Коломейцев Александр Эдуардович
кандидат технических наук, сотрудник
Краснодарское высшее военное орденов Жукова и Октябрьской Революции
Краснознамённого училище имени генерала армии С.М. Штеменко

Юрчак Владислав Витальевич, курсант
Краснодарское высшее военное орденов Жукова и Октябрьской Революции
Краснознамённое училище имени генерала армии С.М. Штеменко

Богомолов Степан Александрович, сотрудник
Краснодарское высшее военное орденов Жукова и Октябрьской Революции
Краснознамённого училище имени генерала армии С.М. Штеменко

Чукин Александр Юрьевич, сотрудник
Краснодарское высшее военное орденов Жукова и Октябрьской Революции
Краснознамённого училище имени генерала армии С.М. Штеменко

Миронов Максим Евгеньевич, курсант
Краснодарское высшее военное орденов Жукова и Октябрьской Революции
Краснознамённое училище имени генерала армии С.М. Штеменко

ЗАЩИТА ОТ АТАК С ИСПОЛЬЗОВАНИЕМ ВРЕДОНОСНОГО ПО НУЛЕВОГО ДНЯ (ZERO-DAY): МЕТОДЫ ПОВЕДЕНЧЕСКОГО АНАЛИЗА И ПЕСОЧНИЦЫ

Аннотация. Статья, опирающаяся на анализ актуальных угроз 2025–2026 годов и передовых методов защиты, посвящена проблеме обнаружения и предотвращения атак с использованием неизвестных ранее уязвимостей (zero-day). Рассматриваются фундаментальные ограничения сигнатурного подхода, вскрывшиеся на фоне лавинообразного роста числа новых образцов вредоносного ПО.

Ключевые слова: Zero-day уязвимость, поведенческий анализ, песочница (sandbox), машинное обучение, искусственный интеллект, обнаружение аномалий.

Введение. Современный ландшафт киберугроз, претерпевая стремительную трансформацию, характеризуется не только количественным ростом инцидентов, но и качественным усложнением методов атак, причём наиболее опасным классом остаются эксплуатации уязвимостей нулевого дня (zero-day), для которых на момент атаки не существует известных сигнатур или выпущенных обновлений. Злоумышленники, всё чаще отказываясь от громких деструктивных действий в пользу долгосрочного необнаружимого присутствия, вынуждают организации пересматривать подходы к защите, поскольку традиционные антивирусные решения, ориентированные на сигнатурный анализ, демонстрируют свою несостоятельность перед лицом полиморфного, обфусцированного и постоянно видоизменяющегося вредоносного кода.

Согласно данным отчёта Google Threat Intelligence Group за 2025 год, в реальных условиях было зафиксировано 90 эксплуатируемых zero-day уязвимостей, что превышает показатель 2024 года (78) и укладывается в диапазон 60–100, установившийся за последние четыре года, однако при этом доля уязвимостей, направленных на корпоративные технологии, достигла рекордных 48%, что свидетельствует о смещении фокуса атак с конечных пользователей на инфраструктурные компоненты, включая сетевые экраны, устройства удалённого доступа и средства безопасности. Ещё более тревожную динамику демонстрирует рынок вредоносного ПО: в 2025 году количество новых уникальных образцов возрастало

каждый квартал, достигнув ошеломляющего скачка на 1548% только в четвёртом квартале, причём 23% обнаруженного вредоносного ПО успешно обходили традиционные сигнатурные детекторы, автоматически классифицируясь как zero-day угрозы. Показательно, что российский бизнес, согласно опросу, проведённому в феврале 2025 года, остаётся в значительной степени недооценивающим эту угрозу: почти две трети компаний утверждают, что не сталкиваются с zero-day атаками, тогда как проактивную защиту используют не более 15% респондентов, причём передовые технологии анализа сетевого трафика и Threat Intelligence внедрены менее чем в 10% организаций.

Цель настоящей статьи, таким образом, заключается в систематизации и сравнительном анализе современных методов защиты от zero-day угроз, акцентируя внимание на двух взаимодополняющих подходах – поведенческом анализе и технологии песочниц (sandbox), – поскольку именно эти методы, будучи свободными от зависимости от сигнатур, способны обнаруживать неизвестное вредоносное ПО на основе его активности. Для достижения этой цели предполагается решить следующие задачи: во-первых, охарактеризовать текущее состояние угроз zero-day, опираясь на статистические данные 2025–2026 годов; во-вторых, проанализировать эволюцию методов поведенческого анализа, включая их реализацию на основе машинного обучения и искусственного интеллекта; в-третьих, исследовать архитектуру и функциональные возможности современных песочниц, уделяя особое внимание методам обхода, которые применяют злоумышленники для маскировки; в-четвёртых, предложить практические рекомендации по построению гибридных систем защиты, объединяющих преимущества поведенческого анализа, песочниц и технологий машинного обучения.

Природа zero-day угроз: масштаб, динамика и структурные сдвиги

Прежде чем переходить к рассмотрению методов защиты, необходимо зафиксировать масштаб проблемы, поскольку понимание количественных и качественных характеристик zero-day атак позволяет обоснованно выбирать приоритетные направления для внедрения защитных механизмов. Анализируя данные Google Threat Intelligence Group, можно выделить несколько устойчивых трендов, сформировавшихся к 2025 году. Во-первых, количество ежегодно эксплуатируемых zero-day уязвимостей стабилизировалось в диапазоне 60–100, что указывает на некий «равновесный уровень», определяемый, с одной стороны, усилиями разработчиков по повышению безопасности, а с другой – наращиванием ресурсов злоумышленников (включая коммерческих поставщиков средств наблюдения) по поиску и приобретению новых уязвимостей. Во-вторых, произошёл структурный сдвиг в сторону корпоративных технологий: 48% всех zero-day эксплойтов в 2025 году были нацелены на enterprise-решения – сетевые устройства, платформы виртуализации, системы управления доступом и приложения для совместной работы. Этот сдвиг, будучи обусловленным экономической целесообразностью (компрометация одного корпоративного устройства открывает доступ к целой сети), требует пересмотра стратегий защиты, смещая фокус с периметра на внутреннюю инфраструктуру.

В-третьих, наблюдается устойчивый рост эксплуатации уязвимостей в сетевых и security-устройствах (брандмауэры, VPN-шлюзы, системы обнаружения вторжений), которые, будучи расположенными на границе сети, предоставляют злоумышленникам идеальную точку входа, причём атаки на этот класс устройств, как правило, остаются незамеченными, поскольку сами средства защиты не могут контролировать собственную целостность. Примером может служить атака группировки Interlock, которая эксплуатировала zero-day уязвимость в Cisco Secure Firewall Management Center (CVE-2026-20131) за 36 дней до официального выпуска патча, что позволило злоумышленникам получать удалённый доступ с правами root к устройствам, оставаясь необнаруженными на протяжении более месяца.

Таблица 1

Динамика эксплуатации zero-day уязвимостей (2021–2025 гг.)

Год	Количество эксплуатируемых zero-day	Доля корпоративных целей (%)	Основные векторы атак
2021	81	~30	Браузеры, мобильные ОС
2022	65	~35	Почтовые серверы, VPN
2023	100	~38	Веб-приложения, гипервизоры
2024	78	~42	Сетевые устройства, облачные API
2025	90	48	Edge-устройства, security-аппараты

В-четвёртых, коммерческие поставщики средств наблюдения (commercial surveillance vendors, CSV) впервые за всю историю наблюдений превзошли традиционные государственные группы по числу используемых zero-day эксплойтов, что свидетельствует о коммерциализации рынка уязвимостей и снижении порога входа для злоумышленников, готовых приобретать эксплойты на чёрном или полулегальном рынках. Иллюстрацией этого тренда является эксплойт-чейн DarkSword, использующий шесть различных уязвимостей (четыре из которых – zero-day) для полной компрометации устройств под управлением iOS 18.4–18.7, причём данный инструментарий, начиная с ноября 2025 года, применялся как минимум тремя различными коммерческими поставщиками и рядом групп, связанных с государствами, в кампаниях, нацеленных на пользователей в Саудовской Аравии, Турции, Малайзии и Украине.

Таблица 2

Распределение zero-day эксплойтов по категориям целей (2025 г.)

Категория цели	Количество zero-day	Доля от общего числа	Основные группы угроз
Корпоративное ПО	43	48%	CSV, государственные группы
Мобильные ОС	15	17%	CSV, киберпреступники
Браузеры	<10	<10%	APT-группы
Операционные системы (ПК)	22	24%	APT-группы, вымогатели

Поведенческий анализ: от сигнатур к аномалиям

Поведенческий анализ, составляющий основу современных систем защиты от zero-day угроз, базируется на фундаментальном предположении: вредоносное ПО, каким бы способом оно ни было замаскировано и обфусцировано, на этапе исполнения неизбежно демонстрирует определённые паттерны активности, отличающие его от легитимных приложений. Эти паттерны, включая обращение к системным вызовам (API), создание процессов, изменение файловой системы, сетевые соединения и модификацию реестра, фиксируются и анализируются в реальном времени, причём ключевым преимуществом поведенческого подхода является его независимость от сигнатур, что позволяет обнаруживать неизвестные образцы, руководствуясь исключительно аномальностью их поведения.

Традиционные методы поведенческого анализа, основанные на правилах (rule-based), предполагают задание экспертных пороговых значений – например, обнаружение процесса, пытающегося зашифровать более 100 файлов за минуту, трактуется как признак ransomware-атаки. Однако такой подход, будучи интуитивно понятным и легко интерпретируемым, страдает двумя недостатками: во-первых, злоумышленники, изучая правила, могут

адаптировать своё ПО так, чтобы не превышать заданные пороги (например, замедляя шифрование); во-вторых, легитимные приложения, выполняющие интенсивные файловые операции (резервное копирование, индексация), могут порождать ложные срабатывания, перегружая аналитиков.

Современным ответом на эти ограничения стало внедрение методов машинного обучения, позволяющих автоматически выявлять сложные, многомерные паттерны вредоносного поведения, недоступные для правилах экспертного анализа. Обзор, проведённый Абдуллой Аль Сиамом и соавторами (2026), охватывает четыре фундаментальных семейства моделей для обнаружения zero-day вторжений: свёрточные нейронные сети (CNN), глубокие нейронные сети (DNN), байесовские сети (BN) и обучение с подкреплением (RL). Авторы показывают, что DNN демонстрируют наилучшую агрегированную производительность на богатых признаковых входах, достигая 99,56% точности на наборе данных CICDDoS2019; CNN показывают преимущество при анализе тензоризированных потоков и байтов (92,17% на Bot-IOT); байесовские сети обеспечивают интерпретируемую неопределённость при приемлемой точности (99,74% на NSL-KDD); а обучение с подкреплением перспективно для адаптивного обнаружения-реагирования, достигая 96,18% на CSE-CIC-IDS2018.

Особого внимания заслуживает подход, реализованный в системе BEACON, которая использует большие языковые модели (LLM) для генерации плотных контекстуальных эмбеддингов из отчётов о поведении, сгенерированных песочницами. Эти эмбеддинги, фиксирующие как семантические, так и структурные паттерны образцов, затем обрабатываются одномерной свёрточной нейронной сетью для многоклассовой классификации вредоносного ПО, причём авторы сообщают, что BEACON последовательно превосходит существующие методы на общедоступном наборе данных Avast-CTU CAPE. Ещё одним перспективным направлением является использование графовых нейронных сетей (GNN), как это сделано в системе GraphShield, преобразующей последовательности API-вызовов во временные графы и применяющей механизмы внимания для извлечения как локальных активностей, так и протяжённых поведенческих корреляций. GraphShield, обученный на 300 000 сбалансированных примерах и протестированный на отдельной выборке из 200 000 образцов, достигает F1-меры 99,5% при уровне ложных срабатываний менее 1%, что представляет собой улучшение более чем на 58% по сравнению с продвинутыми моделями на основе последовательных данных.

Таблица 3

Сравнение эффективности моделей машинного обучения для обнаружения zero-day

Модель	Набор данных	Точность (Acc) / F1	Ложные срабатывания	Лаг при анализе
DNN	CICDDoS2019	99,56%	0,5%	Низкий
BN	NSL-KDD	99,74%	0,3%	Низкий (интерпретируемо)
CNN	Bot-IOT	92,17%	1,2%	Очень низкий (для edge)
RL	CSE-CIC-IDS2018	96,18%	0,8%	Средний (обучение)
GNN (GraphShield)	Собственный (500K)	99,5% (F1)	<1%	Средний
LLM + CNN (BEACON)	Avast-CTU CAPE	>существующих методов	–	Высокий

Технология песочниц (Sandbox): архитектура, эволюция и контрмеры

Песочница, будучи изолированной виртуальной средой для безопасного исполнения подозрительного кода, представляет собой ключевой инструмент динамического поведенческого анализа, позволяющий наблюдать за действиями вредоносного ПО без риска для целевой системы. Классическая архитектура песочницы включает гипервизор или эмулятор, создающий виртуальную машину с заданной конфигурацией (операционная система, установленное ПО, сетевое окружение), детектор, запускающий исследуемый файл и фиксирующий все его действия (системные вызовы, сетевые подключения, доступ к реестру, создание процессов), и анализатор, сопоставляющий зафиксированную активность с правилами обнаружения.

Однако эволюция песочниц, происходящая под давлением постоянно совершенствующихся методов обхода, привела к появлению целого семейства контрамер со стороны злоумышленников, и согласно отчёту Picus Red Report 2026, проанализировавшему более 1,1 миллиона вредоносных файлов и 15,5 миллионов действий, техника «виртуализации и обхода песочниц» (T1497) поднялась на четвёртое место в рейтинге наиболее распространённых методов MITRE ATT&CK, встречаясь в 20% проанализированных образцов, причём этот метод, отсутствовавший в топ-10 на протяжении двух предыдущих лет, демонстрирует взрывной рост популярности.

Методы обхода песочниц можно классифицировать на три основные категории, каждая из которых представляет собой отдельный вызов для систем защиты. *Системные проверки* (T1497.001) предполагают сбор информации об окружении: проверка количества процессоров (если меньше четырёх – аборт), размера экрана (если соответствует стандартным разрешениям песочниц 1024x768 или 800x600 – аборт), наличия драйверов, характерных для инструментов анализа, имён виртуальных дисков (VBOX, VMWare) или MAC-адресов гипервизоров. Вредоносное ПО Blitz, проанализированное в июне 2025 года, демонстрирует все эти проверки, выполняясь только при обнаружении реальной рабочей станции с достаточными ресурсами.

Проверки времени (T1497.003) основаны на том факте, что песочницы, как правило, имеют ограниченное время анализа (5–10 минут), поэтому вредоносное ПО может откладывать свою деструктивную активность на 15–20 минут или использовать механизмы ожидания ввода пользователя, предполагая, что в автоматизированной среде такой ввод отсутствует. *Проверки пользовательской активности* (T1497.002) представляют собой наиболее изощрённый класс методов обхода: вредоносное ПО анализирует движение мыши, нажатия клавиш и даже – как показано в новейших исследованиях – геометрические параметры траектории курсора, пытаясь отличить реального человека от автоматизированного скрипта. Злоумышленники, использующие этот метод, требуют от жертвы решить CAPTCHA-подобные задачи или совершить определённые действия мышью, доказывая «человечность», прежде чем активировать основную нагрузку.

В ответ на эти вызовы современные песочницы эволюционируют в нескольких направлениях. Во-первых, развиваются технологии *голого железа (bare-metal)*, исключающие признаки виртуализации: вредоносное ПО исполняется на реальном оборудовании, а не в гипервизоре, что устраняет большинство системных проверок. Во-вторых, внедряются *поведенческие стимуляторы* – системы, эмулирующие активность реального пользователя (случайные движения мыши, нажатия клавиш, открытие документов), что позволяет «обмануть» проверки пользовательской активности. В-третьих, используются *многоуровневые очереди детонации*, где файл последовательно обрабатывается в средах с возрастающим уровнем реалистичности: от быстрых эмуляторов до полноценных bare-metal систем. Примером такого подхода является VMRay Sandbox, сочетающий высокоинтерактивную песочницу с возможностями bare-metal исполнения.

Особого внимания заслуживает платформа MetaDefender Aether от OPSWAT, представляющая собой AI-нативный движок для быстрого обнаружения zero-day, который обрабатывает файлы через четыре прогрессивных уровня: репутационный анализ, адаптивную

песочницу, ML-оценку угроз и схожестный охотник за угрозами (similarity-based threat hunting). Цепочка этих уровней обеспечивает 99,9% эффективность обнаружения zero-day при ресурсной эффективности, в 100 раз превышающей традиционные VM-песочницы, причём почти половина угроз отсеивается уже на первом уровне репутации, а остальные проходят последовательную углублённую проверку.

Таблица 4

Сравнение типов песочниц и их устойчивости к методам обхода

Тип песочницы	Признаки виртуализации	Эмуляция пользователя	Временные ограничения	Устойчивость к обходу
Эмулятор (QEMU)	Высокие	Отсутствует	Есть (жесткие)	Низкая
VM-песочница (VMWare)	Средние	Базовая	Есть	Средняя
Bare-metal песочница	Отсутствуют	Продвинутая	Минимальные	Высокая
Гибридная (AI + VM)	Низкие	Интеллектуальная	Адаптивные	Очень высокая

Гибридные подходы: интеграция поведенческого анализа, песочниц и искусственного интеллекта

Ни один из рассмотренных методов, будучи использованным в изоляции, не обеспечивает полной защиты от zero-day угроз, поскольку поведенческий анализ на конечных точках ограничен производительностью и может быть обойдён через задержки активации, а песочницы, несмотря на свою глубину, не всегда масштабируются на весь поток входящих файлов в реальном времени. Оптимальным решением, признанным ведущими экспертами, является построение многоуровневой гибридной системы, где файлы проходят последовательную фильтрацию через слои с возрастающей вычислительной стоимостью, но также с возрастающей точностью.

Первый уровень – **репутационный анализ и статические сигнатуры** – отсеивает известные угрозы за миллисекунды, используя глобальные базы репутации файлов и хэшей. Второй уровень – **лёгкая эмуляция и поведенческие правила** – анализирует структуру файла и начальные признаки активности, выявляя простые обфускации и упаковщики. Третий уровень – **песочница с ML-анализом** – запускает файл в изолированной среде, фиксирует полный трейс активности и передаёт его в модели машинного обучения для классификации. Четвёртый уровень – **продвинутый анализ с участием человека** – резервируется для наиболее сложных случаев, когда автоматические системы выдают неопределённый результат.

Ключевым элементом этой архитектуры является *цикл обратной связи*, когда результаты анализа из песочниц используются для переобучения моделей поведенческого анализа на конечных точках, что позволяет постепенно смещать «тяжёлые» проверки влево, к более ранним уровням фильтрации. Более того, современные платформы (такие как MetaDefender Aether) внедряют механизмы *схожестного охотника за угрозами*, когда поведенческие отпечатки файла сравниваются с базой данных из более чем 100 миллионов проанализированных образцов, позволяя автоматически атрибутировать файл к известным семействам вредоносного ПО и кампаниям на основе сходства поведения, даже если конкретный образец ранее не встречался.

Практические рекомендации по построению системы защиты от zero-day угроз.

Исходя из проведённого анализа, можно сформулировать следующие практические рекомендации, адресованные как корпоративным центрам безопасности (SOC), так и разработчикам средств защиты.

Во-первых, необходимо внедрить **многоуровневую архитектуру детонации**, где файлы проходят последовательную обработку: репутационный слой → статические сигнатуры → эмуляция → VM-песочница → bare-metal анализ. Такая стратификация, обеспечивая баланс между скоростью и точностью, позволяет отсеивать до 90% угроз на первых двух уровнях, резервируя ресурсоёмкий анализ только для подозрительных файлов.

Во-вторых, следует использовать **AI-модели для поведенческого анализа**, причём предпочтение следует отдавать моделям с объяснимостью (explainable AI), таким как графовые нейронные сети или байесовские сети, позволяющие аналитику понимать, почему файл был классифицирован как вредоносный. Модели на основе графов (GraphShield) демонстрируют наилучшее соотношение точности и интерпретируемости.

В-третьих, **песочницы должны быть защищены от методов обхода**: использование bare-metal сред для критических анализов, внедрение стимуляторов пользовательской активности (рандомизированные движения мыши, нажатия клавиш), а также применение технологии «временных ловушек» – заведомо завышенных таймаутов, чтобы вредоносное ПО, использующее временные проверки, не успело «заснуть».

В-четвёртых, **цикл обратной связи** между песочницами и конечными точками должен быть автоматизирован: поведенческие сигнатуры, выявленные в песочницах, должны в течение минут распространяться на все конечные точки, позволяя блокировать аналогичные атаки на ранних стадиях.

В-пятых, **обучение персонала** остаётся критическим фактором: поскольку даже самая совершенная автоматическая система даёт ложные срабатывания, аналитики должны уметь интерпретировать результаты ML-моделей, различать истинные аномалии и легитимные отклонения. Рекомендуется внедрение практики «регулярных красных командно-синих учений», в ходе которых проверяется устойчивость как технических средств, так и процедур реагирования.

Заключение.

Проведённое исследование позволяет сделать вывод о том, что защита от атак с использованием zero-day уязвимостей требует комплексного, многоуровневого подхода, объединяющего поведенческий анализ, технологии песочниц и искусственный интеллект. Основные выводы таковы.

Во-первых, масштаб угрозы zero-day продолжает расти: 90 эксплуатируемых уязвимостей в 2025 году, рекордная доля корпоративных целей (48%) и коммерциализация рынка эксплойтов создают беспрецедентные риски для организаций. Во-вторых, традиционные сигнатурные методы защиты окончательно утратили свою эффективность на фоне 1548% роста новых уникальных образцов вредоносного ПО и 23% zero-day, обходящих классические детекторы. В-третьих, поведенческий анализ на основе машинного обучения, особенно с использованием графовых нейронных сетей и LLM-эмбеддингов, демонстрирует высокую эффективность (F1 до 99,5%) при контролируемом уровне ложных срабатываний (<1%). В-четвёртых, песочницы эволюционируют от простых VM-сред к bare-metal системам с AI-усилением, однако злоумышленники активно используют методы обхода (системные проверки, временные задержки, проверки пользовательской активности), что требует постоянного совершенствования защитных механизмов.

Перспективы дальнейших исследований связаны с развитием квантово-устойчивых методов поведенческого анализа, интеграцией LLM-агентов для автономного реверс-инжиниринга неизвестных образцов, а также созданием децентрализованных систем обмена поведенческими сигнатурами (threat intelligence на блокчейне), позволяющих сокращать время между обнаружением zero-day в одной организации и защитой всех остальных.

Список литературы:

1. Google Threat Intelligence Group. Look What You Made Us Patch: 2025 Zero-Days in Review. – Google Cloud Blog, March 6, 2026.

2. WatchGuard Technologies. Over 1500% Increase in New, Unique Malware Highlights Growing Security Complexity. – GlobeNewsWire, Feb. 19, 2026.

3. Группа компаний «Гарда». Опрос российских компаний об угрозах zero-day. – Компьютерра, 13 марта 2025.

4. Al Siam, A., Faruqui, N., Azad, A. et al. Securing the unseen: A comprehensive exploration review of AI-powered models for zero-day attack detection. – Expert Systems, Vol. 43, Issue 3, 2026.

5. Perera, W.S., Jiang, H. BEACON: Behavioral Malware Classification with Large Language Model Embeddings and Deep Learning. – Computing Research Repository, 2025