УДК 004

Биктубаева К.С.
Уфимский университет науки и технологий, Уфа

Научный руководитель:
Шафиков М.Р.
Уфимский университет науки и технологий, Уфа

АДВЕРСАРНЫЕ АТАКИ НА ZERO TRUST-СИСТЕМЫ: УЯЗВИМОСТИ И МЕТОДЫ ЗАЩИТЫ

Аннотация. Архитектура Zero Trust считается перспективной для защиты от современных киберугроз. Однако ее внедрение не гарантирует абсолютной защиты: злоумышленники разрабатывают адверсарные атаки, направленные на обход механизмов непрерывной аутентификации, манипуляцию контекстно-зависимыми политиками и эксплуатацию уязвимостей машинного обучения, используемого в ZT-системах. В данной статье анализируются основные уязвимости архитектуры

нулевого доверия и предлагаются стратегии защиты, такие как адаптивная аутентификация, устойчивые к манипуляциям модели машинного обучения и гибридные методы обнаружения аномалий.

Ключевые слова: Zero Trust, адверсарные атаки, машинное обучение, динамический контроль доступа, кибербезопасность.

Zero Trust (ZT, нулевое доверие) – концепция информационной безопасности, предполагающая отсутствие доверия к каким бы то ни было объектам IT-инфраструктуры организации, будь то пользователи, устройства или программы. Концепция ZT радикально меняет парадигму кибербезопасности, заменяя традиционные модели на постоянную верификацию пользователей, устройств и транзакций [1].

Адверсарные атаки (adversarial attacks, состязательные атаки) – это тип атак, направленных на обман или вывод из строя моделей машинного обучения. В отличие от традиционных кибератак, которые нацелены на взлом систем или кражу данных, адверсарные атаки манипулируют входными данными модели, заставляя ее выдавать неправильные результаты, совершать ошибки или демонстрировать неожиданное поведение [2].

С 2021 по 2024 г. наблюдается значительный рост числа кибератак, что затрагивает и системы, основанные на концепции ZT. По данным FACCT с 2021 по 2023 г. число инцидентов увеличилось в 3 раза. Пик пришелся на 2022–2023 г., что было связано с массовым переходом компаний на ZT-модель без должной адаптации механизмов защиты. В 2024 г. темпы роста снижаются, однако значения остаются высокими. Это свидетельствует о том, что, несмотря на постепенную адаптацию ZT, угрозы эволюционируют.

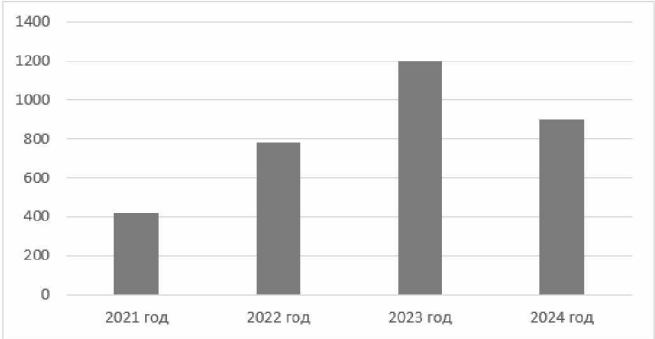

Рис. 1. Количество адверсарных атак на Zero Trust-системы за 2021–2024 гг.

Архитектура (ZT) демонстрирует высокую устойчивость к традиционным кибератакам, но все же подвержена и остается уязвимой к ряду адверсарных атак [3]. Рассмотрим основные векторы атак, которые можно классифицировать на три основные категории:

1. Уязвимости аутентификации и авторизации. Механизмы аутентификации и авторизации являются основными уязвимостями в архитектурах ZT. Ключевые угрозы включают компрометацию учетных данных через фишинг и атаки типа Credential Stuffing, обход многофакторной аутентификации (например, с помощью SIM-свопинга и MITM-атак), а также кражу токенов доступа (JWT/OAuth через XSS/CSRF). Особенно опасной является подмена refresh_token, которая позволяет сохранять доступ даже после изменения учетных данных. В среде ZT, где каждый запрос требует аутентификации, эти уязвимости могут стать критическими точками отказа.

2. Уязвимости микросервисной архитектуры и API. Переход на микросервисную архитектуру в Zero Trust-системах, несмотря на свои преимущества, создает новые векторы для потенциальных атак, особенно через уязвимости API. Согласно OWASP API Top 10, наиболее опасными являются атаки типа BOLA (Broken Object Level Authorization), позволяющие злоумышленникам получать несанкционированный доступ к данным через манипуляцию идентификаторами объектов, а также различные виды инъекций (SQLi, NoSQLi), эксплуатирующие недостаточную валидацию входных параметров при запросах к базам данных [4]. Не менее серьезной угрозой является подмена параметров API, когда злоумышленники изменяют HTTP-заголовки или query-параметры с целью эскалации привилегий. Эти проблемы становятся серьезнее из-за плохого разделения сети между микросервисами. Взломав один сервис, злоумышленник может атаковать другие.

3. Уязвимости систем мониторинга и машинного обучения. Злоумышленники стали активно использовать стандартные инструменты администрирования (PowerShell, RDP) для маскировки вредоносных действий под обычную работу администраторов. В тоже время стремительно развиваются adversarial ML-атаки, злоумышленники искусно подделывают данные, чтобы аномалии оставались незамеченными [5]. Атакующие специально создают множество ложных тревог, чтобы перегрузить аналитиков и скрыть настоящие атаки. Также они используют медленный подбор данных, который часто проходит незамеченным из-за неправильных настроек систем защиты. Данные методы демонстрируют, что даже современные системы мониторинга на основе машинного обучения (ML) не обладают абсолютной устойчивостью, поэтому необходимо постоянно совершенствовать алгоритмы обнаружения [6].

Для эффективного противодействия выявленным уязвимостям Zero Trust-систем, выделяют следующие меры защиты:

Для защиты механизмов аутентификации и авторизации в Zero Trust-системах необходимо реализовать многоуровневый подход. Во-первых, использовать современные методы входа – биометрию (сканирование отпечатков пальцев или лица) и физические ключи безопасности вместо ненадежных SMS-кодов, дополняя это проверкой привычного

местоположения и устройств пользователей для выявления подозрительных попыток доступа; во-вторых, усиленно защищать токены доступа – устанавливать короткий срок их действия (не более 15 минут), применять электронную подпись и систему мгновенной блокировки при компрометации; в-третьих, регулярно проводить комплексные проверки – контролировать настройки системы входа, проверять ссылки перенаправления, шифровать серверы выдачи токенов и своевременно устранять обнаруженные уязвимости.

Для защиты микросервисов и API в ZT-средах необходимо реализовать комплекс мер. На уровне API-шлюзов критически важны: проверка прав доступа к каждому объекту, блокировать вредоносные команды, ограничивать частоту обращений [7]. Использовать защищенные соединения, выдавать минимальные права, применять принцип «запрещено все, что не разрешено». Обязательным элементом является мониторинг всех API-вызовов с анализом аномалий и автоматической блокировкой массовых или подозрительных запросов.

Для эффективного противодействия сложным атакам, таким как Adversarial ML, необходимо совершенствовать системы мониторинга и машинного обучения. Ключевыми мерами являются: внедрение устой--чивых ML-моделей, включая алгоритмы обнаружения аномалий и регулярного проверять их на уязвимость, совмещать поведенческий анализ с проверкой известных шаблонов атак, а также автоматически адаптировать чувствительность систем для выявления медленных и скрытых атак.

Исследование адверсарных атак на системы ZT-системы имеет важное практическое значение для современной кибербезопасности, поскольку выявляет ключевые уязвимости перспективной архитектуры и предлагает конкретные меры защиты. Полученные результаты позволяют организациям, внедряющим ZT, заранее устранять слабые места в механизмах аутентификации, защите API и системах мониторинга. Но, несмотря на существующие уязвимости, концепция ZT при правильной реализации и постоянном совершенствовании остается одним из наиболее перспективных подходов к обеспечению кибербезопасности в современных условиях.

Список использованных источников:

1. Национальный институт стандартов и технологий (NIST). Zero Trust Architecture (NIST Special Publication 800-207). 2020. URL: https://newsletter.radensa.ru/wp-content/uploads/2025/ 01/NIST-SP800-207.pdf.

2. Krivonogov A.A., & Petrov V.S. (2023). Explainable Adversarial Mitigation Framework for Zero-Trust Cyber Warfare. ResearchGate. URL: https://www.researchgate.net/publication/375071924_Explainable_Adversarial_Mitigation_Framework_for_Zero-Trust_Cyber_Warfare.

3. Многоагентные системы как технологическая база реализации концепции нулевого доверия / С.С. Валеев, Н.В. Кондратьева,

М.Б. Гузаиров, А.С. Исмагилова // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление. 2024. № 3. С. 116–123.

4. Open Web Application Security Project (OWASP). OWASP API Security Top 10 2023. 2023. URL: https://owasp.org/API-Security.

5. Smith J., Johnson A. Advanced Adversarial Attacks on Machine Learning Systems. Springer, 2022. 315 p. URL: https://link.springer.com/book/10.1007/978-3-030-99772-4.

6. TechRepublic. Zero Trust Security: A Cheat Sheet (Free PDF). 2023. URL: https://www.techrepublic.com/resource-library/downloads/zero-trust-security-a-cheat-sheet-free-pdf/.

7. Zhang L., Wang Y. Adversarial Attacks on Zero Trust Security Systems: Methods and Countermeasures // arXiv preprint. 2021. URL: https://arxiv.org/pdf/2112.02797.

Biktubaeva K.S.
Ufa University of Science and Technology, Ufa

Scientific supervisor:
Shafikov M.R.
Ufa University of Science and Technology, Ufa

ADVERSARIAL ATTACKS ON ZERO TRUST SYSTEMS: VULNERABILITIES AND PROTECTION METHODS

Abstract. The Zero Trust architecture is considered promising for protection against modern cyber threats. However, its implementation does not guarantee absolute protection: attackers develop adversarial attacks aimed at bypassing continuous authentication mechanisms, manipulating context-sensitive policies, and exploiting machine learning vulnerabilities used in ZT systems. This article analyzes the main vulnerabilities of the zero-trust architecture and suggests protection strategies such as adaptive authentication, manipulation-resistant machine learning models, and hybrid anomaly detection methods.

Keywords: Zero Trust, adversarial attacks, machine learning, dynamic access control, cybersecurity.