Министерство образования Республики Беларусь
Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

УДК 004.85

Шараев Никита Петрович

Обнаружение признаков сетевой разведки с использованием методов машинного обучения

АВТОРЕФЕРАТ
на соискание степени магистра

по специальности 1-98 80 01 «Информационная безопасность»

Научный руководитель

Петров Сергей Николаевич
кандидат технических наук,
доцент

Минск 2021

ВВЕДЕНИЕ

В последнее время наблюдается тенденция перехода от массовых кибератак отдельных злоумышленников к масштабным атакам киберпреступных группировок на конкретные организации (таргетированные или APT атаки). Данные атаки в значительной мере опасны для организаций, что связано в первую очередь с созданием злоумышленниками вредоносного программного обеспечения с учетом специфики работы и сетевой инфраструктуры организации. В общем случае, APT атаки состоят из четырех этапов: подготовка, проникновение, распространение и достижение цели. Обнаружить подобный тип атак на поздней стадии крайне сложно, а в отдельных случаях невозможно. По данной причине целесообразно провести обнаружение и анализ таргетированной атаки на этапе подготовки. На указанном этапе злоумышленники проводят процедуру сетевой разведки инфраструктуры организации, которую можно выявить с помощью методов машинного обучения.

Машинное обучение является наиболее динамически развивающейся областью науки в последние десятилетия. Основными направлениями деятельности при этом остаются задачи классификации, регрессии, компьютерного зрения, машинного перевода и иные, что применимо для обнаружения признаков сетевой разведки.

Первый раздел диссертационной работы посвящен анализу и классификации угроз информационной безопасности, выделению существующих признаков сетевой разведки. Во втором разделе подробно рассмотрены существующие наборы данных, выделены метрики и создан новый набор данных. В третьем разделе проводится формализация и математическое описание задачи обнаружения признаков сетевой разведки, детальный анализ выбранных в первом разделе алгоритмов машинного обучения с математическим описанием отдельных методов и исследование применения методов на практике с фиксированием эффективности работы. Четвертый раздел представляет собой результат разработки модуля обнаружения признаков сетевой разведки, базируясь на полученных в предыдущих разделах данных.

В результате проверки магистерской диссертации в системе «Антиплагиат» был получен результат в 94,82% оригинальности и 5,18% заимствований из различных источников, что эквивалентно использованию общепринятых определений, терминов и другой информации. Результат проверки представлен в Приложении Ж.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи работы

Целью диссертационной работы является разработка модуля обнаружения признаков сетевой разведки с использованием методов машинного обучения.

В соответствии с поставленной целью, в работе сформированы и решены следующие задачи:

– провести обнаружение и анализ признаков сетевой разведки;

– сформировать метрики, характерные для сетевой разведки;

– разработать набор данных на основе определенных метрик;

– исследовать существующие методы машинного обучения для задачи классификации;

– провести практический анализ методов машинного обучения и выбрать наилучший;

– разработать программный модуль обнаружения признаков сетевой разведки, базируясь на полученных в результате исследования данных.

Положения, выносимые на защиту

– набор данных, сформированный на основе полученных метрик;

– результаты исследования эффективности использования различных методов машинного обучения для обнаружения признаков сетевой разведки в сетевой инфраструктуре РУП «Белтелеком»;

– прототип модуля обнаружения признаков сетевой разведки.

Связь с приоритетными направлениями научных исследований и запросами реального сектора экономики

Тема диссертационной работы соответствует п. 3.8 «Обеспечение цифрового доверия, защита информационных ресурсов и информационно-коммуникационной инфраструктуры» Стратегии развития информатизации в Республике Беларусь на 2016 – 2022 годы, утвержденной на заседании Президиума Совета Министров от 3 ноября 2015 г. № 26.

В диссертации поставлена и решена актуальная задача по поиску оптимальной стратегии защиты информационных ресурсов с использованием свободного программного обеспечения и языка программирования Python версии 3. Научную новизну содержат способ

расчета метрики сетевой разведки, анализ эффективности использования различных методов машинного обучения для обнаружения признаков сетевой разведки, а также сам модуль обнаружения признаков сетевой разведки.

Практическая ценность работы состоит в том, что предложенное решение обеспечивает приемлемый уровень обеспечения информационной безопасности, без затрат на приобретение дорогостоящего лицензионного программного обеспечения.

Личный вклад соискателя

Содержание диссертации отображает личный вклад автора. Он заключается в изучении существующих подходов обнаружения признаков сетевой разведки, выделении их недостатков, формировании метрик и создании на основе их набора данных, выборе эффективного и производительного метода машинного обучения, проектировании и разработке модуля обнаружения признаков сетевой разведки, а также анализе его эффективности.

Определение цели и задач исследований, интерпретация и обобщение полученных результатов проводились с научным руководителем, кандидатом технических наук, доцентом С.Н. Петровым.

Апробация результатов диссертации

Теоретические результаты диссертационных исследований представлены в виде тезисов на следующих научных конференциях:

– XXV Международная научно-техническая конференция «Современные средства связи», Минск, 22–23 октября 2020 года;

– XVII Международная научно-практическая конференция «Управление информационными ресурсами», Минск, 12 марта 2021 года.

Практические результаты диссертационных исследований представлены в виде тезисов на следующих научных конференциях:

– 57-ой конференции аспирантов, магистрантов и студентов БГУИР, Минск, 19 – 23 апреля 2021 г;

– 7-ой конференции Maltepe University International Student Congress (MUISC), Стамбул, 5-7 мая 2021 г;

– 19-ой Белорусско-Российской научно-технической конференции «Технические средства защиты информации», Минск, 8 июня 2021 г.

Опубликованность результатов диссертации

По результатам исследований, представленных в диссертации, опубликовано 6 печатных работ в сборниках: «Инфокоммуникации: 57-я конференция аспирантов, магистрантов и студентов», «Современные средства связи: материалы XXV международной научно-технической конференции», «Управление информационными ресурсами: материалы XVII Международной научно-практической конференции», «Education & applications on design and engineering during pandemic 2021» и «Технические средства защиты информации: тезисы докладов XIX Белорусско-Российской научно-технической конференции».

Структура и объем диссертации

Диссертационная работа состоит из введения, общей характеристики работы, основной части из четырех разделов, заключения, списка использованных источников, списка собственных источников, 8 приложений, графического материала. Полный объем диссертационной работы составляет 73 страниц, включая 26 иллюстраций, список использованных источников из 31 наименования, список собственных источников из 6 наименований, 9 приложений объемом 16 страниц, 1 акт внедрения, 1 справка внедрения результатов диссертационной работы, графические материалы из 12 слайдов презентации в формате А4.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Введение содержит краткое описание работы и обоснование необходимости исследований.

В первом разделе проведен анализ существующего в Республике Беларусь понятийного аппарата. Предложена классификация угроз информационной безопасности, отражающая законодательство, характер угрозы, вид воздействия угрозы и причину угрозы. Рассмотрены возможные существующие признаки сетевой разведки, а также способы их обнаружения.

Раздел два содержит анализ существующих открытых наборов данных, предназначенных для обнаружения аномалий сетевого трафика и сетевой разведки в частности. В связи с отсутствием качественных, актуальных и пригодных для указанных целей датасетов произведено определение метрик, характерных для сетевой разведки, а также представлены формулы их расчета. Базируюсь на данных метриках, спроектирована топология и создан тестовый сегмент в корпоративной сети РУП «Белтелеком», в рамках которого проводилось создание датасета.

В третьем разделе представлена основные вычисления. Проведено описание математической задачи классификации с помощью методов машинного обучения признаков сетевой разведки, а также представлены основные формулы в данном направлении. Рассмотрены существующие методы классификации, их особенности и, частично, способы их расчета. Выделены наиболее перспективные для целей диссертационной работы методы. Для всех выбранных методов проведено обучение и тестирование с использованием различных гиперпараметров. Дополнительно проведено улучшение отдельных алгоритмов с помощью процедур бэггинга и бустинга.

Четвертый раздел описывает требований, необходимых для стабильной работы разрабатываемой программы. Представлены необходимые компоненты и зависимости Спроектирован и разработан модуль обнаружения признаков сетевой разведки. Внесены оптимизационные правки в программный код модуля. Проведен анализ работы модуля.

В заключении сформулированы основные выводы по диссертационной работе и представлены полученные результаты.

ЗАКЛЮЧЕНИЕ

В результате выполнения работы: проведен анализ признаков сетевой разведки и существующих методов машинного обучения; сформированы метрики, характерные для сетевой разведки; разработан набор данных на основе определенных метрик; выбран подходящий с практической точки зрения метод машинного обучения, показавший наилучший результат; разработан программный модуль обнаружения признаков сетевой разведки.

Анализ существующих открытых наборов данных, предназначенных для обнаружения аномалий сетевого трафика и сетевой разведки в частности показал отсутствие качественных и актуальных датасетов. В этой связи произведена разработка нового датасета, метрики и способы расчета которых представлены в диссертационной работе.

Базируясь на созданном датасете, проведена тренировка и тестирование алгоритмов машинного обучения с использованием выделенных гиперпараметров. Выделены наиболее перспективные для целей диссертационной работы методы: логистическая регрессия, квадратичный дискриминантный анализ, метод опорных векторов, метод К-ближайших соседей, «наивный» байесовский классификатор, дерево принятия решений и многослойный персептрон (нейронная сеть прямого распределения). Наилучшие результаты показал метод дерева принятия решений с параметрами criterion = «gini» и splitter = «random», с точностью 100 % и скоростью работы 0,912 мс. Дополнительно проведено улучшение отдельных алгоритмов с помощью процедур бэггинга и бустинга, а также представление алгоритма с наилучшими параметрами в виде программного кода, что позволило увеличить скорость работы приблизительно в 2 раза.

На основе полученных данных спроектирован и разработан модуль обнаружения признаков сетевой разведки. Проведено исследование работы модуля, показавшее 100% обнаружения попыток проведения сетевой разведки, а также ошибки второго рода (falsepositive) при наличии подключенного к прослушиваемому интерфейсу DNS или DHCP сервера. При этом добиться исключения ошибок второго рода можно, применяя параметр «excluded_ips» при конфигурации программного обеспечения.

СПИСОК СОБСТВЕННЫХ ПУБЛИКАЦИЙ

1–А. Шараев, Н. П. Профессиональный стандарт в сфере информационной безопасности инфокоммуникационных систем / Н. П. Шараев // Современные средства связи : материалы XXV Междунар. науч.-техн. конф., 22–23 окт. 2020 года, Минск / Белорусская государственная академия связи ; редкол.: А. О. Зеневич [и др.]. – Минск : БГАС, 2020. – С. 201-202.

2–А. Шараев, Н. П. Обнаружение признаков сетевой разведки с использованием машинного обучения / Н. П. Шараев, С. Н. Петров // Современные средства связи : материалы XXV Междунар. науч.-техн. конф., 22–23 окт. 2020 года, Минск / Белорусская государственная академия связи ; редкол.: А. О. Зеневич [и др.]. – Минск : БГАС, 2020. – С. 209-210.

3–А. Шараев, Н. П. Выявление и анализ признаков сетевой разведки методом машинного обучения / Н. П. Шараев, С. Н. Петров // Управление информационными ресурсами: материалы XVII Междунар. науч.-практ. конф., 12 мар. 2021 года, Минск / Акад. упр. при Президенте Респ. Беларусь ; редкол. : А. С. Лаптенок. – Минск: Академия управления при Президенте Республики Беларусь, 2021. – С. 238-240.

4–А. Шараев, Н. П. Выявление сетевой разведки методами машинного обучения / Н. П. Шараев, С. Н. Петров // Защита информации : сборник материалов 57-й научной конференции аспирантов, магистрантов и студентов БГУИР, 19-23 апреля 2021 г., БГУИР, Минск, Беларусь: тезисы докладов. – Мн. – 2021. – С. 34-37.

5–A. Sharaev N. Identification of Network Intelligence Features by Machine Learning Methods // Education & applications on design and engineering during pandemic 2021: 7th Maltepe University International Student Congress (MUISC), 5-7 May 2021, Istanbul, Turkey / Maltepe University; Editors: Dr. Öğr. Üyesi Emre Atlıer Olca and other – Istanbul: Maltepe University, 2021. – P. 47.

6–А. Шараев, Н. П. Сравнительный анализ методов машинного обучения для решения задачи обнаружения признаков сетевой разведки / Н. П. Шараев, С. Н. Петров // Технические средства защиты информации: тез. докл. XIX Белорусско-Российской науч.-техн. конф, 8 июня 2021 г., Минск / редкол. : Т. В. Борботько [и др.]. – Минск : БГУИР, 2021. – С. 104.