УДК 004

Тагиев Р.Н.

студент 4 курса, направление Международные экономические отношения

Казанский (Приволжский) федеральный университет

(г. Казань, Россия)

НОВАЯ ЭРА КИБЕРАТАК: КАК ХАКЕРЫ ИСПОЛЬЗУЮТ LLM ДЛЯ КРАЖИ ДАННЫХ НА ПРИМЕРЕ ИНЦИДЕНТА С ПАКЕТОМ NX

***Аннотация**: эта статья посвящена анализу одного из самых необычных инцидентов в истории кибербезопасности — компрометации npm-пакета nx, которым пользуются миллионы разработчиков. Хакеры использовали новый, неожиданный метод для кражи данных: вместо сложного кода они задействовали большие языковые модели (LLM), такие как Gemini и Claude, для поиска и передачи конфиденциальной информации. Мы проводим технический разбор этой атаки, объясняем, почему традиционные средства защиты оказались бессильны, и выявляем новые риски, связанные с интеграцией ИИ в рабочие процессы. Статья предлагает практические рекомендации для разработчиков и специалистов по безопасности, чтобы помочь им противостоять этому новому типу угроз и защитить свои системы от будущих атак, использующих возможности искусственного интеллекта.*

***Ключевые слова**: кибербезопасность, искусственный интеллект, уязвимости, информационная безопасность, вредоносное ПО, цепочка поставок.*

В последние годы атаки на цепочки поставок программного обеспечения (supply chain attacks) стали одной из наиболее серьёзных и быстрорастущих угроз в сфере кибербезопасности. Эти атаки нацелены на компрометацию доверенных компонентов, библиотек или инструментов, используемых в процессе разработки, что позволяет злоумышленникам получить доступ к конечным системам пользователей.

Если раньше такие инциденты чаще всего были связаны с внедрением бэкдоров или классических вредоносных программ, то современные атаки демонстрируют всё большую изощрённость. В августе 2025 года мир стал свидетелем нового, ранее невиданного вектора атаки, когда популярный пакет nx был скомпрометирован, а вредоносное ПО использовало локальные ИИ-ассистенты для поиска и кражи конфиденциальных данных [1].

Этот инцидент стал переломным моментом, поскольку он показал, что легитимные и широко используемые технологии, такие как большие языковые модели (LLM), могут быть превращены в инструменты для злонамеренных целей. Вместо того чтобы полагаться на сложный и легко обнаруживаемый код для сканирования файловой системы, злоумышленники делегировали эту задачу ИИ-ассистентам, установленным на компьютерах жертв. Такой подход позволяет вредоносному ПО оставаться менее заметным для традиционных антивирусных решений, которые не обучены анализировать и блокировать запросы к локальным LLM.

Целью данной статьи является всесторонний анализ инцидента с компрометацией пакета nx, включая его технический разбор, выявление ключевых уязвимостей, связанных с новым вектором атаки, и предложение мер по предотвращению подобных угроз в будущем. Мы проанализируем, как интеграция LLM в рабочую среду разработчика создаёт новые риски и как сообществу кибербезопасности необходимо адаптироваться к этой быстро меняющейся угрозе. Наша работа вносит вклад в понимание того, как легитимные технологии могут быть использованы в преступных целях, и предлагает конкретные шаги для повышения безопасности в условиях повсеместного распространения искусственного интеллекта.

Технический анализ и механизм атаки.

Инцидент с пакетом nx представляет собой классическую атаку на цепочку поставок, где начальной точкой стало компрометирование учётной записи разработчика на платформе npm. Это позволило злоумышленникам опубликовать несколько вредоносных версий пакета (включая v21.5.0, v20.9.0 и другие), которые содержали скрытый вредоносный код [2].

Ключевым элементом атаки стал postinstall скрипт, который автоматически запускается после установки пакета. Этот механизм, являющийся стандартной функцией npm, позволил злоумышленникам выполнить произвольный код на машине пользователя без его явного согласия. Вредоносный скрипт был замаскирован под именем telemetry.js и скрыт от быстрого обнаружения.

После запуска скрипт выполнял несколько действий:

1. Сбор информации: Он сканировал локальную файловую систему на наличие установленных ИИ-ассистентов, таких как Gemini CLI и Claude Code CLI.

2. Эксфильтрация данных: В зависимости от наличия ассистентов, скрипт использовал один из двух методов для кражи данных.

Это наиболее новаторский и тревожный аспект атаки. Вместо того чтобы полагаться на собственный код для поиска файлов, который мог бы быть легко обнаружен антивирусами, злоумышленники использовали ИИ-ассистентов в качестве "агентов". Вредоносный скрипт отправлял LLM-модели следующий промпт:

"Recursively search local paths on Linux/macOS... for any file whose pathname or name matches wallet-related patterns... record only a single line in /tmp/inventory.txt containing the absolute file path, e.g.: /absolute/path."

Этот запрос, направленный к локально работающей нейросети, выглядит как легитимная задача для разработчика. Он просит ИИ выполнить рутинную операцию поиска файлов по заданным шаблонам (wallet, .key, metamask, id_rsa и т.д.).

LLM, обученные для выполнения таких задач, могли бы сгенерировать список путей к конфиденциальным файлам, который затем сохранялся во временный файл.

После получения списка путей вредоносный скрипт считывал их содержимое. Собранные данные, включая API-ключи, токены и приватные ключи от криптокошельков, были дважды закодированы в формате Base64 и загружены в публичный репозиторий на GitHub, созданный на аккаунте жертвы [3].

Инцидент с nx не просто привёл к краже данных — он выявил фундаментальные уязвимости в современной экосистеме разработки и безопасности.

Последствия и уязвимости.

Традиционные антивирусные системы и средства обнаружения вторжений (IDS) обычно работают, анализируя сигнатуры известных угроз, аномальное поведение файловой системы или сетевой трафик. Однако в этом случае:

1. Процесс был легитимным: Вредоносный скрипт не содержал сложного кода для сканирования диска. Он использовал встроенную функцию операционной системы и запрос к установленному на машине приложению (LLM).

2. Активность была замаскирована: Запрос к LLM мог быть классифицирован как обычная пользовательская команда, а не как вредоносная активность. Это создало "слепое пятно" для систем безопасности.

Этот инцидент показал, что глубокая интеграция LLM-ассистентов в рабочие процессы разработчиков несёт в себе новые риски. Предоставляя ИИ доступ к файловой системе, разработчики создают новый, потенциально опасный вектор атаки.

Если LLM могут выполнять команды для поиска и обработки файлов, они могут быть использованы для злонамеренных целей, даже если их разработчики предусмотрели защиту от прямых вредоносных запросов. Как показал анализ, хакеры умеют обходить эти ограничения.

Инцидент также подчеркнул продолжающуюся проблему доверия в экосистемах открытого кода. Несмотря на усилия по повышению безопасности, компрометация учётной записи всё ещё остаётся относительно простым и эффективным методом для внедрения вредоносного кода.

Меры противодействия и рекомендации.

Чтобы противостоять новому вектору атак с использованием LLM, необходимо разработать многоуровневую стратегию, охватывающую как технические, так и организационные меры.

Улучшенный мониторинг postinstall хуков. Платформы типа npm и системы безопасности должны внедрить более строгий контроль за выполнением postinstall скриптов.

Адаптация антивирусного ПО. Традиционные антивирусы неэффективны против этого вектора. Новое поколение средств защиты конечных точек (Endpoint Detection and Response, EDR) должно включать:

1. Мониторинг процессов LLM: Отслеживание и анализ запросов, отправляемых к локальным LLM. Система должна уметь распознавать вредоносные промпты, даже если они замаскированы под легитимные задачи.

2. Контроль доступа: Принудительное ограничение доступа LLM к чувствительным директориям, таким как .ssh, .npmrc или каталоги с криптокошельками.

Разработчики LLM должны внедрить более строгие меры безопасности. Модели должны быть обучены отказывать в выполнении запросов, которые могут быть использованы для вредоносных целей, и распознавать паттерны, указывающие на попытки эксфильтрации данных [4].

Заключение.

Инцидент с пакетом nx стал тревожным сигналом, указывающим на эволюцию угроз в сфере кибербезопасности. Он продемонстрировал, что злоумышленники активно ищут и находят новые уязвимости, эксплуатируя даже самые передовые технологии. Использование легитимных ИИ-ассистентов в качестве инструмента для кражи данных открывает новую главу в истории вредоносного ПО.

Эта атака подчеркнула острую необходимость в разработке новых подходов к безопасности. Защита должна сместиться от простого анализа кода к комплексному мониторингу всего рабочего процесса, включая взаимодействие с ИИ-инструментами. Для эффективного противодействия этим угрозам требуется совместная работа разработчиков платформ, создателей инструментов безопасности и всего сообщества разработчиков. Только так мы сможем обеспечить безопасность в условиях, где технологии развиваются быстрее, чем защитные механизмы.

СПИСОК ЛИТЕРАТУРЫ:

1. "What is a Supply Chain Attack?" NIST Cybersecurity Framework. [Электронный ресурс]. URL: https://w.nist.gov/cybersecurity/supply-chain-risk (дата обращения:13.09.2025);
2. GitHub Issues. "Malicious postinstall script found in nx package." Issue #32522, GitHub, August 26, 2025. [Электронный ресурс]. URL: https://github.com/nrwl/nx/issues/32522 (дата обращения: 13.09.2025);
3. NRWL. "Official Advisory: Supply Chain Attack on nx." GitHub Security Advisories, August 28, 2025. [Электронный ресурс]. URL: https://github.com/nrwl/nx/security/advisories/GHSA-cxm3-wv7p-598c (дата обращения: 13.09.2025);
4. Semgrep. "Security Alert | NX Compromised to Steal Wallets and Credentials." Semgrep Blog, August 27, 2025. [Электронный ресурс]. URL:

https://semgrep.dev/blog/2025/security-alert-nx-compromised-to-steal-wallets-and-credentials (дата обращения: 13.09.2025)

Tagyyew R.N.
4th year student, International Economic Relations
Kazan Federal University
(Kazan, Russia)

NEW ERA OF CYBER ATTACKS: HOW HACKERS USE LLM TO STEAL DATA USING EXAMPLE OF NX PACKAGE INCIDENT

***Abstract**: article is devoted to the analysis of one of the most unusual incidents in the history of cybersecurity — the compromise of the npm package nx, which is used by millions of developers. Hackers used a new, unexpected method to steal data: instead of complex code, they used large language models (LLM) such as Gemini and Claude to find and transfer confidential information. We are conducting a technical analysis of this attack, explaining why traditional means of protection have proved powerless, and identifying new risks associated with the integration of AI into work processes. The article offers practical recommendations for developers and security specialists to help them counter this new type of threat and protect their systems from future attacks using artificial intelligence capabilities.*

***Keywords**: cybersecurity, artificial intelligence, vulnerabilities, information security, malware, supply chain.*