Инженерный вестник Дона, №3 (2025)
ivdon.ru/ru/magazine/archive/n3y2025/9944

Методы защиты интернета вещей от атак нулевого дня

С.Ю. Рыбаков

Московский технический университет связи и информатики, г. Москва

Аннотация: Атаки нулевого дня представляют одну из наиболее опасных угроз для безопасности современных систем, приложений и инфраструктуры поскольку они являются непредсказуемыми. Традиционные средства защиты, основанные на сигнатурах, не способны выявить атаки "нулевого дня", из-за неизвестных средствам защиты сигнатур подобных атак. Для противодействия таким атакам в сетях интернета вещей (Internet of Things - IoT), необходимы как углубленные исследования, так и внедрение практических мер. Представленный обзор современных исследований, направленных на обнаружение атак "нулевого дня" показал, что наилучшую эффективность в выявлении атак и ботнетов "нулевого дня" в сетях IoT демонстрируют методы глубокого обучения. Эти подходы позволяют анализировать аномалии в сетевом трафике и выявлять новые угрозы, а также атаки "нулевого дня", минимизируя количество ложных срабатываний.
Ключевые слова: атака нулевого дня, уязвимость, интернет вещей, машинное обучение, аномалия, сигнатурный метод защиты, автоэнкодер, сетевой трафик.

Постановка задачи

Термин "Интернет вещей" описывает группу объектов, объединенных через Интернет для обмена и хранения данных, относящихся к конкретным сервисам или приложениям. К таким объектам относятся умные устройства, исполнительные механизмы, сенсоры и другие устройства с встроенной электроникой [1, 2]. Развитие IoT способствует ускоренному внедрению новых и инновационных услуг и приложений благодаря стремительному прогрессу в области информационных технологий [3, 4]. Интеллектуальные устройства IoT находят применение в различных сложных сценариях, поскольку их функциональные возможности постоянно расширяются за счет усовершенствования обработки данных, сенсорных технологий и вычислительных мощностей.

Атака нулевого дня Zero-day attacks — это случайная атака, которую невозможно искоренить, ее можно только выявить и избежать. Ее также называют атакой одного дня, которая происходит в день нулевой осведомленности.

Атака нулевого дня происходит, когда хакеры используют брешь в безопасности до того, как поставщик обнаружит ее и поспешит ликвидировать. Атаки нулевого дня включают в себя проникновение вредоносных программ, шпионских программ и предоставление несанкционированного доступа к пользовательской информации.

Атаки нулевого дня, которые представляют собой новые (аномальные) атаки, использующие ранее известные системные дефекты, являются серьезной проблемой "степени знания системы" как различия между легитимными и нелегитимными пользователями.

Уязвимость нулевого дня — это дефект системы, котором пользователь совершенно не знает. Эта слабость часто используется хакерами для изменения компьютерных программ, кражи данных и заражения сетей. Атака нулевого дня — это атака, направленная на использование только что обнаруженной уязвимости. Когда обнаруживается недостаток в системе безопасности, создаются патчи, закрывающие дыры в системе. Атаки "нулевого дня" вызывают серьезную озабоченность, поскольку они настолько неожиданны. Для использования этих слабостей применяется ряд стратегий.

В IoT массовое распространение и длительное физическое время жизни нарушат парадигму безопасности "проникнуть и поставить заплатку", которая помогает смягчить последствия уязвимостей, присущих отдельным системам. Уязвимости от "нулевого дня" до "вечных дней" изобилуют в современных встраиваемых устройствах. Исправление этих уязвимостей и в нормальных условиях достаточно сложно, а с развитием Интернета вещей оно стало еще сложнее.

Типичные методы защиты от атак нулевого дня

Системы обнаружения вторжений СОВ (IDS - Intrusion Detection Systems) исследуются уже несколько десятилетий, и они продолжают

развиваться как основной элемент компьютерной и сетевой безопасности. Обнаружение ранее обнаруженных уязвимостей, часто известных как угрозы "нулевого дня", тем не менее, является сложной задачей.

Обнаружение атак нулевого дня путем классификации сетевого трафика в IoT предполагает выявление и устранение ранее неизвестных или нераскрытых уязвимостей и методов атак, использующих эти уязвимости. Такой подход крайне важен для безопасности IoT, поскольку он обеспечивает дополнительный уровень защиты от возникающих угроз, гарантируя своевременное обнаружение и предотвращение атак, использующих уязвимости, которые еще не были исправлены или устранены с помощью обновлений безопасности

Атаки "нулевого дня" используют неизвестные уязвимости, чтобы обходить системы мониторинга кибербезопасности.

Типичным методом обнаружения атак нулевого дня является создание базы данных на основе сигнатур и документирование таких атак. При этом не учитывается, сопоставима ли стратегия атаки с группой других атак.

Так в [5] считают, что вариации атак могут быть легко обнаружены и добавлены в онтологию базы данных благодаря пониманию стратегии нападения, которая обеспечивает семантическую связь между частями атаки.

Под онтологией обычно понимается формат хранения структурированных данных. Онтология используется для создания базы знаний об атаках XML-инъекций на веб-сервисы. В работе предлагается использовать уникальный идентификатор (Extended Identifier — XID) в качестве гибридного метода обнаружения, который сочетает в себе обнаружение на основе сигнатур и знаний.

Поскольку многие новые и неизвестные атаки генерируются с использованием хорошо известных стратегий (известных сигнатур), следует ожидать низкого уровня ложноположительных обнаружений.

Другими словами, в выводах учитывались только те действия атаки, которые удовлетворяли ограничениям аксиом общих классов - первые, проверенные в процедуре обнаружения. Рекомендуется расширить онтологию, чтобы включить в нее дополнительные виды атак на веб-сервисы, такие как отказ в обслуживании. С ростом числа классов атак и аксиом увеличивается и способность гибридного подхода обнаруживать атаки нулевого дня.

В [6] представлено исследование рисков "нулевого дня" для сетей IoT. В качестве метода принятия решения о нападениях "нулевого дня" предложена стратегия на основе контекстного графа. Используя распределенную систему диагностики, предложенная методика классифицировала контекст как на центральном поставщике услуг, так и на локальном пользовательском сайте. При обнаружении атаки "нулевого дня" для передачи сигналов тревоги и восстановления доверия между сетевыми организациями и IoT-устройствами использовался специальный протокол обмена данными.

Хотя построенные графы экземпляров могут не отражать всех маршрутов атак нулевого дня, когда некоторые действия атакующих уклоняются от системных вызовов или когда временной интервал атаки значительно превышает исследуемый период времени такой метод может выявить хотя бы часть путей. Согласно этому исследованию, для обнаружения путей атак нулевого дня можно использовать байесовские сети. С этой целью был разработан граф экземпляров объектов, который является основой для байесовских сетей. Интегрируя данные о вторжениях и оценивая вероятности заражения объектов, реализованная система ZePro может успешно раскрывать пути атак нулевого дня [7].

Важной проблемой является выявление полиморфных червей "нулевого дня" и разработка сигнатур на пограничных сетевых шлюзах, так чтобы их можно было ликвидировать прямо на входе.

Однако исследования показали, что большинство недавно созданных сетевых сигнатур, как правило не основаны на уязвимостях и легко обходятся атаками. В [8] утверждается, что сигнатуры, основанные на уязвимостях, могут быть созданы на сетевом уровне, не требуя оценки действия червей на уровне хоста или восприимчивых программистов.

Они начинают с разработки сетевого генератора сигнатур на основе длины (Length-based Signature Generator - LESG) для червей, использующих дефекты переполнения буфера. Создаваемые сигнатуры присущи переполнениям буфера, что затрудняет злоумышленникам их обход. Генератор LESG отличается высокой скоростью, устойчивостью к шумам и отличным подбором сигнатур и оказывается способным достичь поставленных целей, основываясь на реальных уязвимостях нескольких протоколов и реальных сетевых данных.

В дальнейшем система была усовершенствована путем создания комплексных сигнатур для обфусцированных атак "нулевого дня" в формате snort.

Обфуска́ция (от лат. obfuscare или запутывание кода) — приведение исходного кода или исполняемого кода программы к виду, сохраняющему её функциональность, но затрудняющему анализ, понимание алгоритмов работы и модификацию при декомпиляции.

Методы интеллектуального анализа данных

В отличие от традиционных систем обнаружения вторжений, которые опираются на известные шаблоны атак, при обнаружении атак нулевого дня используются современные алгоритмы машинного обучения для анализа

данных сетевого трафика и выявления аномального поведения, которое может указывать на новую или неизвестную атаку.

Существующие подходы к борьбе с такими атаками основаны на машинном обучении/глубоких нейронных сетях (machine learning/deep neural networks - ML/DNN) или обнаружении аномалий.

Двумя ключевыми ограничениями рассмотренных предыдущих подходов являются формирование сигналов о неизвестных действиях и ложных тревогах при аномальном поведении. Для решения этих проблем обычно предлагается новая методика анализа и обнаружения атак "нулевого дня", которая позволяет обнаружить эксплойты "нулевого дня" путем исследования сети организации и мониторинга поведенческой активности эксплойтов "нулевого дня" на каждом этапе их жизненного цикла. Чтобы обнаружить наличие эксплойта нулевого дня, в работе [9] предлагается система на основе машинного обучения для анализа сетевого трафика и выявления неожиданного поведения сети. Предложенная система сочетает подходы контролируемой классификации для анализа существующих классов с гибкостью неконтролируемой классификации для раскрытия новых аспектов классификации.

В работе [10] уменьшение количества ошибок классификации повышает производительность моделей машинного обучения bagging и boosting. В любой модели ML-предсказания для оценки истинности вклада анализируемых признаков атак оцениваются значения Шэпли, которые помогают в распознавании лучших признаков с помощью подхода аддитивных объяснений Шэпли (Shapley Additive Explanations — SHAP). SHAP — это популярный метод описания глубоких нейронных сетей, который предоставляет аналитические сведения о вкладе каждой входной функции в заданный прогноз.

Значения Шэпли преобразуются в шкалу вероятностей, чтобы соотнести их со значением предсказания ML-модели и определить лучшие признаки для каждого предсказания, сделанного обученной ML-моделью.

При расчёте вектора Шепли необходимо формировать коалиции из ограниченного набора признаков. Однако, не каждая ML-модель позволяет просто убрать признак без повторного обучения модели «с нуля». Потому для формирования коалиций обычно не убирают «лишние» признаки, а заменяют их на случайные значения из «фонового» набора данных. Усреднённый результат модели со случайными значениями признака эквивалентен результату модели, в которой этот признак вообще отсутствует [6].

Тенденция топ-атрибутов, полученная из ложноотрицательных и ложноположительных предсказаний обученной ML-модели, может быть использована для создания индуктивных правил в зависимости от шкалы вероятности признаков. Обнаруживая ложноотрицательные вредоносные программы "нулевого дня", данный подход помогает повысить уровень кибербезопасности.

Для обнаружения атак нулевого дня в сетях IoT с помощью классификации сетевого трафика было предложено множество моделей.

В [11] предложили надежную модель для обнаружения IoT-ботнетов, основанную на системном обучении с использованием признаков атак в сочетании с методами контролируемого обучения. Экспериментальные исследования с применением различных наборов данных показали, что модель способна выявлять как атаки "нулевого дня", так и IoT-ботнеты.

В [12] предложены новые подходы к использованию автоматического и глубокого обучения для выявления неизвестных угроз в захваченном трафике IoT, что часто называют обнаружением аномалий. С этой целью данные PCAP сначала преобразуются в RGB-изображения, что облегчает их анализ и

позволяет выбрать наиболее подходящий алгоритм машинного или глубокого обучения с контролем для достижения наилучших результатов. Целью исследования является определение лучшего алгоритма контролируемого обучения путем тестирования и сравнения семи различных алгоритмов с различным уровнем сложности. Эти алгоритмы были разделены на две категории:

- Обычные классификаторы: метод ближайших соседей (k-Nearest Neighbors — k-NN), машина опорных векторов (Support Vector Machine — SVM) и логистическая регрессия (Logistic Regression — LR).
- Глубокие классификаторы: сверточные нейронные сети (Convolutional Neural Networks — CNN) с двумя и четырьмя слоями, а также модели VGG16 и MobileNetv2.

Результаты тестирования показали, что алгоритм на основе SVM продемонстрировал наивысшую эффективность, достигнув точности 94%. Этот результат может быть объяснен его способностью к быстрому и стабильному обучению при меньших вычислительных затратах.

В [13] для более быстрого и точного обнаружения уязвимостей "нулевого дня" с минимальным уровнем ложных срабатываний предложен подход на основе машинного обучения. В работе атаки "нулевого дня" включены в набор данных CICIDS, который обрабатывается в потоковом режиме. Для оценки производительности системы использовались такие параметры, как надежность, время обнаружения и объем используемой памяти. Система была реализована с применением методов случайного леса, случайного дерева, байесовского классификатора и дерева Хеффдинга. Результаты показали, что дерево Хеффдинга обеспечивает точность 99,97% при времени обработки 5,94 секунды и использовании всего 0,08 МБ памяти. Это делает его наиболее эффективным из предложенных методов.

Система обнаружения вторжений, представленная в [14], использует методы трансферного обучения и уточнения моделей для повышения точности обнаружения в ограниченных и несбалансированных наборах данных.

Система обнаружения вторжений в [14] демонстрирует отличную точность и низкий процент ложных срабатываний даже для новых семейств атак "нулевого дня". Использование трансферного обучения и тонкой настройки сети повышает уровень обнаружения и превосходит предыдущие системы обнаружения вторжений на основе глубокого обучения.

Система обнаружения вторжений, предложенная в [14], может быть расширена для учета реальных данных из сетей IoT, что позволит более полно оценить ее эффективность и устойчивость.

В [15] разработали систему обнаружения вторжений, способную идентифицировать как неизвестные кибератаки, так и атаки "нулевого дня". Они предложили усовершенствованную модель обнаружения вторжений на основе автоэнкодера. Этот подход отличается своей инновационностью, подчеркивая важность выбора оптимального порога для эффективного выявления атак "нулевого дня". Кроме того, авторы отмечают, что использование единого критерия для определенного типа атак может быть недостаточно эффективным для обнаружения других, менее очевидных угроз. Чтобы показать значимость каждого вида атаки, была проведена независимая оценка точности для каждой из них с использованием различных критериев. Для проверки эффективности модели был использован актуальный набор данных CICIDS2017. Оптимизированная версия автокодировщика показала отличные результаты, достигая общей точности 99,29%. Модель продемонстрировала высокую производительность как при индивидуальном анализе, так и на совокупном уровне, обеспечивая надежное обнаружение кибератак.

В [16] разработана структурированная схема, направленная на создание адаптивной системы киберзащиты от угроз "нулевого дня" с использованием обучения с подкреплением. Этот подход объединяет элементы теории управления и методов машинного обучения. Одним из ключевых преимуществ предложенной схемы является использование вознаграждения в процессе обучения, что позволяет системам защиты обходиться без знания конкретных деталей атак "нулевого дня", таких как цели атак и расположение уязвимых узлов. Получение таких сведений заранее зачастую крайне затруднительно или вовсе невозможно.

Схема обучения с подкреплением позволяет эффективно противостоять различным типам атак, включая:

- Стратегические атаки, где атакующий и защищающийся участвуют в некооперативной игре;
- Нестратегические случайные атаки, при которых действия атакующего определяются заданным распределением вероятностей;
- Байесовские графы атак, в которых атакующий компрометирует сетевые узлы, используя известные уязвимости или уязвимости "нулевого дня".

Таким образом, предложенная схема демонстрирует высокую адаптивность и эффективность в условиях ограниченности информации об атаках "нулевого дня".

В [17] предложена модель вариационного автокодировщика (Variational Auto Encoder — VAE) для обнаружения атак "нулевого дня" с использованием глубокого обучения (Deep Learning — DL). Основной целью проекта было создание IDS с минимальным уровнем ложноотрицательных срабатываний и высокой точностью. Для обеспечения корректной работы модели автокодировщика требуется предварительная обработка необработанных данных, чтобы они могли быть корректно

интерпретированы алгоритмами глубокого обучения. После этого модель VAE применяется для поиска уязвимостей "нулевого дня" в сетевых данных. Эффективность модели автокодировщика была подтверждена в ряде исследований, где результаты анализировались с различных точек зрения. По результатам моделирования, модель автокодировщика продемонстрировала высокие показатели: коэффициент каппа составил 0,973, F-score — 0,982, точность — 0,989, специфичность — 0,977, а чувствительность — 0,985.

В [18] разработали современную методику, известную как переданная глубоко-конволюционная генеративная состязательная сеть (Transferred Deep-Convolutional Generative Adversarial Network — tDCGAN), предназначенную для различения настоящего и поддельного вредоносного программного обеспечения. Несмотря на различия между образцами вредоносного ПО, данные, полученные из случайного распределения, имеют множество сходств с реальными данными. Модель tDCGAN, основанная на глубоком автокодировщике (Deep Autoencoder — DAE), позволяет выделять значимые характеристики и стабилизировать процесс обучения генеративной состязательной сети (Generative Adversarial Network — GAN). DAE играет ключевую роль на начальном этапе, собирая общие данные, выявляя признаки вредоносного ПО и передавая эту информацию генератору GAN. Такой подход обеспечивает постепенное и точное обучение GAN до этапа генерации результатов. Дискриминатор, обученный на основе GAN, передает детектору возможность определять характеристики вредоносного ПО, используя процедуру обучения с переносом. Экспериментальные результаты показали, что tDCGAN достигла среднего показателя классификации 95,74%, демонстрируя превосходство над другими моделями как в устойчивости обучения, так и в защите от симулированных атак "нулевого дня".

Выводы

Атаки "нулевого дня" представляют собой серьезную угрозу безопасности, особенно в условиях быстрого роста Интернета вещей (IoT), который становится все более уязвимым для кибератак. Разрабатываемые для защиты IoT-приложений системы обнаружения вторжений, обычно ориентируются на известные атаки. Однако атаки "нулевого дня" — те, которые еще не были выявлены СОВ, — остаются значительным вызовом, вызывая серьезные опасения по поводу безопасности и конфиденциальности пользовательских данных в IoT-приложениях.

Одной из ключевых проблем является необходимость передачи больших объемов разрозненных данных сетевого трафика IoT на удаленные центральные облачные серверы для обработки.

Подходы, основанные на централизованном глубоком обучении, требуют значительных ресурсов, включая большой объем памяти для хранения данных, длительное время обучения и высокие затраты на коммуникацию. Кроме того, облачные центры обработки данных зачастую расположены далеко от мест установки IoT-устройств. Это приводит к значительным задержкам в работе систем обнаружения ботнетов, построенных на базе централизованного глубокого обучения, что ограничивает их эффективность в реальных условиях.

Будущие исследования в этой области будут сосредоточены на применении усовершенствованных техник глубокого обучения, которые смогут более эффективно адаптироваться к новым угрозам и минимизировать уровень ложных срабатываний, что позволит сделать СОВ более надежными и эффективными в условиях современных киберугроз.

Литература (References)

1. Saharkhizan M., Azmoodeh A., Dehghantanha A., Choo K.K.R. and Parizi R.M. An ensemble of deep recurrent neural networks for detecting IoT cyber-attacks using network traffic. IEEE Internet of Things Journal. 2020. 7. 9. Pp. 8852-8859.

2. Huong T.T., Bac T.P., Long D.M., Thang B.D., Binh N.T., Luong T.D. and Phuc T.K. Lockedge: Low complexity cyberattack detection in IoT edge computing. IEEE Access. 2021. 9. Pp. 29696-29710.

3. Eskandari M., Janjua Z.H., Vecchio M. and Antonelli F. Passban IDS: An intelligent anomaly-based intrusion detection system for IoT edge devices. IEEE Internet of Things Journal. 2020. 7. 8. Pp. 6882-6897.

4. Diro A.A. and Chilamkurti N. Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems. 2018. 82. Pp. 761-768.

5. Rosa T.M., Santin A.O. and Malucelli A. Mitigating XML Injection 0-Day Attacks through Strategy-Based Detection Systems. IEEE Security & Privacy. 2013. 11. 4. Pp.46-53.

6. Sharma V., Kim J., Kwon S., You I., Lee K. and Yim K. A framework for mitigating zero-day attacks in IoT. Conference on Information Security and Cryptography (CISC-S'17). 2017. Pp. 1-6.

7. Sun X., Dai J., Liu P., Singhal A. and Yen J. Towards probabilistic identification of zero-day attack paths. 2016 IEEE Conference on Communications and Network Security (CNS). 2016. Pp.64-72.

8. Li Z., Wang L., Chen Y. and Fu Z. Network-based and Attack-resilient Length Signature Generation for Zero-day Polymorphic Worms. 2007 IEEE International Conference on Network Protocols. 2007. Pp.164-173.

9. Singh U.K., Joshi C. and Singh S.K. Zero-day Attacks Defense Technique for Protecting System against Unknown Vulnerabilities. International Journal of Scientific Research in Computer Science and Engineering. 2017. 5. 1. Pp.13-18.

10. Rajesh K. and Geetha S. Zero-Day Malware Detection and Effective Malware Analysis Using Shapley Ensemble Boosting and Bagging Approach. Sensors. 2022. 22. 7.

11. Ngo Q.D. and Nguyen Q.H. A Reinforcement Learning-Based Approach for Detection Zero Day Malware Attacks on IoT System. Computer Science Online Conference. Cham. Springer International Publishing. 2022. Pp. 381-394.

12. El-Sayed R., El-Ghamry A., Gaber T. and Hassanien A.E. Zero-day malware classification using deep features with support vector machines. 2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS). 2021. Pp. 311-317.

13. Seraphim B.I. and Poovammal E. Zero-Day Attack Detection Analysis in Streaming Data Using Supervised Learning Techniques. Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021. Singapore. Springer Nature Singapore. 2022. Pp. 517-530.

14. Rodríguez E., Valls P., Otero B., Costa J.J., Verdu J., Pajuelo M.A. and Canal R. Transfer-Learning-Based Intrusion Detection Framework in IoT Networks. Sensors. 2022. 22. 15. URL:doi.org/10.3390/s22155621.

15. Roshan K. and Zafar A. An Optimized Auto-Encoder based Approach for Detecting Zero Day Cyber-Attacks in Computer Network. 2021 5th International Conference on Information Systems and Computer Networks (ISCON). 2021. Pp. 1-6.

16. Anand P., Singh Y. and Selwal A. Learning-based techniques for assessing zero-day attacks and vulnerabilities in IoT. Recent Innovations in

Computing: Proceedings of ICRIC 2021. Singapore. Springer Singapore. 2022. Pp. 497-504.

17. Hu Z., Chen P., Zhu M. and Liu P. Reinforcement learning for adaptive cyber defense against zero-day attacks. Adversarial and Uncertain Reasoning for Adaptive Cyber Defense: Control and Game Theoretic Approaches to Cyber Security. 2019. Pp. 54-93.

18. Priya S. and Uthra R. An Effective Deep Learning-Based Variational Autoencoder for Zero Day Attack Detection Model. Inventive Systems and Control: Proceedings of ICISC 2021. Singapore. Springer Singapore. 2021. Pp. 205-212.

Дата поступления: 21.01.2025
Дата публикации: 4.03.2025