Red Teaming with ChatGPT

Finding vulnerabilities within the University infrastructure

Leiden University

March 2023

Introduction

In this assignment, you are going to use ChatGPT to explore potential exploits in applications, websites, and hosts from the University. The key objectives are:

- Learning how to use ChatGPT
- Learning about various vulnerabilities and potential exploits
- Learning about the process of responsible disclosure
- Applying this knowledge in practice

The basic goal of this assignment is to learn red teaming.

> *Red teaming is a simulation-based approach to evaluate and improve the effectiveness of plans, strategies, systems, and security measures. It involves a group of experts who simulate the thinking and actions of an adversary to challenge the assumptions and test the robustness of a plan, strategy or security measures. The objective of red teaming is to identify weaknesses and vulnerabilities and provide recommendations for improvement.*
>
> ChatGPT

While the above quote includes the mention of a group of experts to simulate the actions of an adversary, we don't believe this is necessary. In fact, we would like you to red team a novice adversary. Think of students, attempting to avoid having to do an exam by DDoS-ing the university or changing their grades, script kiddies trying to screw up some institutions for fun or inexperienced malicious actors that want to get money from ransomware attacks.

Tips

- Starting from a service may be easier than from a vulnerability. That way you can tell ChatGPT how it works and what the options are to discover possible weaknesses.
- Try to come up with "forgotten" services or websites. The main important systems like uSis are (hopefully) more protected, while some small side services may have not been updated in a while.
- You may work alone or in pairs. Enroll into a group on Brigthspace (*Groups-Bonus*) with your partner to be able to access the assignment and submit the report.

Responsible Disclosure

For this assignment, it is very important to obey the responsible disclosure guidelines from the university. During this assignment it could be possible that you will take actions that are prohibited by law. If you follow the guidelines set out in this section of the assignment, the university will not take legal action against you.

Leiden University guidelines on responsible disclosure are available here:
`https://www.staff.universiteitleiden.nl/binaries/content/assets/ul2staff/ict/responsible-disclosure-eng.pdf`

This a short summary of these guidelines for responsible disclosure. In case of doubts, please contact the teacher.

- Please email your findings as soon as possible to `abuse@leidenuniv.nl`. Send your findings encrypted with our PGP key (fingerprint `E69D 56CB FAB3 E0F8 EE00 8623 7E1E ECF9 436E 6A5B`). Do this before you hand in and finish your report
- Do not abuse the found vulnerability; this includes:
  - downloading more data than necessary for showing the vulnerability
  - changing or removing data
- Be extra cautious when encountering personal data and don't cause a data breach
- Do not share the vulnerability with others until it is resolved. Await approval from the ISSC before including it in your report.
- Certain types of vulnerability may **not** be tested:
  - Do not test the physical security of the university
  - Do not test third-party applications (those that don't have an `.leidenuniv.nl` or `.universiteitleiden.nl` domain)
  - No social engineering
  - No DDoS
  - No malware distribution
  - No credential phishing or spam.
- Describe the issue found as explicitly and in detail as possible, and provide any evidence you might have.
- Do provide sufficient information to reproduce the problem so that we can resolve it as quickly as possible. Usually the IP address or the URL of the affected system and a description of the vulnerability is sufficient, but complex vulnerabilities may require further explanation.

Task 1: Explore vulnerabilities with ChatGPT

The first part of the assignment is about exploring potential vulnerabilities in services or hosts of the University. Of course, you can choose these vulnerabilities with any method, but one suggestion is finding some good ones in conversation with ChatGPT.

You will need to discuss at least **three** different vulnerabilities in your report and we want most of your information to come from ChatGPT or be sourced from ChatGPT (for example, by asking it for a good source to read up on it). The reason for this is that it may be how many novice adversaries may get information about vulnerabilities in the future. Besides the red teaming role play reason, it will also teach the ISSC how to use this tool for the same purpose.

Extensively discuss these vulnerabilities, what causes them and how they can be exploited. Include an example of an exploit, as offered by ChatGPT.

Task 2: Attempt to exploit vulnerabilities

After you have looked into vulnerabilities and exploits for them, it is time to see if they work. For this, it is allowed to test these exploits on actual University systems under the guise of responsible disclosure. Keep the guidelines in mind when you do this.

Try each of the three vulnerabilities on at least one service or system. Keep a good record of everything you have done in the attempt to exploit it. This part of the assignment can also be successful if you don't manage to find a true vulnerability. Reporting to what type of exploit a service is secure, is also a good result.

When you do find a way to exploit a vulnerability, report this to the `abuse@leidenuniv.nl` mailbox **immediately**. Don't include this in your report until you get permission from the ISSC, which will come after it has been resolved. If the deadline for the report comes before the vulnerability has been resolved, the ISSC will provide you with a document stating this, which can be included instead.

Report guidelines

The report should be submitted via Brigthspace (assignment *Read Teaming with ChatGPT*). It should be in English and in pdf. It is targeted at the Security Operations department of the Leiden University ICT Shared Services Centre, so make sure it looks professional and it is clear, easy to read and concise. Explain your method of using ChatGPT in this report and how you asked it about exploits.

Out of scope

Some vulnerabilities are rather trivial and others cannot be abused. The following are examples of known and accepted vulnerabilities and risks that are outside the scope of the assignment and responsible disclosure policy:

- `HTTP` 404 codes/pages or other `HTTP` non-200 codes/pages and Content Spoofing/Text Injection on these pages.
- fingerprint version banner disclosure on common/public services.
- disclosure of known public files or directories or non-sensitive information, (e.g. `robots.txt`).
- clickjacking and issues only exploitable through clickjacking.
- lack of `Secure/HTTPOnly` flags on non-sensitive Cookies.
- `OPTIONS HTTP` method enabled.
- anything related to `HTTP` security headers, e.g.:
  - `Strict-Transport-Security`
  - `X-Frame-Options`
  - `X-XSS-Protection`
  - `X-Content-Type-Options`
  - `Content-Security-Policy`
- `SSL` Configuration Issues:
  - `SSL` forward secrecy not enabled.
  - Weak / insecure cipher suites.
- `SPF, DKIM, DMARC` issues.
- host header injection.
- reporting older versions of any software without proof of concept or working exploit.
- information leakage in metadata.

Grading

Any discovered vulnerability that the ISSC assesses with high severity will give you 10 points to be used to substitute one of the assignment grades. Medium and low severity vulnerabilities will be assessed similarly (but for a somewhat smaller number of points; in discussion with the ISSC).

No discovered vulnerabilities: We will only evaluate your report, assessing how much effort you have put into interacting with ChatGPT, learning from it, interacting with the University systems, writing up the report, and the lessons that the ISSC can learn from your report. This will result in bonus points (**up to** 2.0) that will be added to the total assignment grade. So if your average assignment grade is 7.8 and you get 2.0 bonus points, your total assignment grade for the course will be 9.8. Alternatively, you can interpret this bonus as (**at most**) 1.0 being added to your final course grade.

These options can be combined (i.e., if you have discovered a vulnerability and have submitted a great report, we will give you both a high grade to substitute one of the assignment grades `and` up to 2.0 bonus points to add to the total assignment grade).