Сычев Д.И. Методы машинного и глубокого обучения для систем обнаружения вторжений: обзор и анализ // Международный журнал информационных технологий и энергоэффективности. – 2023. –
Т. 8 № 4(30) с. 9–17

Международный журнал информационных технологий и энергоэффективности

Сайт журнала:
http://www.openaccessscience.ru/index.php/ijcse/

УДК 62

МЕТОДЫ МАШИННОГО И ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ СИСТЕМ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ: ОБЗОР И АНАЛИЗ

Сычев Д.И.
Санкт-Петербургский государственный университет телекоммуникаций имени профессора М.А. Бонч-Бруевича, Санкт-Петербург, Россия (193232, г. Санкт-Петербург, пр. Большевиков, 22, к. 1), e-mail: s.denis_2001@mail.ru

В современном мире, с развитием информационных технологий, исследования в области кибербезопасности играют все большую роль. Одной из важных систем в кибербезопасности, является система обнаружения вторжений (англ. *Intrusion Detection Systems - IDS*). IDS мониторит состояние программного и аппаратного обеспечения, работающего в сети. Несмотря на прошедшие десятки лет разработки, существующие IDS по прежнему сталкиваются с трудностями в точности определения вторжения, обнаружении новых атак и ложных срабатываний. Для решения вышеописанных проблем ведутся исследования в области разработки IDS, использующую методы машинного обучения. Машинное обучение может автоматически определять существенные различия между общими и аномальными данными с высокой точностью. Системы обнаружения вторжений можно разделить на 2 типа: на основе сигнатур и на основе аномалий. IDS на основе сигнатур опираются на предопределенные шаблоны или сигнатуры известных атак, в то время как IDS на основе аномалий выявляют аномальное поведение, отклоняющееся от нормальной сетевой активности.
Методы машинного обучения (ML) и глубокого обучения (DL) широко используются в IDS для повышения точности и эффективности обнаружения вторжений. В этом тексте мы рассмотрим таксономию методов ML и DL, используемых для IDS.

Ключевые слова: Системы обнаружения вторжений, Машинное обучение, Глубокое обучение, Сетевая безопасность.

MACHINE AND DEEP LEARNING METHODS FOR INTRUSION DETECTION SYSTEMS: OVERVIEW AND ANALYSIS

Sychev D.I.
St. Petersburg State University of Telecommunications named after Professor M.A. Bonch-Bruevich, St. Petersburg, Russia (193232, St. Petersburg, Bolshevikov Ave., 22, room 1), e-mail: s.denis_2001@mail.ru

In the modern world, with the development of information technology, research in the field of cybersecurity is playing an increasingly important role. One of the important systems in cybersecurity is the intrusion Detection System (English Intrusion Detection Systems - IDS). IDS monitors the status of the software and hardware running on the network. Despite the past decades of development, existing IDS still face difficulties in accurately detecting intrusion, detecting new attacks and false positives. To solve the problems described above, research is underway in the field of IDS development using machine learning methods. Machine learning can automatically detect significant differences between general and anomalous data with high accuracy. Intrusion detection systems can be divided into 2 types: signature-based and anomaly-based. Signature-based IDS rely on predefined patterns or

signatures of known attacks, while anomaly-based IDS detect abnormal behavior that deviates from normal network activity.

Machine learning (ML) and deep learning (DL) techniques are widely used in IDS to improve the accuracy and efficiency of intrusion detection. In this text, we will look at the taxonomy of ML and DL methods used for IDS.

Keywords: Intrusion detection systems, Machine learning, Deep learning, Network security.

Введение

Развитие сетевых устройств оказывает все большее влияние на современную жизнь, что делает кибербезопасность важной областью для исследований[1,2]. Основные методы борьбы с киберпреступностью в основном составляют антивирусное программное обеспечение, брандмауэры и системы обнаружения вторжений (IDS). Эти методы защищают устройства в сети от внутренних и внешних атак. Среди них IDS — это тип системы обнаружения, которая играет ключевую роль в обеспечении защиты и отслеживания состояния программного и аппаратного обеспечения, работающего в сети.

Первая система обнаружения вторжений была предложена в 1980 году [1]. С тех пор появилось много различных IDS продуктов. Тем не менее, многие IDS по-прежнему страдают от высокого уровня ложных срабатываний, генерируя множество предупреждений для мало опасных ситуаций, что увеличивает нагрузку на систему аналитики и может привести к тому, что серьезные вредоносные атаки будут проигнорированы. Еще одна проблема существующих IDS заключается в том, что они уязвимы к появляющимся новым типам атак. Поскольку сетевые системы продолжают развиваться и изменяться, злоумышленники регулярно разрабатывают новые варианты атак. Таким образом, в современном мире существует необходимость в IDS, способных обнаруживать ранее неизвестные атаки.

Чтобы решить вышеуказанные проблемы, начали разрабатываться IDS с использованием методов машинного обучения. Машинное обучение — это метод искусственного интеллекта, который может автоматически извлекать полезную информацию из больших наборов данных [2]. При наличии большого датасета, IDS на основе машинного обучения показывают могут показывать хороший результат в обнаружении проникновений, а за счет генерализации, справиться с некоторыми вариантами новых атак. Кроме того, IDS на основе машинного обучения не имеют сильной зависимости от предметной области, поэтому их легко спроектировать.

Ниже, представлена классификация методов машинного обучения и глубокого обучения для IDS. Выделены различные подходы на основе алгоритмов, источников данных и архитектур. Таким образом, предоставлен четкий и структурированный обзор последних достижений в области обнаружения вторжений с использованием машинного и глубокого обучения. Эта таксономия будет полезна как для исследователей, так и для практиков, которые стремятся лучше понять сильные и слабые стороны различных подходов и выбрать наиболее подходящий метод для конкретного использования.

1. Обзор систем обнаружения вторжений

Системы обнаружения вторжений могут быть разделены на две основные категории: сетевые IDS (NIDS) и хост-ориентированные IDS (HIDS). NIDS анализируют сетевой трафик, чтобы обнаруживать аномалии и атаки, в то время как HIDS сосредоточены на мониторинге отдельных хостов, таких как серверы или рабочие станции[3].

Традиционные подходы к обнаружению вторжений включают использование сигнатурных и правиловых методов. Сигнатурные методы заключаются в сравнении сетевого трафика с известными шаблонами атак, называемыми сигнатурами. Если сетевой трафик соответствует определенной сигнатуре, IDS генерирует предупреждение о возможной атаке. Правиловые методы основаны на анализе сетевого трафика или системных событий с использованием заранее определенных правил, которые указывают на подозрительную активность.

Традиционные методы обнаружения вторжений имеют несколько ограничений:

- Ложные срабатывания: Сигнатурные и правиловые методы склонны к ложным срабатываниям, когда допустимое поведение или сетевой трафик ошибочно определяются как атака[3,4]. Это может привести к перегрузке системы предупреждений и ухудшению эффективности работы специалистов по безопасности.
- Обнаружение атак "нулевого дня": Традиционные IDS слабо справляются с обнаружением атак "нулевого дня", которые представляют собой новые или ранее неизвестные угрозы, не имеющие сигнатур или явных признаков. Это создает слепые пятна в обнаружении и оставляет системы уязвимыми перед новыми атаками.
- Ресурсоемкость: Поддержка актуальной базы сигнатур и правил для традиционных IDS требует постоянного обновления и значительных ресурсов на поддержание эффективности системы. Кроме того, обработка больших объемов сетевого трафика может вызвать задержки и замедления в работе IDS.
- Отсутствие адаптивности: Традиционные методы обнаружения вторжений статичны и не способны адаптироваться к изменяющимся сценариям угроз или условиям сетевого трафика.

Для преодоления ограничений традиционных подходов к обнаружению вторжений, активно применяются методы машинного обучения и глубокого обучения. Эти методы способны автоматически обучаться на основе данных, выявлять закономерности и адаптироваться к новым сценариям угроз. Таким образом, ML и DL методы могут снизить количество ложных срабатываний, обеспечить обнаружение атак "нулевого дня" и улучшить адаптивность системы к изменяющимся условиям[5].

Машинное обучение и глубокое обучение в IDS применяются для решения различных задач, таких как классификация трафика, аномалий и атак, а также для анализа поведения сетевых узлов

2. Методы машинного обучения для IDS

Среди наиболее распространенных подходов к машинному обучению в системах обнаружения вторжений можно выделить обучение с учителем. В данном методе, модель обучается на основе размеченных заранее данных, содержащих примеры нормального поведения и возможных атак[6]. Целью обучения является настройка модели таким образом, чтобы она смогла классифицировать наблюдаемое поведение в системе как нормальное или аномальное. К популярным методам обучения с учителем, которые используются для IDS, относятся:

- **Метод логистической регрессии**: Это статистический метод для анализа набора данных, в котором одна или несколько независимых переменных используются для предсказания вероятности принадлежности наблюдения к одному из двух классов (например, нормальный или аномальный). Логистическая регрессия является простым и интерпретируемым подходом, однако для сложных и нелинейных зависимостей, она может быть менее эффективной.
- **Метод опорных векторов** (SVM): SVM – это мощный алгоритм классификации, который стремится найти оптимальную разделяющую гиперплоскость между двумя классами. SVM хорошо справляется с задачами, которые имеют большое количество признаков, и сложными зависимостями, но может быть ресурсоемким при больших объемах данных.
- **Дерево решений и ансамблевое обучение**: Дерево решений – это иерархические структуры, которые последовательно разделяют данные на основе определенных критериев. Ансамблевые методы, такие как метод случайного леса (Random Forest) и градиентный бустинг (Gradient Boosting), объединяют множество деревьев решений для улучшения производительности и устойчивости к переобучению. Деревья решений и ансамблевые методы характеризуются высокой точностью и интерпретируемостью, но могут столкнуться с проблемами масштабируемости при обработке больших данных.

Обучение без учителя используется в случаях, когда размеченные данные недоступны или их очень мало. Эти методы пытаются выявить аномалии или кластеры, опираясь на структуру и распределение данных. Некоторые популярные методы обучения без учителя для IDS включают:

- **K-средних**: Это итеративный алгоритм кластеризации, который разделяет данные на K кластеров на основе расстояния между точками данных. K-средних может использоваться для выявления групп аномального поведения или атак, но подвержен влиянию выбросов и чувствителен к исходному выбору центроидов кластеров.
- **DBSCAN** (Density-Based Spatial Clustering of Applications with Noise) – это алгоритм кластеризации, основанный на плотности, который группирует точки данных на основе их плотности и расстояния. DBSCAN хорошо справляется с аномалиями и может обнаруживать кластеры произвольной формы, но требует настройки гиперпараметров для определения плотности и расстояния.
- **Автоэнкодеры** – это нейронные сети, которые сначала сжимают данные в низкоразмерное представление (кодирование), а затем восстанавливают исходные данные из этого представления (декодирование). В контексте IDS, автоэнкодеры могут обучаться на нормализованных данных и использоваться для обнаружения аномалий, сравнивая восстановленные данные с исходными. Если разница между восстановленными и исходными данными велика, это может указывать на аномальное поведение или атаку.

3. Методы глубокого обучения для IDS

Глубокое обучение (Deep Learning, DL) – это подраздел машинного обучения, который использует нейронные сети с большим количеством слоев для обработки и анализа данных. Глубокое обучение может автоматически извлекать сложные признаки из сырых данных, что делает его особенно полезным для применения в системах обнаружения вторжений[6,7]. Одним из основных методов глубокого обучения, который используется в IDS является способ обучения Сверточной нейронной сети (Convolutional Neural Networks, CNN). CNN - это тип глубоких нейронных сетей, разработанный специально для обработки изображений и временных рядов. В контексте IDS, CNN могут использоваться для анализа сетевого трафика и системных журналов, автоматически извлекая признаки, связанные с атаками. CNN состоят из сверточных, пулинговых (субдискретизирующих) и полносвязных слоев, которые обрабатывают и объединяют признаки на разных уровнях абстракции[7]. Сверточные нейронные сети зарекомендовали себя как отличный инструмент для различных приложений и сетевых инструментов, однако, при использовании их с системами обнаружения вторжений, нужно учитывать ряд возможных подводных камней, которые могут появится при разработке:

- **Высокая вычислительная сложность**: CNN обычно состоят из нескольких уровней с многочисленными параметрами, что делает их обучение дорогостоящими в вычислительном отношении. Это может быть серьезным ограничением, особенно в при работе в реальном времени, где важны малая задержка и эффективная обработка.
- **Потребность в больших наборах размеченных данных**: Чтобы достичь высокой производительности, CNN требуют значительных объемов размеченных данных. Получение большого набора размеченных данных для обнаружения вторжений может быть затруднено из-за динамического характера киберугроз и сложности получения достоверных меток для сетевого трафика.
- **Чувствительность к входному представлению функций**: CNN предназначены для работы с сеткообразными структурами данных (например, изображениями). Применение их к IDS может потребовать преобразования данных сетевого трафика в подходящее представление, которое не всегда может быть простым или оптимальным.
- **Ограниченная интерпретируемость**: CNN часто считают моделями «черного ящика» из-за их сложной структуры и отсутствия интерпретируемости. Может быть трудно понять, почему CNN классифицировала конкретное сетевое событие как злонамеренное или неопасное, что может привести к недоверию со стороны экспертов по безопасности.

Также, среди популярных методов глубокого обучения, стоит рассмоетрть рекуррентные нейронные сети и автоэнкодер. Рекуррентные нейронные сети (Recurrent Neural Networks, RNN) – это класс глубоких нейронных сетей, специально разработанный для работы с последовательными данными, такими как временные ряды или текст. RNN обладают внутренней памятью и могут учитывать контекст и порядок событий при анализе данных. В IDS, RNN могут использоваться для обнаружения атак, основанных на аномальных последовательностях действий или сетевых запросов. Одним из распространенных вариантов RNN являются сети долгой краткосрочные памяти (LSTM) и еправляемый рекуррентный блок

(GRU), которые способны эффективно обрабатывать долгосрочные зависимости между событиями.

Как уже упоминалось в предыдущих разделах, автоэнкодеры являются нейронными сетями, которые сначала кодируют данные в компактное представление, а затем восстанавливают их из этого представления. Вариационные автоэнкодеры (VAE) – это расширение автоэнкодеров, которое вводит стохастический слой в кодирование, позволяя модели генерировать новые данные, похожие на обучающую выборку. И автоэнкодеры, и VAE могут использоваться в IDS для обнаружения аномалий и атак, основанныхна разнице между восстановленными данными и исходными данными, а также для создания репрезентативных эмбеддингов, которые могут использоваться в других моделях машинного обучения.

Методы глубокого обучения предлагают множество инновационных подходов для систем обнаружения вторжений. Они позволяют автоматически извлекать сложные признаки из данных, обеспечивая высокую точность и эффективность в обнаружении атак и аномалий. Однако эти методы требуют больших вычислительных ресурсов и обучающих данных для достижения оптимальной производительности. Важно выбирать подходящие методы глубокого обучения в зависимости от конкретной задачи IDS, доступных данных и вычислительных возможностей.

4. Проблемы и вызовы в применении ML и DL для IDS

Хотя машинное обучение и глубокое обучение предлагают множество преимуществ для систем обнаружения вторжений, их применение также связано с определенными проблемами и вызовами. В этом разделе мы рассмотрим некоторые общие ключевые аспекты, которые необходимо учитывать при использовании ML и DL в IDS.

Для эффективного обучения большинства моделей машинного и глубокого обучения требуются большие объемы размеченных данных. Однако в контексте IDS, сбор и разметка данных о вторжениях может быть трудоемким и дорогостоящим процессом. Это может привести к использованию небольших или несбалансированных обучающих наборов данных, что ухудшает качество обучения моделей и их способность обнаруживать атаки[8]. Так же, в современном мире злоумышленники постоянно разрабатывают новые и изменяющиеся стратегии атак, чтобы обойти системы обнаружения вторжений. Это требует от IDS с использованием ML и DL умения быстро адаптироваться к новым видам атак и обеспечивать надежное обнаружение. Однако обновление и повторное обучение моделей может быть ресурсоемким процессом, особенно в случае сложных архитектур глубокого обучения.

Модели глубокого обучения, требуют больших вычислительных ресурсов для обучения и инференции. Такие ресурсы, как графические процессоры (GPU) и специализированные аппаратные ускорители, могут быть дорогостоящими и недоступными для некоторых организаций. В результате, выбор подходящих моделей и оптимизация вычислительных процессов становятся критически важными для успешного использования ML и DL в IDS. Модели машинного обучения и, в особенности, глубокого обучения часто называют "черными ящиками" из-за их сложности и отсутствия интерпретируемости. Это может создать трудности для понимания и объяснения причин, по которым модель считает определенное поведение аномальным или атакующим. В результате, это может привести к ошибкам и недоверию со стороны пользователей и экспертов в области безопасности.

Использование ML и DL в IDS может потребовать сбора и обработки большого количества сетевых данных и журналов, которые могут содержать конфиденциальную и чувствительную информацию. Защита этих данных от утечек и злоупотреблений является важным аспектом применения ML и DL в системах обнаружения вторжений. Это может потребовать разработки дополнительных методов защиты данных и обеспечения соблюдения принципов конфиденциальности и соответствия законодательству.

Применение методов машинного и глубокого обучения для IDS представляет собой мощный инструмент в борьбе с киберугрозами, однако его успешная реализация связана с рядом проблем и вызовов. Здравый учет этих аспектов и разработка стратегий является важным аспектом, на который кампаниям, разрабатывающие IDS, стоит уделить внимание.

Вывод

В этой статье мы рассмотрели основы систем обнаружения вторжений (IDS) и их развитие с течением времени. Мы подробно обсудили применение методов машинного обучения и глубокого обучения в контексте IDS, а также рассмотрели некоторые из вызовов и проблем, связанных с их использованием. Машинное обучение и глубокое обучение предлагают значительные преимущества для повышения эффективности и точности систем обнаружения вторжений, позволяя адаптироваться к изменяющимся атакам и автоматически извлекать сложные признаки из данных. Однако успешное внедрение этих технологий требует учета ряда проблем и вызовов, таких как недостаток размеченных данных, вычислительные ресурсы, уязвимость к противодействию и манипуляциям, а также проблемы конфиденциальности и безопасности данных.

Интеграция машинного и глубокого обучения в системы обнаружения вторжений представляет собой значительный шаг вперед в борьбе с киберугрозами. Однако для успешного применения этих подходов необходимо тщательно учесть ряд проблем и вызовов, а также разработать адекватные стратегии и решения для преодоления потенциальных препятствий на пути к обеспечению надежной и эффективной кибербезопасности. Сотрудничество между исследователями, специалистами в области безопасности и разработчиками программного обеспечения является ключевым фактором для достижения этих целей и создания новых, инновационных решений для борьбы с киберпреступностью. Важно продолжать исследования и разработки, направленные на повышение эффективности IDS и устойчивости к различным видам атак, чтобы обеспечить безопасность и защиту киберпространства в будущем.

Список литературы

1. Гельфанд А. М. и др. Области применения аналитики больших данных в критических информационных инфраструктурах //Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2022). – 2022. – С. 438-440.
2. Бударный Г. С. и др. Разновидности нарушений безопасности и типовые атаки на операционную систему //Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2022). – 2022. – С. 406-411.

3. Ковалев И. А., Косов Н. А. Состязательные атаки в нейронных сетях //Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2021). – 2021. – С. 490-492.
4. Тимофеев Р. С., Косов Н. А. Сравнение методов обучения сверточных нейронных сетей //Актуальные научные исследования в современном мире. – 2021. – №. 6-1. – С. 97-102.
5. Косов Н. А. и др. Анализ методов машинного обучения для детектирования аномалий в сетевом трафике //Цифровизация образования: теоретические и прикладные исследования современной науки. – 2021. – С. 33-37.
6. Синельщиков В. С., Цветков А. Ю. Защита персональных данных на предприятии //Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2021). – 2021. – С. 653-657.
7. Цветков А. Ю., Рузманов Е. Ю. Рассмотрение тестирования на проникновение в задачах защиты информации //ББК 3 П27. – 2021. – С. 55.
8. Шемякин С. Н. и др. Использование теории графов для моделирования безопасности облачных систем //Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. – 2021. – №. 2. – С. 31-35.
9. Гельфанд А. М., Гвоздев Ю. В., Штеренберг С. И. Исследования недостатков языков высокоуровнего программирования для осуществления скрытого вложения в исполнимые файлы //Актуальные проблемы инфотелекоммуникаций в науке и образовании. – 2015. – С. 295-297.
10. Штеренберг, С. И. Компьютерные вирусы / С. И. Штеренберг, А. В. Красов, А. Ю. Цветков. Том Часть 1. – Санкт-Петербург : Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, 2015. – 63 с.
11. Зимин А. Е., Косов Н. А. Обеспечение информационной безопасности в процессе создания и использования программ для ЭВМ //Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2017). – 2017. – С. 343-348.

References

1. Gelfand A. M. et al. Applications of big data analytics in critical information infrastructures // Actual problems of infotelecommunications in science and education (APINO 2022). - 2022. - . pp. 438-440.
2. Budarny G. S. et al. Variety of security violations and typical attacks on the operating system // Actual problems of infotelecommunications in science and education (APINO 2022). - 2022. - pp. 406-411.
3. Kovalev I. A., Kosov N. A. Competitive attacks in neural networks // Actual problems of infotelecommunications in science and education (APINO 2021). - 2021. - pp. 490-492.
4. Timofeev R. S., Kosov N. A. Comparison of training methods for convolutional neural networks // Actual scientific research in the modern world. – 2021. – no. 6-1. - pp. 97-102.
5. Kosov N. A. et al. Analysis of machine learning methods for detecting anomalies in network traffic //Digitalization of education: theoretical and applied research of modern science. - 2021. - pp. 33-37.

6. Sinelshchikov V. S., Tsvetkov A. Yu. Protection of personal data at the enterprise // Actual problems of infotelecommunications in science and education (APINO 2021). - 2021. - pp. 653-657.
7. Tsvetkov A. Yu., Ruzmanov E. Yu. Consideration of penetration testing for information protection //BBK 3 P27. - 2021. - pp. 55.
8. Shemyakin S. N. et al. Using graph theory to model the security of cloud systems // Bulletin of the St. Petersburg State University of Technology and Design. Series 1: Natural and technical sciences. – 2021. – no. 2. - pp. 31-35.
9. Gelfand A. M., Gvozdev Yu. V., Shterenberg S. I. Investigation of the shortcomings of high-level programming languages for the implementation of hidden attachments to executable files // Actual problems of infotelecommunications in science and education. - 2015. - S. 295-297.
10. Shterenberg, S. I. Computer viruses / S. I. Shterenberg, A. V. Krasov, A. Yu. Tsvetkov. Volume Part 1. - St. Petersburg: St. Petersburg State University of Telecommunications. prof. M.A. Bonch-Bruevich, 2015. – p.63.
11. Zimin A. E., Kosov N. A. Ensuring information security in the process of creating and using computer programs // Actual problems of infotelecommunications in science and education (APINO 2017). - 2017. - pp. 343-348.