УДК 004.056.53

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ В СЕТЕВОМ ТРАФИКЕ: ПОДХОДЫ, ОГРАНИЧЕНИЯ И ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ

Шлянов Д.А.[1]

студент,

Поволжский государственный университет телекоммуникаций и информатики,

Самара, Россия

Аннотация

Статья посвящена анализу современных методов машинного обучения, применяемых для обнаружения аномалий и вторжений в сетевом трафике. Рассматриваются три группы подходов: сигнатурные системы обнаружения вторжений, классические алгоритмы (метод опорных векторов, случайный лес, наивный байесовский классификатор, метод k-ближайших соседей) и модели глубокого обучения (многослойный персептрон, свёрточные и рекуррентные нейронные сети, включая LSTM и GRU). Обсуждаются практические ограничения подходов: высокая доля ложноположительных срабатываний, зависимость качества моделей от репрезентативности обучающих датасетов, уязвимость к состязательным атакам и сложности эксплуатации в реальном времени. Формулируются принципы практического применения рассмотренных методов в системах сетевой безопасности.

[1] Научный руководитель: **Куваева Е.Н**., ст. преподаватель, Поволжский государственный университет телекоммуникаций и информатики, Самара, Россия.
Kuvaeva E.N., Senior Lecturer, Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia.

Ключевые слова: машинное обучение, обнаружение аномалий, сетевой трафик, обнаружение вторжений, глубокое обучение, информационная безопасность, кибербезопасность.

MACHINE LEARNING METHODS FOR NETWORK TRAFFIC ANOMALY DETECTION: APPROACHES, LIMITATIONS AND APPLICATION PROSPECTS

Shlyanov D.A.

Student,

Povolzhskiy State University of Telecommunication and Informatics,

Russia, Samara

Abstract

The article analyzes modern machine learning methods used for anomaly and intrusion detection in network traffic. Three groups of approaches are examined: signature-based intrusion detection systems, classical algorithms (support vector machines, random forest, naive Bayes classifier, k-nearest neighbors) and deep learning models (multilayer perceptron, convolutional and recurrent neural networks, including LSTM and GRU). Practical limitations are discussed: high false-positive rates, dependence of model quality on the representativeness of training datasets, vulnerability to adversarial attacks, and challenges of real-time deployment. Principles for the practical application of the reviewed methods in network security systems are proposed.

Keywords: machine learning, anomaly detection, network traffic, intrusion detection, deep learning, information security, cybersecurity.

Стремительный рост объёмов сетевого трафика и усложнение атак на инфраструктуру - от DDoS до целенаправленных вторжений - делают традиционные сигнатурные средства защиты недостаточными: они эффективно блокируют уже известные угрозы, но плохо справляются с новыми и видоизменёнными атаками [5]. Это стимулирует применение методов машинного обучения для построения систем обнаружения аномалий, способных выявлять отклонения от нормального поведения сети без опоры на заранее известные сигнатуры [1]. Тем не менее внедрение подобных систем в промышленную эксплуатацию сталкивается с рядом технических ограничений, обзору которых посвящена данная статья. Цель работы - систематизировать основные группы методов машинного обучения, применяемых для обнаружения сетевых аномалий, и проанализировать их практические ограничения.

Системы обнаружения вторжений принято делить на сигнатурные и поведенческие (аномальные). Сигнатурные системы, классическим примером которых является Snort, сопоставляют трафик с базой известных шаблонов атак и обеспечивают низкий уровень ложных срабатываний, но бессильны против ранее неизвестных угроз [5]. Поведенческие системы строят модель «нормального» состояния сети и помечают как аномальные любые существенные отклонения от неё; именно в этом классе систем находят применение методы машинного обучения [1].

Среди классических алгоритмов машинного обучения для обнаружения сетевых аномалий наиболее востребованы метод опорных векторов, случайный лес, наивный байесовский классификатор, метод k-ближайших соседей и градиентный бустинг [3]. Эти методы требуют предварительного извлечения признаков из сетевых потоков (длительность соединения, объём переданных байт, частота пакетов и т. д.) и, как правило, показывают приемлемое качество

классификации при относительно небольших вычислительных затратах, что делает их привлекательными для практического внедрения.

Модели глубокого обучения - многослойный персептрон, свёрточные нейронные сети, рекуррентные нейронные сети и их модификации (LSTM, GRU) - позволяют работать непосредственно с последовательностями сырых данных трафика, автоматически извлекая признаки без ручного конструирования [6]. Комбинированные свёрточно-рекуррентные архитектуры показывают, что совместное использование пространственных и временных зависимостей повышает точность обнаружения по сравнению с отдельно взятыми моделями, однако требует существенно больших вычислительных ресурсов и объёмов размеченных данных для обучения.

Значимой проблемой остаётся качество и репрезентативность обучающих наборов данных. Наиболее широко используемый в исследованиях датасет KDD Cup 99 и его усовершенствованная версия NSL-KDD страдают от несбалансированности классов и устаревшей структуры атак, что ставит под сомнение применимость обученных на них моделей к современному трафику [4]. Использование устаревших датасетов приводит к завышенным и не всегда воспроизводимым в реальных условиях показателям качества моделей.

Качество моделей принято оценивать по метрикам точности, полноты, аккуратности и F-меры, однако именно доля ложноположительных срабатываний (False Positive Rate) остаётся ключевым практическим ограничением: даже при формально высокой точности классификации абсолютное число ложных тревог на реальном объёме трафика крупной сети может исчисляться тысячами в сутки, что приводит к усталости операторов и снижению доверия к системе [3].

Отдельную угрозу представляют состязательные атаки, при которых злоумышленник целенаправленно изменяет параметры трафика (тайминги, размер пакетов, фрагментацию) таким образом, чтобы остаться в пределах

«нормального» поведения с точки зрения модели, оставаясь незамеченным [6]. Устойчивость моделей машинного обучения к подобным целенаправленным манипуляциям существенно ниже, чем к случайному шуму, что систематически отмечается в работах по безопасности систем на основе машинного обучения [8].

Дополнительным ограничением служит рост доли зашифрованного трафика, при котором классические признаки на основе содержимого пакетов недоступны, а модели вынуждены опираться исключительно на метаданные потоков (объём, тайминги, порядок пакетов) [8]. В сочетании с требованиями к обработке трафика в режиме реального времени это накладывает жёсткие ограничения на сложность архитектур, применимых в промышленных системах.

Соммер и Пэксон [2] обращают внимание на фундаментальное отличие задачи обнаружения вторжений от типичных задач машинного обучения: в сетевой безопасности отсутствует стабильное понятие «нормы», среда нестационарна, а последствия ошибок первого и второго рода несопоставимы по цене. Это объясняет, почему многие модели, демонстрирующие высокие метрики на тестовых датасетах, показывают существенно худшие результаты при развёртывании в реальной инфраструктуре.

По мере эволюции сетевой инфраструктуры и появления новых типов легитимных сервисов статистические характеристики «нормального» трафика со временем смещаются (concept drift), что приводит к постепенной деградации качества обученных моделей и требует их регулярного переобучения на актуальных данных [1].

Существенным препятствием для внедрения моделей глубокого обучения в системы сетевой безопасности остаётся низкая интерпретируемость их решений: оператору сложно понять, какие именно признаки трафика

послужили основанием для срабатывания, что затрудняет как расследование инцидентов, так и доверие персонала к автоматизированной системе [8].

Ряд исследователей указывает, что автоматические детекторы аномалий эффективнее работают не как самостоятельное решение, а как инструмент приоритизации алертов для аналитика центра мониторинга информационной безопасности (SOC), при котором финальное решение остаётся за человеком [2].

Сбор и хранение сетевого трафика для целей обучения и эксплуатации моделей поднимает вопросы конфиденциальности пользовательских данных и соответствия требованиям законодательства о персональных данных, особенно при передаче трафика в облачные системы анализа [3].

На основании проведённого анализа можно сформулировать четыре принципа практического применения методов машинного обучения в задачах обнаружения сетевых аномалий. Принцип 1 - комбинирование подходов: сигнатурные и поведенческие методы должны применяться совместно, компенсируя ограничения друг друга [5]. Принцип 2 - регулярное переобучение моделей на актуальных данных с учётом эффекта смещения распределения. Принцип 3 - человек в контуре принятия решений: результаты работы модели должны рассматриваться как приоритизированные алерты для аналитика, а не как автоматический вердикт. Принцип 4 - оценка устойчивости к состязательным воздействиям на этапе разработки, а не только после внедрения системы в эксплуатацию.

Методы машинного обучения существенно расширяют возможности обнаружения аномалий в сетевом трафике по сравнению с исключительно сигнатурными подходами, однако не заменяют их полностью. Практическая эффективность подобных систем определяется не только точностью отдельно взятой модели, но и качеством обучающих данных, устойчивостью к целенаправленным атакам и продуманной интеграцией в процессы

мониторинга с участием специалистов по информационной безопасности. Дальнейшее развитие области связано с созданием адаптивных моделей, устойчивых к смещению распределения данных, и с разработкой интерпретируемых архитектур, пригодных для использования в реальных SOC.

Библиографический список

1. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey // ACM Computing Surveys. – 2009. – Т. 41, № 3. – Article 15.
2. Sommer R., Paxson V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection // Proceedings of the 2010 IEEE Symposium on Security and Privacy. – 2010. – P. 305–316.-
3. Buczak A.L., Guven E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection // IEEE Communications Surveys & Tutorials. – 2016. – Т. 18, № 2. – P. 1153–1176.
4. Tavallaee M., Bagheri E., Lu W., Ghorbani A.A. A Detailed Analysis of the KDD CUP 99 Data Set // Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA). – 2009. – P. 1–6.-
5. Roesch M. Snort: Lightweight Intrusion Detection for Networks // Proceedings of the 13th USENIX Conference on System Administration (LISA). – 1999. – P. 229–238.
6. Ahmad R., Alsmadi I. Machine Learning Approaches to IoT Security: A Systematic Literature Review // Internet of Things. – 2021. – Т. 14. – Article 100365.-
7. Chalapathy R., Chawla S. Deep Learning for Anomaly Detection: A Survey // arXiv preprint arXiv:1901.03407. – 2019.
8. Ruff L., Kauffmann J.R., Vandermeulen R.A. et al. A Unifying Review of Deep and Shallow Anomaly Detection // Proceedings of the IEEE. – 2021. – Т.