DOI 10.37882/2223-2966.2025.02-2.19

АНАЛИЗ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ В СИСТЕМАХ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ

ANALYSIS OF THE EFFECTIVENESS OF MACHINE LEARNING MODELS IN INTRUSION DETECTION SYSTEMS

O. Karelova
O. Kostrova
K. Kurbanova

Summary. This article examines the classification of intrusion detection systems (IDS) and the effectiveness of applying various machine learning and deep learning algorithms in these systems. The purpose of intrusion detection systems, their main functions, methods for detecting intrusions, types, and operating principles are also described. Statistics on the performance of traditional intrusion detection systems based on open-source solutions are provided, as well as an examination of the effectiveness of applying different machine learning and deep learning algorithms in detecting various types of attacks on network infrastructures.

Keywords: intrusion detection system, attack, machine learning, traffic analysis, Suricata, Snort.

Карелова Оксана Леонидовна
Доктор физико-математических наук, доцент, профессор, Московский Государственный Лингвистический Университет;
Профессор, Российская академия народного хозяйства и государственной службы при Президенте РФ (Москва),
okarelova@yandex.ru
Кострова Ольга Евгеньевна
Институт информационных наук ФГБОУ ВО МГЛУ
olkeruac@gmail.com
Курбанова Камиля Маратовна
Институт информационных наук ФГБОУ ВО МГЛУ
kamila.business03@gmail.com

Аннотация. В статье рассматривается классификация систем обнаружения вторжений (СОВ/IDS) и эффективность применения в таких системах различных алгоритмов машинного и глубокого обучения. Описано назначение систем обнаружения вторжений, их основные функции, методы выявления вторжений, типы и принципы работы. Приведена статистика эффективности работы традиционных систем обнаружения вторжений на базе open-source решений, а также рассмотрена эффективность применения различных алгоритмов машинного и глубокого обучения в обнаружении разных типов атак на сетевые инфраструктуры.

Ключевые слова: система обнаружения вторжений, атака, машинное обучение, анализ трафика, Suricata, Snort.

Введение

С развитием технологий связи, все больший объем различных и разнородных данных передается через сетевые системы. Как правило, эти данные поступают из различных источников, таких как датчики, компьютеры и Интернет вещей (IoT). По мере того, как расширяется область применения передающих устройств, расширяется и зона атаки, что делает сетевые системы более уязвимыми для потенциальных угроз. Методы кибератак становятся все более сложными и изощренными, а также возрастает их частота: во II квартале 2024 года зафиксировано в 2,6 раза больше атак по сравнению с аналогичным периодом в 2023 году [1].

Одной из важнейших задач в области кибербезопасности является обнаружение сетевых угроз. В последнее время многие исследования сосредоточены на применении технологии искусственного интеллекта (ИИ) в системах обнаружения сетевых вторжений.

В ИСО МЭК 18028-1-2008 система обнаружения вторжений (СОВ) (intrusion detection system — IDS) определяется, как техническая система, используемая для идентификации того, что была предпринята попытка вторжения, что вторжение происходит или произошло, а также для возможного реагирования на вторжения в информационные системы и сети. В данной статье рассмотрены виды СОВ и сравниваются методы применения глубокого и машинного обучения в данной области.

Классификация традиционных систем обнаружения вторжений

Система предотвращения и обнаружения вторжений (IPS/IDS) представляет собой совокупность программных и аппаратных средств, предназначенных для выявления и предотвращения несанкционированного доступа к сетевой инфраструктуре. Данные системы можно разделить на две основные категории: систему обнаружения вторжений (далее — СОВ или IDS) и систему предотвращения вторжений (далее — СПВ или IPS).

Ключевые функции IDS включают в себя выявление вторжений и атак, определение их источника, регистрацию инцидентов, анализ уязвимостей, прогнозирование атак, а также оперативное информирование ответственных должностных лиц и формирование отчетов.

Основной механизм действия IDS заключается в анализе сетевого трафика для поиска потенциальных угроз, хотя специфика анализа может различаться. Наиболее распространенными методами являются сигнатурный и поведенческий анализ. Сигнатурные IDS функционируют по принципу, аналогичному антивирусному программному обеспечению: они сопоставляют сигнатуры угроз с обновляемой базой данных для идентификации атак на информационную систему, минимизируя количество ложных срабатываний. Тем не менее, этот метод имеет ограничение: он не может обнаружить атаки, сигнатуры которых отсутствуют в базе.

В отличии от сигнатурных методов, поведенческий анализ основан на моделировании «нормального» функционирования системы. Этот метод опирается на выявление отклонений от эталонного режима, рассматривая любые несоответствия как потенциальные вторжения или аномалии.

Преимущества поведенческих методов заключаются в их способности обнаруживать атаки без знания конкретных сигнатур и высокой чувствительности к изменениям в состоянии системы. Однако, возможны частые ложные срабатывания, особенно в условиях реального сетевого поведения, что также требует временных затрат на этап обучения. Кроме того, нельзя не подчеркнуть, что задача создания эталонной модели представляется трудоемкой и комплексной.

Функции IPS заключаются в активном блокировании атак, прекращении доступа к хостам и изменении конфигурации сети для предотвращения атак, а также фильтрации инфицированных файлов.

Поскольку функциональные возможности IPS не обеспечивают обнаружение внешних и внутренних атак в реальном времени, их почти не имеет смысла использовать без IDS. На современном рынке практически не осталось исключительно IPS-решений, вместо этого предлагаются системы IDPS (Intrusion Detection and Prevention Systems), которые одновременно выявляют атаки и выполняют предустановленные действия, такие как Pass, Alert, Drop и Reject.

Системы обнаружения вторжений в широком смысле подразделяются на категории в зависимости от положения датчиков IDS: в сети или на хосте.

Сетевая система обнаружения вторжений (NIDS) отслеживает и анализирует сетевой трафик на предмет подозрительного поведения и реальных угроз с помощью датчиков NIDS, которые анализируют информацию о заголовках всех пакетов, передаваемых по сети.

Датчики NIDS устанавливаются в критических точках сети для проверки трафика со всех устройств в сети, например, в подсети, где расположены брандмауэры, для обнаружения отказа в обслуживании (DoS) и других подобных атак.

Система обнаружения вторжений на базе хоста (HIDS) отслеживает и анализирует конфигурацию системы и активность приложений на устройствах, работающих в корпоративной сети. Датчики HIDS могут быть установлены на любом устройстве, независимо от того, является ли это настольным ПК или сервером.

Датчики HIDS фиксируют существующие системные файлы и сравнивают их с ранее зафиксированными файлами. Они отслеживают нестандартные изменения, такие как перезапись, удаление и доступ к определенным портам. В результате администраторам отправляются уведомления для расследования действий, которые кажутся подозрительными.

Одним из недостатков систем обнаружения вторжений (IDS) является возможность замедления работы сети или отдельных устройств. Из-за ресурсозатратных процессов, связанных с анализом состояния сети, может происходить снижение производительности системы. Кроме того, если говорить о недостатках эффективности анализа всей сети, это в первую очередь касается хостовых систем (HIDS). Такие системы могут не зафиксировать изменения в трафике, которые могут оказаться безопасными для конкретного устройства, но при этом представлять серьезную угрозу для функционирования всей сети.

Применение ИИ в системах обнаружении вторжений

Чистых IDS с модулями искусственного интеллекта в данный момент на рынке не представлено, но есть решения, включающие в себя эту технологию, например, XDR (Extended detection and response — система обнаружения угроз и реагирования и NDR (Network Detection and Response — сетевое обнаружение и реагирование). Среди таких решений можно выделить Darktrace NDR, Cisco XDR, Palo Alto Networks Cortex XDR. На российском рынке же отметим Kaspersky Next XDR Expert.

Однако по этим решениям не представлена в открытом доступе статистика по точности обнаружения различных атак, поэтому далее мы будем опираться на исследования по эффективности сигнатурных open-source IDS Snort и Suricata, а также отдельных алгоритмов машинного обучения.

Например, Mehak Usmani, Misbah Anwar и др. удалось провести эксперимент с обучением рекуррентных нейронных сетей долгой краткосрочной памяти (LSTM, DT driven LSTMl) прогнозированию ARP-spoofing атак, в ходе которого модели предсказывали атаки с точность 99,9 % и 100 % соответственно [2].

Конечно, спектр моделей, обучаемых с целью выявления вторжений в сетях, гораздо шире. В него входят такие модели, как деревья решений (далее — Decision tree, DT), Метод случайного леса (Random forest, RF), Метод k-ближайших соседей (k-nearest neighbors, KNN), Наивный байесовский классификатор (Naive Bayes, NB), Метод опорных векторов (Support vector machine, SVM), Многослойный перцептрон (Multilayer perceptron, MLP), Байесовские сети (BayesNet), Нечеткая логика (Fuzzy Logic). Разные модели по-разному проявили себя в экспериментах в датасетах с различными типами атак (Таблица 1).

Таблица 1.

Точность обнаружения DDoS, R2L, U2R различными моделями, предложенными в исследованиях

Алгоритм МО; автор	DDoS	R2L, U2R
Самоорганизующаяся карта Кохонена; Braga, R.; Mota, E.; Passito, A. [3]	98 %	ND
Самоорганизующаяся карта Кохонена; Ibrahim, L.M.; Basheer, D.T.; Mahmod, M.S. [4]	75,49	75,49 %
SVM; Kokila, R.; Selvi, S.T.; Govindarajan, K. [5]	95 %	ND
Naive Bayes; Mukherjee, S.; Sharma, N. [6]	98,7 %	64 % (U2R), 96 % (R2L)
Self-taught Deep Learning; Javaid, A.; Niyaz, Q.; Sun,W.; Alam, M. [7]	ND	92,98 %
J48, Naive Bayesian, Random Forest, Multi-layer Perceptron, Support Vector Machine; Yin, C.; Zhu, Y.; Fei, J.; He, X. [8]	83,28 %	83,28 %
Глубокое обучение; Tuan Anh Tang, Lotfi Mhamdi, Des McLernon, Syed Ali Raza Zaidi, Mounir Ghogho [9]	ND	80,7 %
Gated Recurrent Neural Network; Tuan Anh Tang, Lotfi Mhamdi, Des McLernon, Syed Ali Raza Zaidi, Mounir Ghogho	ND	89 %

Исходя из данных Таблицы 1, самыми эффективными моделями для обнаружения DDoS-атак стал Метод опорных векторов (Support vector machine, SVM), а для атак R2L, U2R — STDL.

Далее будут использоваться следующие аббревиатуры:

FPR — Частота ложноположительных срабатываний (False Positive Rate)

FNR — Частота ложноотрицательных срабатываний (False Negative Rate)

TPR — Частота истинно положительных срабатываний (True Positive Rate)

DR – Частота обнаружения (Detection Rate)

DA – Точность обнаружения (Detection Accuracy)

В исследовании Syed Ali Raza Shah и Biju Issac [10] в ходе одного из экспериментов следующие алгоритмы машинного обучения показали результаты, представленные в Таблице 2.

Таблица 2.

Точность обнаружения атак различными алгоритмами машинного обучения

MAC Spoofing, DNS Poisoning, IP Spoofing			
Алгоритм машинного обучения	DR, %	FPR, %	DA, %
Support Vector Machines	96,8	0,7	95,6
Decision Trees	79,2	2,9	82
Fuzzy Logic	94,5	0,2	92,3
BayesNet	65	3,5	73
NaiveBayes	62	3	70
Атаки на SSH, FTP, сканирование			
Алгоритм машинного обучения	DR, %	FPR, %	DA, %
Support Vector Machines	97	0,5	94,2
Decision Trees	81,1	1,9	85
Fuzzy Logic	92	1,6	94
BayesNet	63	5,1	71,2
NaiveBayes	65	6	71
Отказ в обслуживании (DoS), Повышение привилегий (U2R), Несанкционированный удаленный доступ (R2L), Атаки по сторонним каналам			
Алгоритм машинного обучения	DR, %	FPR, %	DA, %
Support Vector Machines	97,3	3,1	95,4
Decision Trees	78	10	81,2
Fuzzy Logic	95	4	94
BayesNet	69	8	74
NaiveBayes	70	7,6	79

По данным из Таблицы 2, самую высокую точность обнаружения для всех приведенных типов атак демонстрируют SVM и Fuzzy Logic.

В том же исследовании на примере Snort и Suricata в одном из экспериментов анализировалась точность обнаружения вредоносных пакетов при обработке трафика. Результаты представлены в Таблице 3.

Таблица 3.
Точность обнаружения вредоносных пакетов Suricata и Snort

Трафик	Snort			Suricata		
	FPR, %	FNR, %	TPR, %	FPR, %	FNR, %	TPR, %
UDP	11	0	0	23	3	0
TCP	10	0	0	32	9	0
ICMP	3	0	0	39	27	3

По приведенным в табл. 3 данным, Snort произвел меньше ложноположительных сигналов. Хотя ложноотрицательные срабатывания наблюдались в обоих IDS, в исследовании было установлено, что точность обнаружения Snort в эксперименте была выше, чем у Suricata

В другом эксперименте проводился анализ Snort и Suricata с гибридным плагином SVM и Fuzzy Logic, а также с оптимизированным SVM. Результаты представлены в Таблице 4.

Таблица 4.
Обнаружение различных атак с помощью Snort с плагином SVM и Fuzzy Logic, и Snort с плагином оптимизированного SVM

Вредоносный трафик	Snort с SVM и Fuzzy Logic, %		Snort с оптимизированным SVM, %	
	FPR	FNR	FPR	FNR
SSH	2	0	1,6	0,1
DoS/DDoS	1	0,5	1	0,2
FTP	3	0,5	2	0,2
HTTP	2	1	1,5	0,9
ICMP	2	0,7	1	0,5
ARP	2	0	1	0
Scan	1	0,5	0,5	0,3
Всего	13	3,2	8,6	2,2

Таким образом, частота ложноположительных и ложноотрицательных срабатываний значительно снизилась.

В исследовании А.И. Гетьмана, М.Н. Горюнова, А.Г. Мацкевича, Д.А. Рыболовлева [11] также сопоставлялись IDS на основе машинного обучения с сигнатурными IDS на примере фаерволла WAF ModSecurity, IDS Suricata и IDS на основе машинного обучения (далее — ML COB).

Результаты эксперимента показали, что NIDS Suricata не распознает атаки, связанные с внедрением и эксплуатацией Shell-кода в зашифрованном HTTPS-трафике. WAF ModSecurity и ML COB показывают схожие результаты в обнаружении этих атак, но ML COB оказывается более эффективным при выявлении попыток загрузки Shell-кода, обнаруживая атаки на более ранних стадиях, включая этап внедрения команд операционной системы.

В сценарии несанкционированного доступа через подбор или скрытое изменение пароля, NIDS Suricata также не фиксирует атаки в зашифрованном трафике. ML COB превосходит WAF ModSecurity по эффективности в обнаружении атак подбора пароля и CSRF-атак.

Когда речь идет о моделировании эксплуатации 0-day уязвимостей, NIDS Suricata снова не справляется с обнаружением атак в зашифрованном трафике, тогда как ML COB успешно выявляет ранее неизвестные угрозы.

Комплексное сравнение средств защиты показало, что ML COB в тестовых условиях превосходит NIDS Suricata по всем показателям и работает на уровне WAF ModSecurity. Также стоит отметить, что эффективность WAF ModSecurity зависит от полноты базы правил, тогда как ML COB зависит от качества обучающего трафика.

Одним из главных ограничений IDS на основе машинного обучения является время и сложность обучения, в том числе на реальном трафике, а также играют роль задержки трафика при обучении.

Преимуществами IDS, основанными на моделях машинного обучения, по сравнению с сигнатурными IDS, являются: возможность выявления 0-day уязвимостей; работа с большим объемом сетевого трафика; детекция скрытых паттернов в сетевом поведении; сокращение ложноотрицательных и ложноположительных срабатываний по сравнению с сигнатурными IDS.

Заключение

В статье рассмотрены различные модели машинного обучения и их применение в сигнатурных системах обнаружения вторжений. Опираясь на описанные преимущества и результаты исследований по данной тематике, можно сделать вывод, что средства с искусственным интеллектом определенно могут внести новшества в функционал существующих средств IDS, однако, не могут выступать в качестве самостоятельных программно-аппаратных средств защиты. Использование модулей машинного обучения осложнено необходимостью трудоемкого обучения и «каллибровкой» модели, однако, целесообразным является их использование в качестве дополнения к существующим сигнатурным средствам защиты и дальнейшее исследование потенциала в данной области, т.к. в комплексе средство повысить общую эффективность выявления вторжений и приближает разработчиков средств кибербезопасности к разрешению проблемы невозможности обнаружения ранее неизвестных атак.

ЛИТЕРАТУРА

1. Авезова Я.Э. Актуальные киберугрозы в странах СНГ 2023—2024 [Электронный ресурс] // Positive Technologies : [сайт]. [2024]. URL: https://www.ptsecurity.com/ru-ru/research/analytics/aktualnye-kiberugrozy-v-stranah-sng-2023-2024/ (дата обращения: 19.10.2024)
2. Usmani M., Anwar M., Komal F., Ghufran A., Shahbaz A. Predicting ARP spoofing with Machine Learning. 2022 International Conference on Emerging Trends in Smart Technologies (ICETST), Karachi, Pakistan, 2022, pp. 1–6. URL: https://ieeexplore.ieee.org/document/9922925 (дата обращения: 19.10.2024)
3. Braga R., Mota E., Passito, A. Lightweight DDoS flooding attack detection using NOX/OpenFlow // The 35th Annual IEEE Conference on Local Computer Networks, Denver, Colorado, USA, Proceedings, 10–14 October 2010, с. 408–415. URL: https://ieeexplore.ieee.org/document/5735752 (дата обращения: 19.10.2024)
4. Ibrahim L.M., Basheer D.T., Mahmod M.S. A comparison study for intrusion database (KDD99, NSL-KDD) based on self-organization map (SOM) artificial neural network // Journal of Engineering Science and Technology, 2018, с.107–119. URL: https://www.researchgate.net/publication/ 329450947_A_comparison _study_for_intrusion_database_KDD99_NSL-KDD_based_on_self_organization_map_SOM_artificial_neural_network (дата обращения: 19.10.2024)
5. Kokila R., Selvi S.T.; Govindarajan K. DDoS detection and analysis in SDN-based environment using support vector machine classifier // International Conference on Advanced Computing, Chennai, India, 2014, с. 205–210. URL: https://ieeexplore.ieee.org/document/7229711 (дата обращения: 19.20.2024)
6. Mukherjee S., Sharma N. Intrusion Detection using Naive Bayes Classifier with Feature Reduction. // Procedia Technology, 2012, Том 4, с. 119–128. URL: https://www.sciencedirect.com/science/article/pii/S2212017312002964 (дата обращения: 19.20.2024)
7. Shone N., Ngoc T.N., Phai V.D., Shi,Q. A Deep Learning Approach to Network Intrusion Detection // IEEE Transactions on Emerging Topics in Computational Intelligence, 2018, Том 2, № 1, с. 41–50. URL: https://eudl.eu/pdf/10.4108/eai.3-12-2015.2262516 (дата обращения 19.10.2024)
8. Yin C., Zhu Y., Fei J., He X. A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks // IEEE Access, 2017, Том 5, с. 21954–21961. URL: https://ieeexplore.ieee.org/document/8066291 (дата обращения: 19.10.2024)
9. Tang T.A., Mhamdi L., McLernon D., Zaidi S.A.R., Ghogho M. DeepIDS: Deep Learning Approach for Intrusion Detection in Software Defined Networking. // Electronics, 2020, № 9, с. 1533. URL: https://www.researchgate.net/publication/344450711_DeepIDS_Deep_Learning_Approach_for_Intrusion_Detection_in_Software_Defined_Networking (дата обращения: 19.10.2024)
10. Shah S.A.R., Issac B. Performance Comparison of Intrusion Detection Systems and Application of Machine Learning to Snort System. // Future Generation Computer Systems, 2018, Том 80, с. 157–170. URL: https://nrl.northumbria.ac.uk/id/eprint/35851/1/Shah,%20Issac%20-%20Performance%20comparison%20of%20intrusion%20detection%20systems%20and%20application%20of%20machine%20learning%20to%20Snort%20system%20AAM.pdf (дата обращения: 19.10.2024)
11. Гетьман А.И., Горюнов М.Н., Мацкевич А.Г., Рыболовлев Д.А. Сравнение системы обнаружения вторжений на основе машинного обучения с сигнатурными средствами защиты информации. Труды Института системного программирования РАН. 2022;34(5):111–126. https://doi.org/10.15514/ISPRAS-2022-34(5)-7 (дата обращения: 19.10.2024)

© Карелова Оксана Леонидовна (okarelova@yandex.ru); Кострова Ольга Евгеньевна (olkeruac@gmail.com);
Курбанова Камиля Маратовна (kamila.business03@gmail.com)
Журнал «Современная наука: актуальные проблемы теории и практики»