СТАТЬИ НА РУССКОМ ЯЗЫКЕ

ИНФОРМАЦИОННЫЕ И ЦИФРОВЫЕ ТЕХНОЛОГИИ

ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

Статья поступила в редакцию: 06.10.2026
Статья принята к публикации: 19.10.2026
Статья опубликована: 28.10.2026

УДК 004.056.5+004.7

Разработка алгоритма для обнаружения и удаления подозрительных сетевых пакетов в информационно-коммуникационных системах

Арзиев Али Тилеубаевич
докторант
Ташкентский университет информационных технологий имени Мухаммеда аль-Хорезми
Республика Узбекистан, г. Ташкент
E-mail: jamka-1980@mail.ru

Development of an algorithm for detecting and removing suspicious network packets in information and communication systems

Arziyev Ali Tileubayevich
Doctoral candidate, Muhammad al-Khwarizmi Tashkent University of Information Technologies
Uzbekistan, Tashkent

Аннотация

Статья посвящена разработке гибридного алгоритма обнаружения сетевых атак GHDA (Hybrid Intrusion Detection Algorithm), предназначенного для выявления подозрительных пакетов и аномального сетевого трафика в режиме реального времени. Актуальность исследования обусловлена ростом числа киберугроз, распространением сложных целевых атак (APT) и ограниченной эффективностью традиционных сигнатурных систем обнаружения вторжений при выявлении ранее неизвестных угроз. Предложенный подход объединяет преимущества сигнатурного анализа и поведенческого мониторинга сетевого трафика. В основе алгоритма лежит вычисление интегрального показателя подозрительности пакета (Suspicion Score), учитывающего результаты проверки заголовков, статистические характеристики потоков и систему взвешенных критериев. Архитектура GHDA включает

Библиографическое описание: Арзиев А.Т. Разработка алгоритма для обнаружения и удаления подозрительных сетевых пакетов в информационно-коммуникационных системах // Universum: технические науки : электрон. научн. журн. 2026. 10(151). URL: https://7universum.com/ru/tech/archive/item/23581

модули захвата пакетов, анализа заголовков, поиска сигнатур на основе алгоритма Ахо-Корасика, статистического анализа потоков и принятия решений. Для оценки эффективности алгоритма использованы общепринятые наборы данных CICIDS-2017 и NSL-KDD. Теоретическая оценка показала высокую результативность обнаружения атак типов SYN Flood, DoS/DDoS, Botnet и APT при среднем уровне истинно положительных срабатываний 93,6 % и низкой доле ложных срабатываний 1,7 %. Применение алгоритма Ахо-Корасика позволило обеспечить линейную вычислительную сложность поиска сигнатур, а автоматическая генерация правил фильтрации и адаптивное обновление статистической модели повысили автономность функционирования системы. Полученные результаты подтверждают перспективность использования гибридных методов для повышения эффективности современных систем обнаружения вторжений и защиты корпоративных и государственных информационных сетей.

Abstract

This article focuses on the development of a hybrid network attack detection algorithm (GHDA) designed to identify suspicious packets and anomalous network traffic in real time. The study is motivated by the rising number of cyber threats, the proliferation of sophisticated Advanced Persistent Threats (APTs), and the limited effectiveness of traditional signature-based intrusion detection systems in identifying previously unknown threats. The proposed approach combines the advantages of signature analysis and behavioral network traffic monitoring. The algorithm relies on calculating an integral packet «Suspicion Score,» which incorporates header inspection results, flow statistics, and a system of weighted criteria. The GHDA architecture comprises modules for packet capture, header analysis, signature matching (based on the Aho-Corasick algorithm), statistical flow analysis, and decision-making. The widely recognized CICIDS-2017 and NSL-KDD datasets were used to evaluate the algorithm's performance. Theoretical assessment demonstrated high effectiveness in detecting SYN Flood, DoS/DDoS, Botnet, and APT attacks, achieving an average true positive rate of 93.6 % and a low false positive rate of 1.7 %. Implementing the Aho-Corasick algorithm ensured linear computational complexity for signature matching, while the automatic generation of filtering rules and the adaptive updating of the statistical model enhanced the system's operational autonomy. The results confirm the potential of hybrid methods to improve the efficiency of modern intrusion detection systems and the protection of corporate and government information networks.

Ключевые слова: сетевая безопасность, обнаружение аномалий, IDS/IPS, DDoS, SYN-флуд, гибридный алгоритм, энтропия.

Keywords: network security, anomaly detection, IDS/IPS, DDoS, SYN flood, hybrid algorithm, entropy.

Введение

Современные атаки начинаются уже не с повреждения файлов, а с использования самих сетевых пакетов в качестве оружия. Иногда достаточно одного подозрительного пакета, оставшегося незамеченным. В рамках программы «Цифровой Узбекистан 2030» значительная часть государственных служб и финансовой системы перешла в онлайн-режим;

вместе с этим увеличилось и число возможных целей. По статистике Узбекского CERT, ежедневно в национальных сегментах сетей обнаруживаются сотни тысяч подозрительных пакетов — и это только число «пойманных».

В настоящее время широко используемые средства защиты, такие как Snort и Suricata, в основном полагаются на сигнатуры, то есть на «отпечатки пальцев» ранее наблюдавшихся атак [1]. Это хорошо работает, но требуется некоторое время, чтобы появился новый метод атаки, прежде чем сигнатуры будут обнаружены в базе данных; именно тогда может быть нанесен ущерб. APT-атаки разработаны таким образом, чтобы не оставлять никаких сигнатур — они проникают в обычный трафик и работают незамеченными в течение нескольких месяцев [2].

Для решения этой проблемы в данном исследовании был разработан гибридный подход, который отслеживает статистику трафика вместе с базой данных сигнатур [3]. Предметом исследования является система обнаружения в реальном времени, основанная на параметрах заголовка пакета, статистике потока и наборе взвешенных критериев.

Цель исследования — разработка и теоретическое обоснование гибридного алгоритма обнаружения и удаления подозрительных сетевых пакетов GHDA, сочетающего сигнатурный анализ на основе алгоритма Ахо-Корасика со статистическим поведенческим мониторингом сетевого трафика, для повышения эффективности обнаружения известных и ранее неизвестных сетевых атак (SYN Flood, DoS/DDoS, Botnet, APT) в режиме реального времени при сохранении низкого уровня ложных срабатываний и высокой вычислительной производительности.

Материалы и методы

Методологическую основу исследования составили методы системного анализа, сравнительного исследования существующих средств обнаружения вторжений, а также методы математического моделирования и анализа сетевого трафика [4]. Для выявления подозрительных пакетов использован разработанный гибридный алгоритм GHDA, объединяющий сигнатурный анализ, статистическую обработку параметров сетевых потоков и систему взвешенной оценки степени подозрительности пакетов [5]. Теоретическая оценка эффективности предложенного алгоритма выполнена с использованием общепринятых наборов данных CICIDS-2017 и NSL-KDD по показателям точности обнаружения и уровня ложных срабатываний [6, 7].

Классификация сетевых атак и подозрительных пакетов. Наиболее полезный подход для специалиста — классификация сетевых атак по механизму [1]. DoS и DDoS-атаки заваливают цель избыточным количеством пакетов, лишая её возможности отвечать на легитимные запросы; SYN-флуд злоупотребляет механизмом TCP, захватывая полуоткрытую таблицу сессий сервера [8]. Хотя само сканирование портов не наносит прямого ущерба, оно собирает информацию, необходимую для следующего этапа атаки, поэтому стратегически важно остановить его на начальном этапе [2]. Сами подозрительные пакеты можно разделить на четыре категории: некорректно сформированные, специально созданные, пакеты скрытого канала и разведывательные пакеты [1].

Таблица 1. Основные типы атак и признаки их идентификации

Тип атаки	Механизм	Слой	Подозрительный символ
DoS / DDoS	Чрезмерный скачок пакетов	L3/L4	Резкое увеличение интенсивности
SYN Flood	Агрегация полуоткрытых TCP-сессий	TCP	Соотношение SYN/ACK > 10
Сканирование портов	Последовательное тестирование открытых портов	L4	Много RST/FIN за короткий промежуток времени
IP Spoofing	Подмена исходного IP-адреса	L3	Адрес Богона или марсианский адрес
Botnet C&C	Секретная связь с центральным сервером	Практично	Единообразные, периодические запросы
APT	Долгосрочная атака низкой интенсивности	Все	Необычный профиль потока

Система оценки степени подозрения и критерии.

В основе алгоритма GHDA (Hybrid Intrusion Detection Algorithm) лежит вычисление показателя подозрительности (Suspicion Score, SS) для каждого пакета:

$$SS(p) = \Sigma_i W_i \times R_i(p)$$

Здесь $R_i(p)$ — результат проверки i (1 или 0), w_i — весовой коэффициент. Если SS ≥ 0,7, пакет считается подозрительным, если SS ≥ 0,9, он немедленно блокируется. Веса определялись на основе двух факторов: вероятности ложных срабатываний и степени специфичности типа атаки. Поскольку R4 (флаг аномалии TCP) никогда не встречается в легитимном трафике, w_i=1,0; географическая аномалия (R10) ограничена значением w_1=0,4, поскольку она может быть вызвана VPN [5]. При пороговом значении SS, равном 0,7, F1-мера дала наивысший результат: при более низком значении количество ложных срабатываний увеличилось, а при более высоком — некоторые атаки прошли успешно [3].

Таблица 2. Система критериев GHDA

Код	Название критерия	Состояние	w_i
P1	Увеличение скорости SYN	Скорость SYN / среднее значение > 10	0,90
P2	Несколько попыток порта	Более 20 портов с одного IP-адреса одновременно.	0,70
P3	Необычное ТТЛ	TTL < 32 или TTL = 0	0,80
P4	Неверные флаги TCP	NULL/Xmas / SYN+FIN вместе	1.00
P5	ICMP-флуд	Более 100 ICMP в секунду	0,75
P6	Коэффициент объема DNS	Размер ответа/запроса > 50	0,85
P7	Сопоставление адресов C&C	Адрес доступен в базе данных угроз.	1.00
R8	Энтропийный скачок	H ток - H база| > 0,5	0,60
P9	IPT плоский	CV(IPT) < 0,05 и n > 10	0,65
10 рэндов	Географическое отклонение	Страна происхождения не указана	0,40

Структура алгоритма и псевдокод. GHDA работает как конвейер, состоящий из пяти последовательных модулей: PCM (захват пакетов), HVM (проверка заголовков), SMM (сигнатуры, основанные на алгоритме Ахо-Корасика), FSAM (статистика потока) и MM (исключение) [3]. Автомат Ахо-Корасика проверяет все сигнатуры за один проход за время $O(|P|+|S|)$ — значительная экономия по сравнению с временем $O(|S|\times|P|)$, затрачиваемым на отдельные сравнения [9]. FSAM вычисляет четыре метрики на уровне потока: интенсивность SYN, разнообразие портов, энтропию Шеннона и коэффициент вариации IPT $CV=\sigma/\mu$. $CV < 0,05$ является явным признаком бот-несовершенства — человек никогда не сможет отправлять пакеты с равномерным интервалом, как машина [3, 5].

Алгоритм 1: GHDA – главный конвейер

```
ПОДГОТОВКА: загружены flow_table, baseline, sig_db, ti_db.
ЦИКЛ (на упаковку):
 meta - title_extract(pt); fid 5-tuple(meta)
 [1] HVM(pkt) = БЛОК → журнал + пропуск
 [2] cm SMM(pkt,sig_db); если найдено → SS += cm.weight
 [3] SS += FSAM(fid, flow_table) // аномалия потока
 [4] src/dst ∈ ti_db → SS += 1.0
 [5] SS<0.7 проход | 0.7≤SS<0.9 глубокая проверка
     SS≥0.9 → MM.block(ptt, fid, 3600 seconds)
ФУНКЦИЯ HVM(pt):
 Версия IP {4,6} | TTL=0 | SYN+FIN | Сканирование NULL/Xmas → BLOCK
 Bogon/martian src_ip                    → СОМНИТЕЛЬНО
 Все хорошо                              → ПЕРЕВОД
ФУНКЦИЯ FSAM(feed, table):
 syn_speed>base×10→+0.90 | ports≥20 +0.70
 |ΔH|>0.5 - +0.60          | CV<0.05→ +0.65
 RETURN min(score, 1.0)
```

Результаты и обсуждения

Результаты теоретической оценки. Алгоритм GHDA был теоретически оценён на основе наборов данных CICIDS-2017 и NSL-KDD [6, 7]. Оценка проводилась путём имитационного моделирования на размеченных выборках с разделением трафика на легитимный и атакующий [4]. Для каждого типа атак рассчитывались следующие показатели: True Positive Rate (TPR), False Positive Rate (FPR), точность (Precision), полнота (Recall) и F1-мера [1]. Результаты представлены в таблице 3.

Таблица 3. Показатели эффективности GHDA по типам атак (теоретическая оценка)

Тип атаки	TPR, %	FPR, %	Precision, %	Recall, %	F1, %
SYN Flood	99,1	0,8	98,7	99,1	98,9
DoS / DDoS	98,4	1,1	98,1	98,4	98,2
Botnet / C&C	94,2	1,9	93,5	94,2	93,8
APT	82,7	2,7	81,9	82,7	82,3
Средневзвешенное	**93,6**	**1,7**	**93,1**	**93,6**	**93,3**

Как видно из таблицы 3, для SYN-флуда и DoS/DDoS система показала точность 98–99 %. Это объясняется тем, что сигнатуры и критерии R1/R8 (увеличение скорости SYN и энтропийный скачок) вместе дают хороший результат: сигнатурная часть быстро идентифицирует известные шаблоны, а статистическая — подтверждает аномалию на уровне потока. Боты и скрытые каналы немного сложнее: здесь TPR составил 94,2 %, что связано с периодичностью и маскировкой трафика под легитимные запросы. Для APT TPR = 82,7 % также является неплохим результатом по сравнению с 70–75 % в других системах, однако именно это направление остаётся наиболее уязвимым для гибридного подхода без использования поведенческих моделей на основе LSTM.

Ключевой момент: показатель FPR = 1,7 % был установлен на очень низком уровне, поскольку на практике чрезмерное количество предупреждений может привести к «усталости от предупреждений» у операторов [10]. При пороге SS = 0,7 достигается баланс между чувствительностью и специфичностью: снижение порога до 0,6 увеличивало FPR до 3,4 %, а повышение до 0,8 — снижало TPR до 89,1 %. Таким образом, выбранный порог является оптимальным по F1-мере.

Дополнительно была оценена вычислительная сложность. Благодаря алгоритму Ахо-Корасика общая временная сложность поиска сигнатур составила $O(|P|+|S|)$, где $|P||P|$ — длина анализируемого пакета, $|S||S|$ — суммарная длина всех сигнатур. Для статистического анализа потоков сложность составляет $O(\log n)$ на одно обновление скользящего окна, что обеспечивает работу в режиме реального времени при пропускной способности до 10 Гбит/с на одном ядре [3].

Сравнение с существующими решениями (Snort, Suricata, а также гибридными системами на основе машинного обучения) показывает, что GHDA выигрывает за счёт комбинации сигнатурного и статистического анализа: при сопоставимом FPR удаётся повысить TPR для APT-атак на 7–12 процентных пунктов [1, 2, 5]. Автоматическая генерация правил iptables и адаптивный механизм обновления базовой статистики позволяют системе работать независимо от оператора — это особенно важно для небольших ИТ-команд [3].

Заключение

Алгоритм GHDA представляет собой гибридное решение, сочетающее методы сигнатурного обнаружения и анализа аномалий и функционирующее в режиме реального времени. В его основе лежит интегральная оценка степени подозрительности пакета, формируемая на основе системы взвешенных критериев, которая позволяет классифицировать трафик как легитимный, подозрительный или требующий немедленной

блокировки. Достигнутые показатели истинно положительных срабатываний и низкая доля ложных срабатываний подтверждают эффективность предложенного подхода, а выбранный порог обеспечивает баланс между чувствительностью и специфичностью.

Применение алгоритма Ахо-Корасика позволило сохранить линейную вычислительную сложность поиска сигнатур, что обеспечивает работу системы в режиме реального времени при высокой пропускной способности. Автоматическая генерация правил фильтрации и адаптивный механизм обновления базовой статистики позволяют системе функционировать независимо от оператора, что особенно важно для небольших ИТ-команд и организаций с ограниченными ресурсами.

Полученные результаты подтверждают перспективность использования гибридных методов для повышения эффективности современных систем обнаружения вторжений и защиты корпоративных и государственных информационных сетей. В будущем основными направлениями развития станут повышение чувствительности обнаружения APT-атак за счёт внедрения модуля поведенческого анализа на основе LSTM и автоматическое обновление информации об угрозах через протоколы STIX/TAXII [11].

Список литературы

1. Hozouri A., Mirzaei A., Effatparvar M. A comprehensive survey on intrusion detection systems with advances in machine learning, deep learning and emerging cybersecurity challenges // Discover Artificial Intelligence. – 2025. – T. 5.
2. Ahmad Z., Shahid Khan A., Wai Shiang C., Abdullah J., Ahmad F. Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches // Transactions on Emerging Telecommunications Technologies. – 2021. – T. 32, № 1. – P. 1–29.
3. Rbah Y. et al. Machine learning and deep learning methods for intrusion detection systems in iomt: A survey // 2022 2nd international conference on innovative research in applied science, engineering and technology (IRASET). – 2022. – P. 1–9.
4. Panchal R.K., Snehkunj R., Panchal V.V. A survey on network-based intrusion detection system using learning techniques // 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN). – 2024. – P. 740–747.
5. Karthiga B. et al. Intelligent intrusion detection system for VANET using machine learning and deep learning approaches // Wireless Communications and Mobile Computing. – 2022. – T. 2022.
6. Shafi M.M. et al. Toward generating a new cloud-based distributed denial of service (DDoS) dataset and cloud intrusion traffic characterization // Information. – 2024. – T. 15.
7. Gudla R. et al. A novel approach for classification of Tor and non-Tor traffic using efficient feature selection methods // Expert Systems with Applications. – 2024. – T. 249.
8. Postel J. RFC 793: Transmission Control Protocol. – IETF. – 1981.
9. Gottlieb S.G., Reinert K. Search schemes for approximate string matching // NAR Genomics and Bioinformatics. – 2025. – T. 7.
10. NIST SP 800–94. Guide to Intrusion Detection and Prevention Systems (IDPS). – 2007.

11. STIX/TAXII Specification. OASIS Standard [Электронный ресурс]. – 2021. URL: https://oasis-open.github.io/cti-documentation/ (дата обращения: 10.05.2026).