

Генерация плана проведения penetration-testing по текстовому описанию инфраструктуры с использованием нейросети GPT-5.1

Саяпин Никита Анатольевич

Приамурский государственный университет им.Шолом-Алейхема

Студент

Аннотация

Цель исследования — продемонстрировать процесс автоматической генерации структурированного плана penetration-testing (тестирования на проникновение) по текстовому описанию ИТ-инфраструктуры с использованием мультимодальной нейросети GPT-5.1. В работе рассматриваются принципы интерпретации моделью компонентов инфраструктуры, угроз и возможных векторов атаки. Оценивается корректность и полнота создаваемого плана pentest-a, включая этапы сбора информации, анализа сервисов, моделирования атак и формирования отчётности.

Ключевые слова: GPT-5.1, penetration testing, кибербезопасность, автоматизация аудита, инфраструктура, threat modeling, ИИ.

Generation of a penetration-testing plan based on a text description of the infrastructure using the GPT-5.1 neural network

Sayapin Nikita Anatolyevich

Sholom-Aleichem Priamursky State University

Student

Abstract

The purpose of the study is to demonstrate the process of automatic generation of a structured penetration—testing plan based on a text description of the IT infrastructure using the GPT-5 multimodal neural network.1. The paper discusses the principles of the model's interpretation of infrastructure components, threats, and possible attack vectors. The correctness and completeness of the created pentest plan is evaluated, including the stages of information collection, analysis of services, attack modeling and reporting.

Keywords: GPT-5.1, penetration testing, cybersecurity, audit automation, infrastructure, threat modeling, AI.

1. Введение

1.1 Актуальность

Проведение тестирования на проникновение (pentest) является одним из ключевых этапов оценки безопасности ИТ-инфраструктуры. Традиционно разработка плана pentest-a требует участия специалистов по

кибербезопасности и занимает значительное время, поскольку необходимо учитывать топологию сети, набор сервисов, операционных систем, уровни доступа и потенциальные уязвимости.

Использование мультиагентных нейросетей нового поколения, таких как GPT-5.1, позволяет автоматизировать процесс подготовки плана pentest-а по текстовому описанию инфраструктуры. Модель способна интерпретировать сетевую архитектуру, определять потенциальные векторы атак и формировать структурированный документ, соответствующий стандартам тестирования безопасности (OWASP, PTES, MITRE ATT&CK и др.).

1.2 Обзор исследований

И. А. Трещев предложил подход к проведению тестирования на проникновение, в котором особое внимание уделяется автоматизированному выполнению атак и последующему анализу уязвимостей программного и аппаратного обеспечения, что позволяет формализовать этапы pentest-а и частично унифицировать сценарии проверки безопасности информационных систем [1]. И. А. Белянова и Е. Д. Пойманова рассмотрели методики тестирования на проникновение веб-приложений, выделив основные этапы анализа и показав, что использование специализированных инструментов и шаблонных сценариев существенно повышает эффективность поиска типовых уязвимостей, таких как SQL-инъекции, XSS и другие ошибки уровня приложения [2]. А. С. Рахманов проанализировал возможности автоматизации тестирования безопасности и показал, что интеграция сканеров уязвимостей (XSpider, Zenmap и др.) и фреймворков эксплуатации (Metasploit) в единый процесс позволяет сократить долю ручных операций и приблизить процедуру pentest-а к конвейерной модели DevSecOps [3]. Н. М. Амосов и соавторы исследовали особенности использования методов машинного обучения при проведении pentest-ов, продемонстрировав, что модели машинного обучения могут применяться как для интеллектуального выбора сценариев атаки, так и для автоматизированной обработки результатов сканирования и корреляции обнаруженных уязвимостей [4]. Д. Е. Белянский рассмотрел методы моделирования угроз безопасности и показал, что использование гибридного подхода, сочетающего STRIDE и CVSS, позволяет не только систематизировать потенциальные угрозы, но и приоритизировать их по уровню риска, что делает такие модели удобной основой для построения планов тестирования и выбора наиболее критичных векторов атак [5].

1.3 Цель исследования

Цель исследования — продемонстрировать возможности GPT-5.1 в автоматической генерации профессионального плана pentest-а по текстовому описанию инфраструктуры и показать применимость такого подхода в образовательной и практической деятельности.

2 Материалы и методы

В рамках исследования использовалась мультимодальная модель GPT-5.1, способная обрабатывать текстовые описания технических систем и формировать структурированные документы.

Проверка работы модели проводилась как пошаговый процесс, включающий формирование текстового описания инфраструктуры, создание промптов к GPT-5.1, генерация плана pentest-a, сравнение результата с методологиями PTES и MITRE ATT&CK.

Для демонстрации использовалась типовая корпоративная сеть со следующими компонентами:

- DMZ-зона (web-серверы, reverse proxy);
- внутренняя сеть сотрудников;
- серверная зона (AD, базы данных, API-сервисы);
- VPN-доступ;
- межсетевые экраны;
- сегментация VLAN.

3. Результаты

3.1 Генерация плана внешнего тестирования (External Pentest)

В первом сценарии, связанном с оценкой внешнего периметра, модель уверенно обработала описание инфраструктуры, включающее публичный диапазон IP-адресов, DMZ-зону, веб-сервер и VPN-шлюз. GPT-5.1 корректно сформировала структурированный план, включающий этапы разведки, сканирования сервисов, анализа веб-уязвимостей и проверки стойкости VPN-аутентификации.

Промпт:

«Generate an external penetration testing plan based on PTES methodology.

Infrastructure:

- Public IP range 203.0.113.0/28;
- DMZ with nginx web server;
- Reverse proxy with HTTPS termination;
- VPN gateway with password authentication.

Focus on reconnaissance, service scanning, SSL analysis, VPN security testing. »

Сгенерированный план корректно отражал структуру, заданную в описании. Модель автоматически добавила этап анализа TLS-конфигурации, подбор возможных точек входа через misconfigured веб-серверы, а также инструменты, применяемые на каждом этапе, что улучшило практическую применимость результата.

Итоговый документ отличался ясной структурой и полнотой, что позволило быстро оценить его соответствие профессиональным методикам pentest-a (Рис. 1).

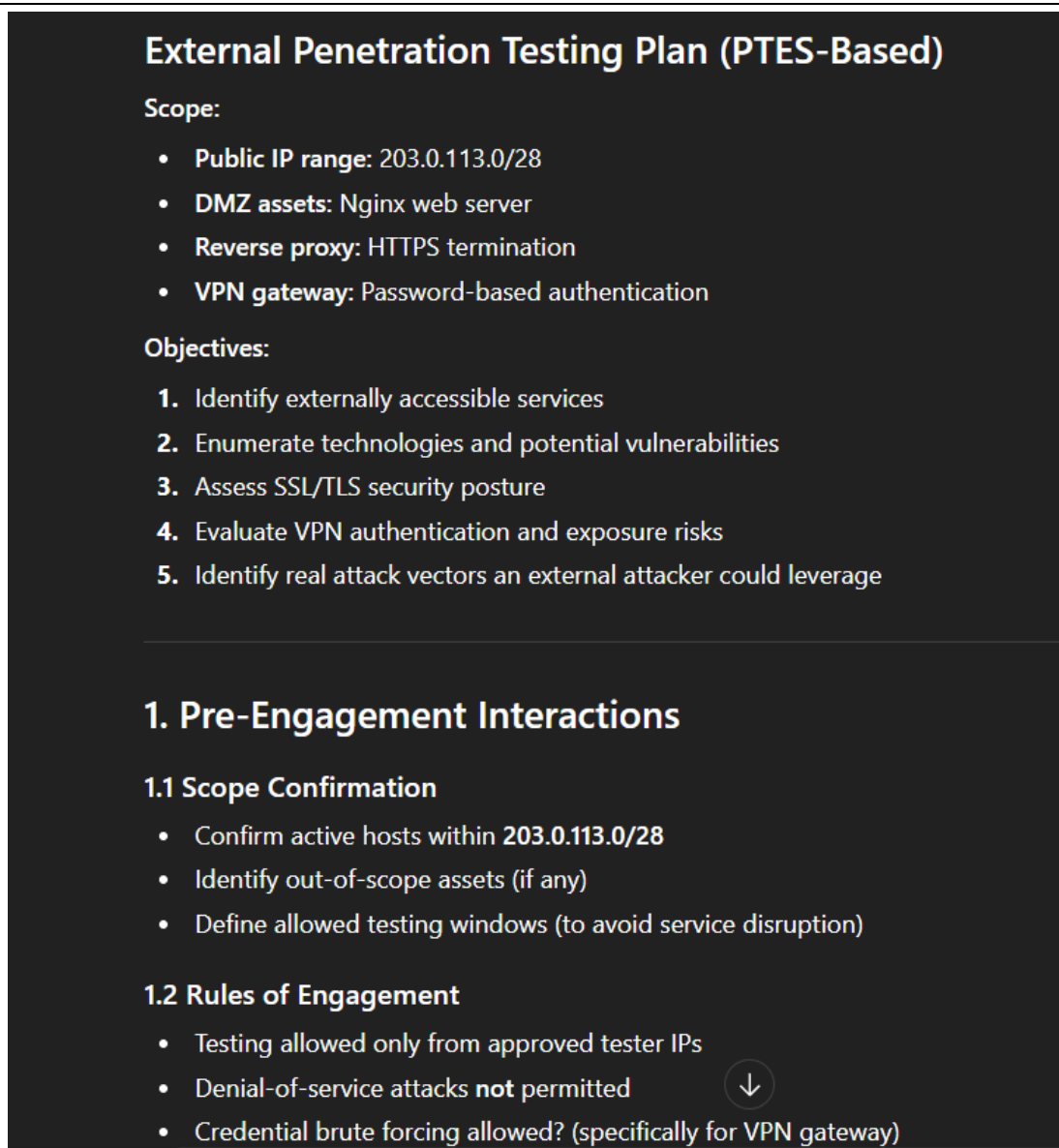


Рисунок 1 — Генерация плана внешнего тестирования безопасности

3.2 План внутреннего тестирования и анализа доменной инфраструктуры (Internal Pentest)

Второй сценарий касался тестирования внутренней сети, содержащей доменную инфраструктуру Active Directory и несколько сегментов пользователей. Модель продемонстрировала понимание ключевых техник атак внутри корпоративной среды, корректно распределив действия по этапам: первичная оценка доступа, исследования сети, повышение привилегий и lateral movement (Рис. 2).

Промпт:

«Design an internal penetration testing plan.

Infrastructure:

- Windows domain (AD DS);
- Windows 10 workstations;
- VLANs for HR, Dev and Admin departments;
- MS SQL database server;

– Low-privileged user account provided.

Include privilege escalation, credential attacks, network reconnaissance and domain lateral movement. »

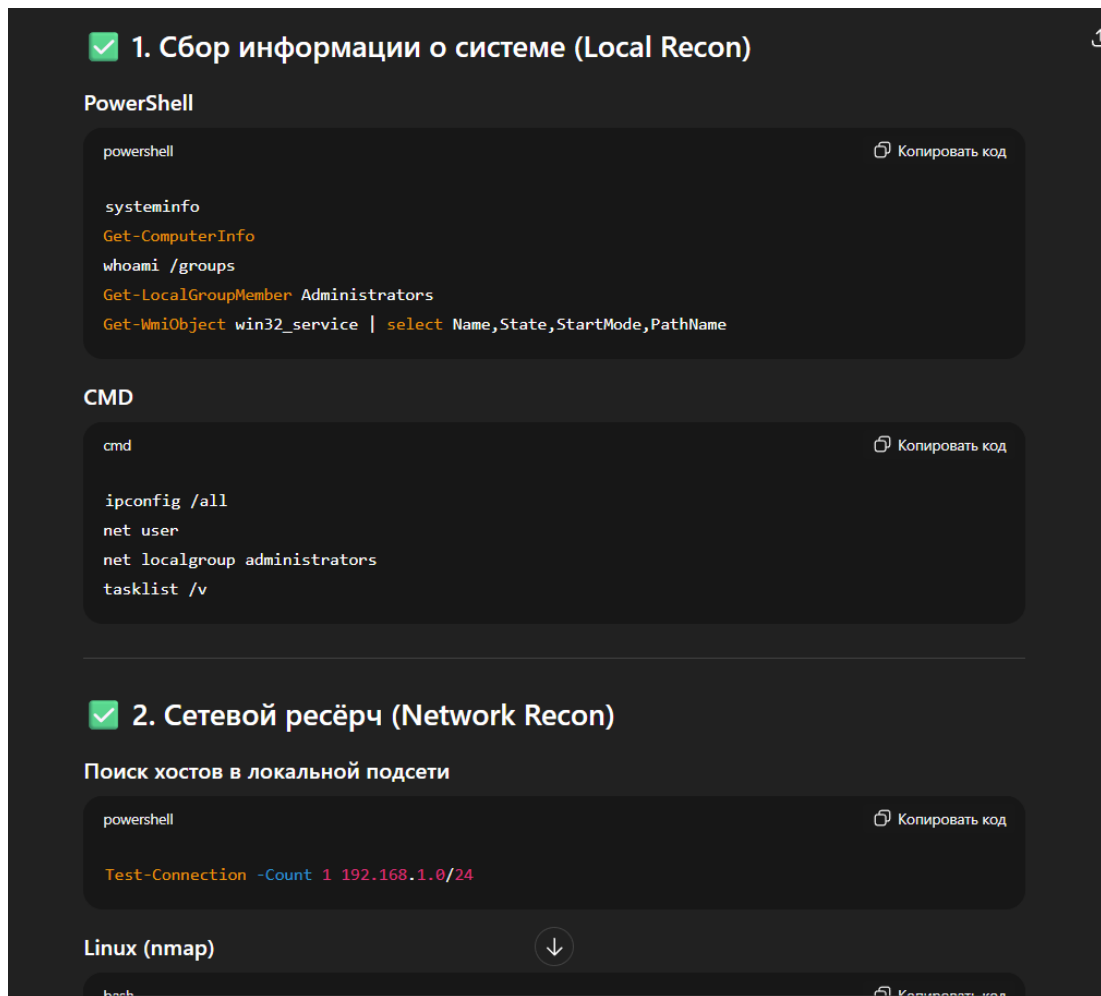


Рисунок 2 — План внутреннего анализа доменной инфраструктуры

Сгенерированный фрагмент содержал полный набор действий, включающий Kerberoasting, анализ GPO, попытки локального повышения привилегий, аудит паролей и перехват сетевых протоколов (LLMNR/NBT-NS). GPT-5.1 показала способность адаптироваться под особенности доменных сред и учитывать типичные векторы атак.

3.3 Анализ маршрутов атаки и threat modeling с использованием STRIDE

Третий сценарий иллюстрировал формирование расширенного аналитического документа, в котором требовалось объединить threat modeling и определение возможных путей атаки внутри инфраструктуры.

Промпт:

«Generate a security assessment plan including STRIDE threat modeling.

Infrastructure:

– Web server in DMZ;

- Internal Windows domain;
- Shared file services;
- VPN gateway;
- Segmented network.

Include attack paths, STRIDE table, and CVSS ranking of risks. »

Модель успешно обработала описание, выделив наиболее вероятные векторы атаки, такие как компрометация VPN, lateral movement через слабые пароли, атаки на AD-сервисы и повышение привилегий через уязвимые конфигурации. GPT-5.1 автоматически создала таблицу STRIDE, показав глубокое понимание классификации угроз (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege), что придало документу практическую аналитическую ценность (Рис. 3).

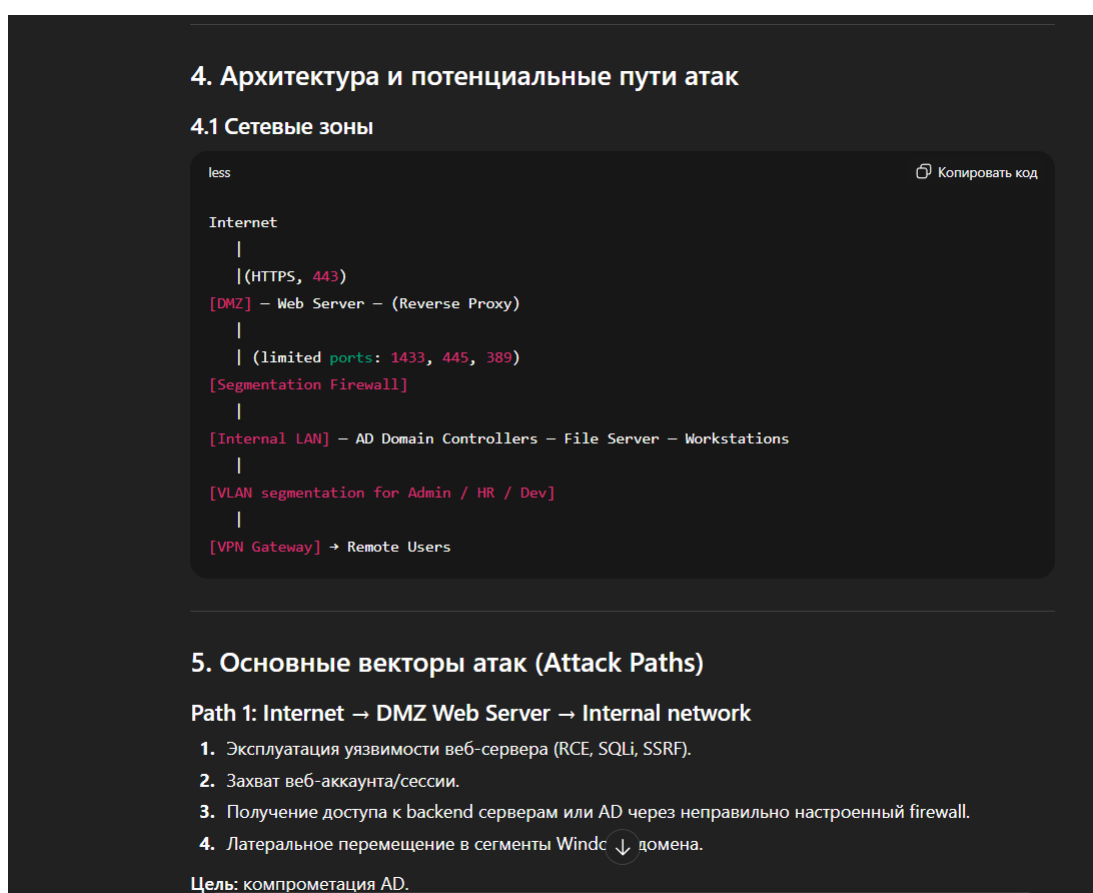


Рисунок 3 — Генерация threat-modeling и путей атак

3.4 Сравнение способов тестирования

Объём и детализация промпта напрямую влияют на глубину анализа. При кратких запросах модель формирует базовый план тестирования, тогда как подробное описание инфраструктуры позволяет получить документ, близкий к профессиональному отчёту по безопасности.

GPT-5.1 корректно учитывает различия между типами pentest-a (внешний, внутренний, инфраструктурный), если это явно указано.

Модель демонстрирует склонность применять best practices: включает threat modeling, CVSS-оценку, рекомендации по снижению рисков, перечисляет инструменты (Nmap, BloodHound, Burp Suite, Responder и др.).

Основные ограничения связаны с тем, что модель не имеет прямого доступа к реальным системам и не может оценить актуальные версии сервисов, поэтому результат требуется дополнять экспертом.

4 Выводы

Проведённое исследование демонстрирует, что использование мультимодальной нейросети GPT-5.1 для автоматической генерации плана тестирования на проникновение по текстовому описанию инфраструктуры является перспективным и практически применимым направлением в области кибербезопасности. Полученные результаты подтверждают высокую способность модели интерпретировать структуру корпоративных сетей, определять потенциальные угрозы и формировать детализированный план действий в соответствии с общепринятыми методологиями (PTES, OWASP, MITRE ATT&CK, STRIDE).

Библиографический список

1. Трещев И. А., Воробьев А. А. О подходе к проведению тестирования на наличие уязвимостей информационных систем // Производственные технологии будущего: от создания к внедрению. отв. ред. С. В. Белых. 2017. С. 175–182.
2. Белянова И. А., Пойманова Е. Д. Тестирование на проникновение веб-приложений // Информационные технологии в образовании : сборник статей науч.-практ. конф. студентов, аспирантов и молодых ученых. 2021. С. 44–47.
3. Рахманов А. С., Тутубалин П. И. Автоматизация тестирования безопасности веб-приложений // Молодежь и наука: актуальные проблемы фундаментальных и прикладных исследований. 2021. Ч. 3. С. 298–301.
4. Амосов Н. М., Тихонова Ю. Б. Особенности использования машинного обучения для проведения пентестов // Региональная информатика и информационная безопасность. Региональная общественная организация «Санкт-Петербургское Общество информатики, вычислительной техники, систем связи и управления». 2020. Вып. 9. С. 244–248.
5. Белянский Д. Е., Заславский Д. Ю., Хетагуров Я. А. Методы моделирования угроз безопасности информационным системам // Становление и развитие новой парадигмы инновационной науки в условиях современного общества. Уфа: ООО «ОМЕГА САЙНС». 2023. С. 27–29.