Министерство образования Республики Беларусь

Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК [004.42:621.395]:004.032.26

Полудворянин
Сергей Михайлович

Модели и алгоритмы программного средства авторизации звонков в IP-телефонии на основе нейронной сети

АВТОРЕФЕРАТ

на соискание академической степени
магистра технических наук

по специальности 1-40 80 05 - Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

Научный руководитель
Нестеренков С.Н.
кандидат технических наук,
доцент

Минск 2020

КРАТКОЕ ВВЕДЕНИЕ

Время аналоговой телефонии (PSTN) подходит к концу. Привычные телефонные линии переводят на работу по сети Интернет. В ближайшем будущем неизбежен полный переход аналоговых телефонных линий на IP-телефонию, а корпоративных АТС на протокол SIP. Согласно аналитическому отчету компании Research and Markets глобальный рынок IP-телефонии вырастет с $6.88 миллиардов в 2017 году до $12.70 миллиардов в 2023 году. Однако вместе с распространением технологии мы все чаще слышим сообщения о взломах и кибер-атаках на телефонные сервисы.

Открытые текстовые протоколы предоставляют всю информацию любому, кто способен перехватывать сетевой трафик. Анализирую полученные данные можно получить доступ для конфиденциальной информации. Возьмем к примеру процесс аутентификации. В SIP-протоколе пароли кодируются с помощью алгоритма MD5, который уже давно считается небезопасным. Множество телекоммуникационного оборудования, а также программные решения, поставляются с известными паролями по умолчанию. Если эти пароли остаются без изменений, злоумышленники могут легко получить доступ к этому оборудованию. Полученные данные аутентификации могут быть использованы для подмены легальной регистрации учетных записей на серверах VoIP провайдера. Полученный доступ к аккаунту может быть использован для совершения огромного количества звонков на специальные дорогие номера, а также рассылки аудио спама. Суммы, которые придется заплатить владельцу скомпрометированного аккаунта, могут быть весьма и весьма значительными.

Системы выявления мошенничества могут быть неудобны в использовании. Многие из них имеют высокий уровень ложных срабатываний, особенно когда нелегитимный трафик составляет небольшой процент от всего трафика, поэтому вероятность разозлить клиента при ложном срабатывании может быть намного выше, чем вероятность выявления мошенничества. Более сложные системы обещают большую точность, но бесполезны для обнаружения в реальном времени для сколько нибудь крупных поставщик услуг. Наконец, мошенничество и легитимное поведение постоянно меняются, поэтому системы, которые не могут развиваться или «учиться», быстро устаревают. Таким образом, существует потребность в инструментах для разработки точных и удобных систем, которые могли бы применяться для обнаружения мошенничества в IP-телефонии, которые способны к масштабированию и могут адаптироваться к изменениям в поведении как легитимных клиентов, так и мошенников.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является анализ существующих алгоритмов и программных средств выявления нелегитимного доступа, применяемых в области IP-телефонии, и на основе проведенного анализа, разработка моделей, алгоритмов и программных средств системы обработки данных с помощью нейронной сети.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести анализ существующих моделей информационных систем авторизации звонков в IP-телефонии.

2. Разработать модели и алгоритмы системы для анализа данных и авторизации звонков в IP-телефонии.

3. Разработать программные средства выявления нелегитимных звонков в IP-телефонии на основе нейронных сетей.

Объектом исследования является безопасность звонков в IP-телефонии на основе нейронных сетей.

Предметом исследования являются модели и алгоритмы программного средства авторизации звонков в IP-телефонии на основе нейронной сети.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность анализа данных VoIP сервисов и биллинговых компаний, предоставляющих услуги в области IP-телефонии. Структура данных этих сервисов хранит множество параметров звонков, на основе которых можно в режиме реального времени выявлять аномальный трафик пользователей и помечать его как нелегитимный. Оперативный и многофункциональный анализ больших объемов данных расширит функциональные возможности систем борьбы с мошенничеством.

Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики

Работа выполнялась в соответствии с научно-техническим заданием и планом работ кафедры «Программное обеспечение информационных технологий» по теме «Разработка моделей, методов, алгоритмов, повышающих показатели проектирования, внедрения и эксплуатации программных средств для перспективных платформ обработки информации, решения интеллектуальных задач, работы с большими массивами данных и внедрение в совре-

менные обучающие комплексы» (ГБ № 16-2004, № ГР 20163588, научный руководитель НИР – Н. В. Лапицкая).

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя С. Н. Нестеренкова, заключается в формулировке целей и задач исследования.

Опубликованность результатов диссертации

По теме диссертации опубликовано 2 тезиса в сборнике трудов и материалов конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе произведен детальный анализ предметной области – какие на данный момент существуют походы к авторизации трафика в телекоммуникационных системах, какие модели и алгоритмы они используют. Также были рассмотрены типовые задачи, решаемые с помощью машинного обучения, современные платформы и фреймворки. Вторая глава содержит описание математической модели исследования и подготовку данных. В третьей главе рассмотрены подходы выявления аномалий в наборе данных и предложена практическая реализация ПО для классификации нелегитимных звонков. Представлены результаты экспериментальных исследований. В заключении подводится краткий итог проделанной работы и полученных результатов.

Общий объем работы составляет 83 страницы, из которых основного текста – 60 страниц, 27 рисунков на 15 страницах, 8 таблиц на 8 страницах, 19 формул, 33 источника на 3 страницах и 2 приложения на 20 страницах.

КРАТКОЕ СОДЕРЖАНИЕ

Во введении рассмотрено современное состояние сервисов IP-телефонии, а также проблема мошенничества в этой области, которая приводит к необходимости ее рассмотрения.

В первой главе приведен анализ подходов к обнаружению мошенничества в сфере телекоммуникаций, а также указаны их недостатки в плане производительности и удобства использования. Помимо этого, также сформулирована и корректно поставлена задача, дан обзор ее предметной области, подробно освещены моменты, с которыми можно столкнуться при решении данной задачи, а именно: адаптация системы к изменениям в поведении клиентов и мошенников, способность к масштабированию, точность классификации. Также были рассмотрены современные платформы и фреймворки машинного обучения.

Во второй главе рассмотрена математическая модель исследования, а также этап анализа и подготовки данных, включающий в себя выделение признаков для анализа, корреляционный анализ и нормализацию данных. Для исследования использовались данные генерируемые телекоммуникационным оборудованием при совершении звонков (Call Detail Record).

В третьей главе разработан подход поиска аномалий в наборе данных. Полученные аномалии были визуализированы в двумерном пространстве с помощью алгоритма t-SNE. Приведена реализация программного средства классификации звонков на легитимные и мошеннические, рассмотрена его архитектурная часть и параметры. Дана оценка полученным результатам исследования и реализации программного средства.

В заключении приведены основные достигнутые результаты, возникшие в ходе исследования трудности и возможные будущие шаги.

ЗАКЛЮЧЕНИЕ

В ходе исследовательской работы был проведен анализ предметной области: проанализированы подходы к обнаружению мошенничества в сфере телекоммуникаций, описаны проблемы, связанные с созданием инструментов для выявления мошенничества, а также свойства и возможности, которыми должны обладать такие инструменты. Были рассмотрены современные платформы и фреймворки машинного обучения.

Для достижения поставленной цели исследования мы реализовали собственный метод борьбы с мошенничеством с использованием нейронной сети. Были собраны и проанализированы данные, генерируемые телекоммуникационным оборудованием при совершении звонков (Call Detail Record). С помощью нескольких методов были выявлены аномальные звонки. Размеченные данные были использованы для того, чтобы обучить нейронную сеть классифицировать объекты по двум классам: мошеннические звонки и легитимные звонки.

Для выявления аномалий мы применяли правило трех сигм, методы кластеризации DBSCAN и Изолирующий лес, а также евклидово расстояние. Суммарно всеми этими методами из набора данных из 150000 объектов к аномалиям были отнесены 2021 точек. Распределения аномалий и нормальных данных были представлены в виде диаграмм рассеивания для пар признаков. Аномалии, найденные каждым из методов, были визуализированы в двумерном пространстве с помощью алгоритма t-SNE.

По результатам анализа данных были выявлены такие аномалии как завышение стоимости звонка, превышение заявленного тарифа в минуту, значительные превышения времени разговора относительно среднего. Было выявлено, что количество звонков в выходные дни и в нерабочее время значительно ниже.

В данном исследовании мы столкнулись с проблемой несбалансированных классов в наборе данных. Мы использовали несбалансированные, взвешенные и дополненные выборки для обучения нейронной сети. Для каждой модели была проведена оценка качества классификатора, в частности построена матрица ошибок, рассчитаны метрики precision, recall, AUC. Также была построена ROC-кривая. Модель, обученная на дополненном наборе данных, смогла выявить практически все мошеннические звонки, однако было получено достаточно много ложноположительных результатов.

Выполненная исследовательская работа помогла сформировать понимание того, какие подходы можно использовать для борьбы с мошенничеством в области телекоммуникаций.

Список публикаций соискателя

1-А Полудворянин, С.М. Проблемы безопасности SIP-телефонии / С.М. Полудворянин, С.Н. Нестеренков // Информационные технологии и системы 2018 (ИТС 2018) : материалы междунар. науч. конф., Минск, 25 окт. 2018 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2018. – С. 54-55.

2-А Полудворянин, С. М. Методы борьбы со спамом в VoIP телефонии / С. М. Полудворянин, С. Н. Нестеренков // Информационные технологии и системы 2019 (ИТС 2019) : материалы междунар. науч. конф., Минск, 30 окт. 2019 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2019. – С. 258-259.