, классификатор LSTM, Random Forest, Feedforward Neural Network и Support Vector Machine (SVM), для анализа предварительно обработанных данных и выявления скрытых каналов. Каждый классификатор обучен на большом наборе данных, включающем как легитимный трафик, так и примеры скрытых каналов, что позволяет им эффективно различать нормальные и аномальные паттерны.

Модуль представления визуализирует результаты анализа и предоставляет отчёты. Этот модуль отображает результаты работы сканера в удобной для восприятия форме, предоставляя пользователю графики, таблицы и отчёты о выявленных скрытых каналах. Визуализация включает информацию о типах и частоте подозрительных кадров, идентифицированных потоках и фреймах, а также о значениях параметров, которые были определены как

аномальные. Пользователь может использовать эти отчёты для дальнейшего расследования и принятия мер по устранению угроз.

В совокупности, эти модули обеспечивают комплексный подход к выявлению скрытых каналов в сетевом трафике. Модуль захвата гарантирует полный сбор данных, модуль анализа выявляет подозрительные признаки, модуль классификации точно определяет наличие скрытых каналов, а модуль представления предоставляет результаты в удобной и понятной форме. Такая архитектура позволяет эффективно обнаруживать и предотвращать использование скрытых каналов, улучшая общую безопасность сетевых систем.

Анализ результативности сканера

Анализ результативности сканера показал его высокую эффективность в условиях реального сетевого трафика. Использование различных классификаторов машинного обучения позволило достичь высокой точности в выявлении скрытых каналов. Экспериментальные данные подтверждают адекватность разработанной модели и её применимость в реальных системах банк-клиент. В ходе испытаний сканер был протестирован на различных наборах данных, имитирующих реальный сетевой трафик, включая трафик с внедрёнными скрытыми каналами. Полученные результаты показали, что сканер успешно идентифицирует скрытые каналы с высокой точностью, минимизируя количество ложных срабатываний. Экспериментальные данные подтвердили, что разработанная модель способна адаптироваться к изменениям в сетевом трафике и эффективно выявлять скрытые каналы в различных условиях. Это делает сканер применимым в реальных системах банк-клиент, где безопасность и точность обнаружения угроз являются критически важными. Высокая результативность сканера в реальных условиях свидетельствует о его потенциале для широкого использования в банковских и других критически важных системах, требующих защиты от скрытых угроз.

Заключение

В рамках данного исследования рассматривалась возможность разработки сканера скрытых каналов в протоколе передачи гипертекста систем банк-клиент. Использование методов машинного обучения позволило значительно повысить эффективность выявления скрытых каналов. В ходе разработки была создана архитектура, включающая модуль захвата для сбора сетевого трафика, модуль анализа для предварительной обработки данных и выявления подозрительных признаков, модуль классификации для опреде-

ления наличия скрытых каналов с помощью различных классификаторов машинного обучения, и модуль представления для визуализации результатов анализа и предоставления отчётов. Экспериментальные данные показали высокую точность работы сканера в реальных условиях, подтверждая его адекватность и применимость в системах банк-клиент. Использование различных классификаторов, таких как наивный байесовский классификатор, LSTM, Random Forest, Feedforward Neural Network и SVM, обеспечило комплексный подход к анализу трафика и выявлению скрытых каналов. Каждый из этих методов внёс свой вклад в общую точность и надёжность системы, что позволило минимизировать количество ложных срабатываний и улучшить выявление скрытых угроз. Будущие исследования будут направлены на дальнейшую оптимизацию алгоритмов, чтобы повысить их производительность и эффективность. Планируется изучение дополнительных факторов, влияющих на безопасность систем банк-клиент, таких как новые виды скрытых каналов, изменение паттернов трафика и развитие технологий шифрования. Также будет уделено внимание адаптации сканера к изменениям в сетевой инфраструктуре и улучшению методов обработки и анализа данных для повышения общей безопасности и устойчивости банковских систем.

Список литературы

1. Hypertext Transfer Protocol Version 2 (HTTP/2): RFC 7540. 2015.
2. BrainJS. JavaScript naive Bayesian classifier. [Электронный ресурс]. https://github.com/BrainJS/classifier/ (Дата обращения: 06.08.2024).
3. Draft-ietf-quic-invariants. [Электронный ресурс]. https://www.rfc-editor.org/auth48/C430 (Дата обращения: 06.08.2024).
4. Vladimir Mladenovic. Evaluation of HTTP/3 Protocol for Internet of Things and Fog Computing Scenarios. [Электронный ресурс]. https://www.researchgate.net/publication/ 354965913_Evaluation_of_HTTP3_Protocol_for_Internet_of_Things_and_Fog_Computing_ Scenarios. (Дата обращения: 06.08.2024).
5. Епишкина А.В., Когос К.Г. Об оценке пропускной способности скрытых информационных каналов, основанных на изменении длин передаваемых пакетов // Информатика, вычислительная техника и управление, 2015, с. 78-82.
6. Симачев А.Ю. Оценка скрытых каналов по памяти в протоколе QUIC и методы их обнаружения // Научно-исследовательский журнал для студентов и преподавателей «StudNet», 2022, Т. 5, № 5, с. 3644-3652.