Generative Artificial Intelligence (AI) in Knowledge Discovery and Generation for Cyber Defense

Ping Wang, PhD, CISSP
University Professor
NCAE Coordinator & POC
Robert Morris University
wangp@rmu.edu

2025 CAE Community Symposium

Overview

2025 CAE Community Symposium

- **Focus**
  Impacts of Gen AI on cyber defense (CD) knowledge discovery
- **Significance**
  - Knowledge is power
  - CD challenge: Lack of knowledge of vulnerabilities/threats
  - Pentesting challenge: Exponential growth of data/complexity
  - Gen AI potential: Improve automation, efficiency, interactivity
- **Goal**
  Contribute an AI-moderated knowledge discovery model and new empirical data using GPT/LLM AI for cyber defense

Research Background

2025 CAE Community Symposium

- ❑ Power of knowledge: Military defense vs. cyber defense
  - ➢ Knowledge of strengths/vulnerabilities of self/opponent (*The Art of War)*
  - ➢ Pentesting for knowledge discovery in cyber defense
- ❑ AI-assisted Pentesting
  - ✓ Early threat detection and response
  - ✓ Automation, efficiency and accuracy in vulnerability and risk analysis
  - ✓ Help human pentesters to predict profiles and make recommendations
- ❑ Risks & Limitations
  - ➢ Malicious misuse for more powerful & automated attacks
  - ➢ Potential disclosures of private/sensitive/copyrighted information
  - ➢ Hallucinations with misleading misinformation
  - ➢ Replace human intelligence/creativity in defense decision making?

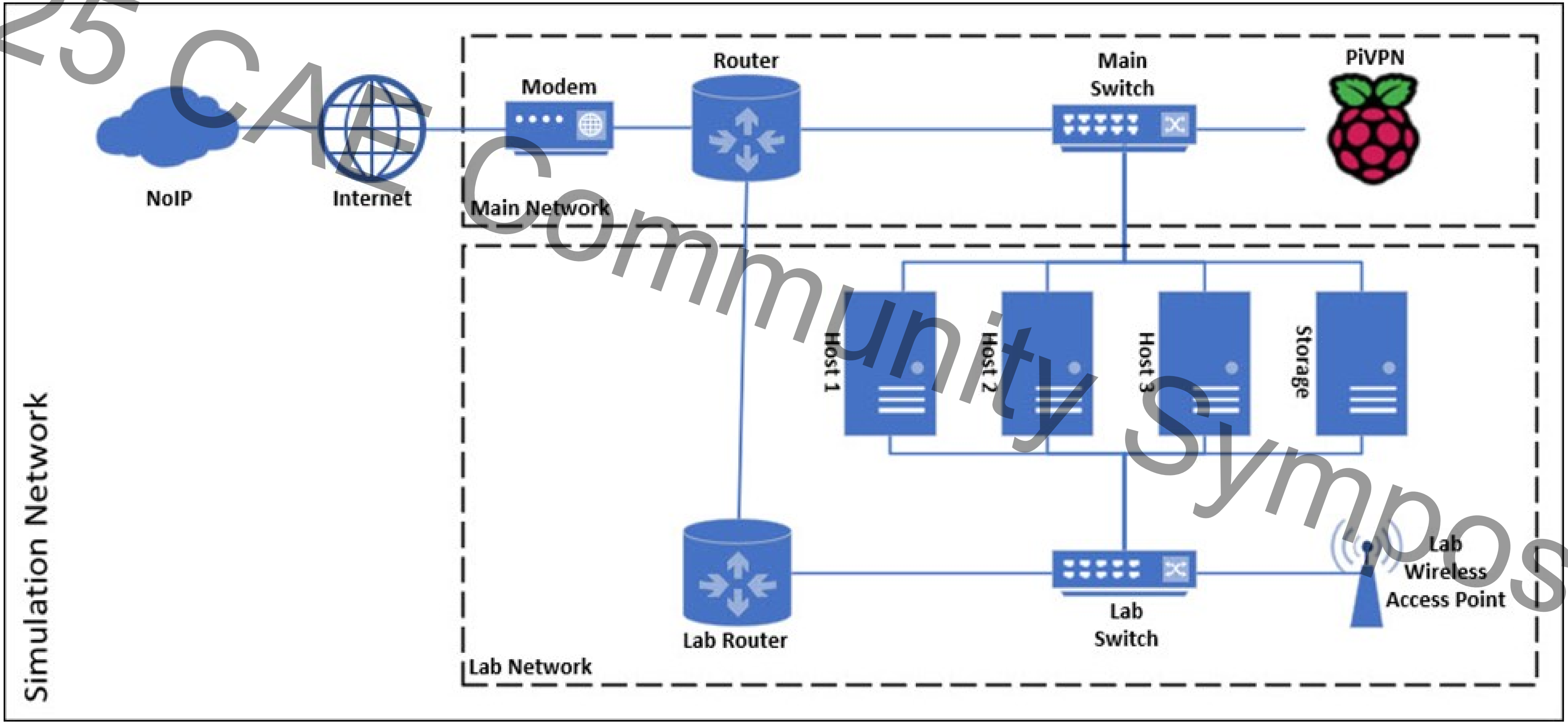
2025 CAE Community Symposium

AI-moderated Knowledge Discovery Model

	Knowledge	Goals	Moderating Effects of AI
Oneself	• Know one's own vulnerabilities • Know how to mitigate one's own vulnerabilities • Know how to hide one's assets and vulnerabilities from the opponent • Know how to set up deceptive vulnerabilities	• To minimize one's vulnerabilities • To assess and manage one's vulnerabilities and risks • To minimize the opponent's knowledge of one's vulnerabilities • To mislead, misinform, distract, and deceive the opponent	• AI may enhance the effects of knowledge discovery • AI may limit the effects of knowledge discovery
Opponent	• Know the opponent's assets and strengths • Know the opponent's vulnerabilities • Know how to discover the opponent's vulnerabilities	• To be aware of threats and avoid striking the strong spots of the opponent • To exploit the vulnerabilities of the opponent • To maximize knowledge of the opponent	• AI may enhance the effects of knowledge discovery • AI may limit the effects of knowledge discovery

Simulation: Lab Diagram

2025 CAE Community Symposium



Simulation: VMs & Tools

VM/Host	IP	OS	Tools
100	192.168.1.204	Windows 11 Pro 23H2	WinSCP, Putty
102	192.168.1.201	Ubuntu Desktop 22.04.4 LTS	ChatGPT Desktop Client
103	192.168.1.203	Kali Linux 2024.1	Metasploitable Framework/Wireshark/Nmap
107	192.168.1.209	Ubuntu Desktop 22.04.4 LTS B	PentBox
300	192.168.1.210	Metasploitable2 VM	Metasploitable 2

2025 CAE Community Symposium

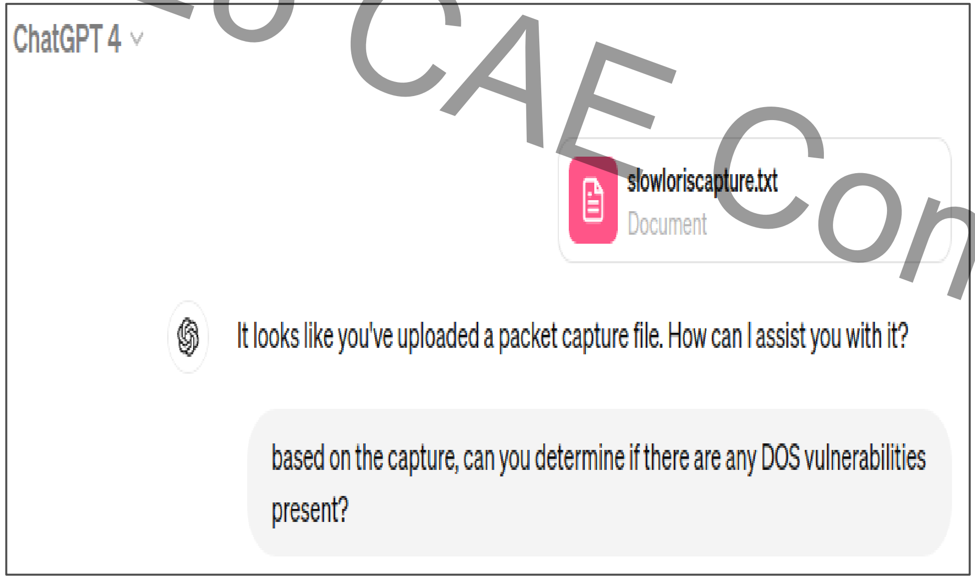
Simulation: Vulnerabilities for Exploitations

2025 CAE Community Symposium

Vulnerability	ID	Packet Capture File
FTP Backdoor	CVE-2011-2523	vsftpdcapture.txt
Slowloris DOS attack	CVE-2007-6750	slowloriscapture.txt
Samba User Name Script	CVE-2007-2447	sambausernamemapscript capture1.txt
Brute Force on VNC	CVE-1999-0502	vncloginscannercapture.txt
Honeypot Pentest	N/A	honeypotcapture.txt

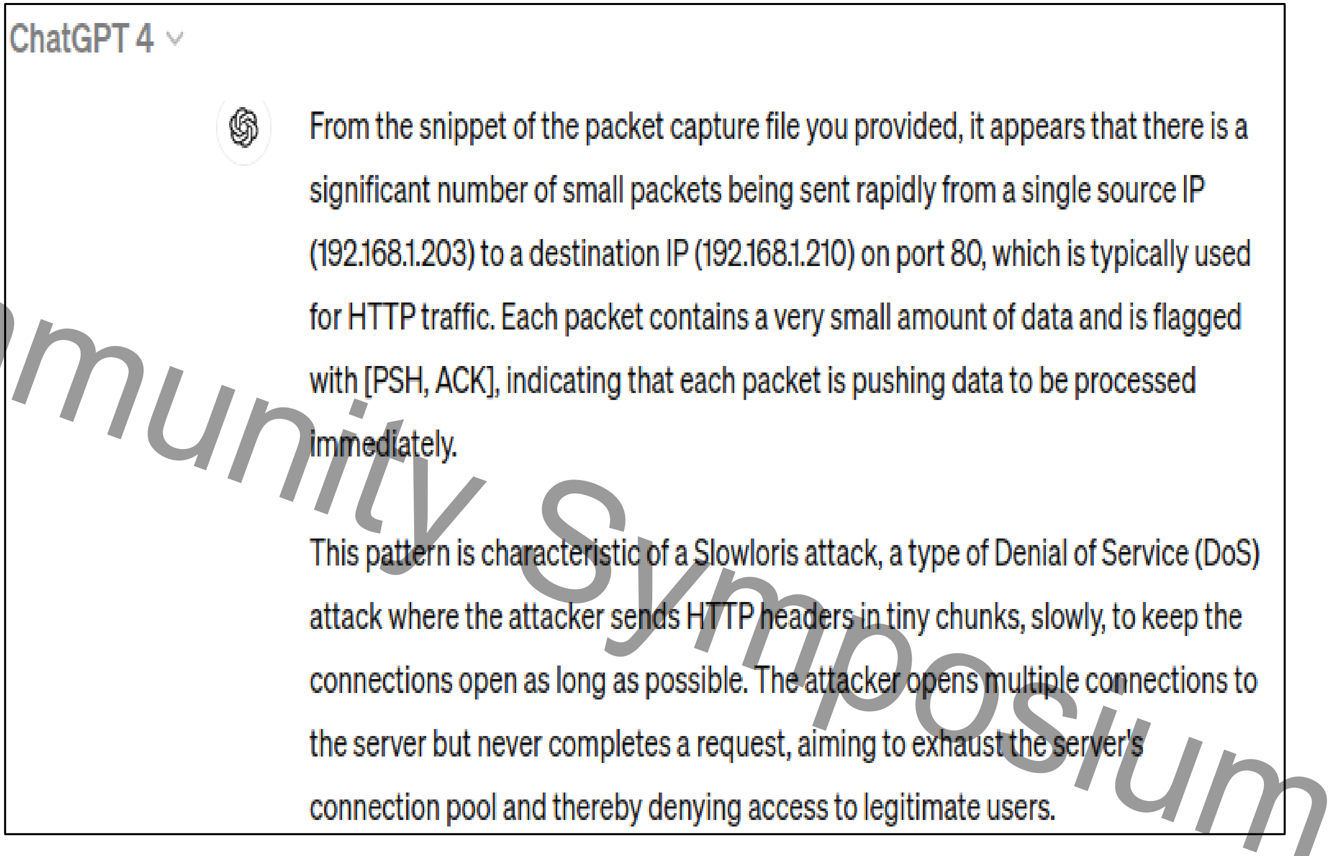
Sample ChatGPT-4 Prompt & Answer

2025 CAE Community Symposium



It looks like you've uploaded a packet capture file. How can I assist you with it?

based on the capture, can you determine if there are any DOS vulnerabilities present?



From the snippet of the packet capture file you provided, it appears that there is a significant number of small packets being sent rapidly from a single source IP (192.168.1.203) to a destination IP (192.168.1.210) on port 80, which is typically used for HTTP traffic. Each packet contains a very small amount of data and is flagged with [PSH, ACK], indicating that each packet is pushing data to be processed immediately.

This pattern is characteristic of a Slowloris attack, a type of Denial of Service (DoS) attack where the attacker sends HTTP headers in tiny chunks, slowly, to keep the connections open as long as possible. The attacker opens multiple connections to the server but never completes a request, aiming to exhaust the server's connection pool and thereby denying access to legitimate users.

Summary Findings & Knowledge

2025 CAE Community Symposium

Vulnerability	Key Indicators	Mitigation Strategies
ftp-vsftpd-backdoor	● Unexpected or suspicions outbound connections ● Non-standard port usage ● Unusual IP addresses or domains in network traffic	● Ensure services are up-to-date ● Use strong, non-default credentials ● Secure critical protocols with secure encryption
Slowloris DOS attack	● Repeated small payload sizes ● Rapid succession of packets to the same destination on a web service port	● Implement rate limiting ● Implement connection timeouts ● Use reverse proxies, or Specialized firewall rules
Samba User Name Script	● Outdated Samba versions ● Increased SMB traffic, especially involving unusual ports or high volumes of data transfer ● Misconfigurations, such as weak passwords and lack of encryption	● Check versions of Samba and compare against known vulnerabilities listed in databases like CVE ● Review system and security logs for signs of unauthorized access or suspicious activities
Brute Force on VNC	● Repeated failed login attempts from the same source ● Different credentials being used from the same IP address	● Password policies requiring complex and lengthy passwords ● Implement account lockout policies after a certain number of failed login attempts

Discussions & Conclusions

2025 CAE Community Symposium

- AI Help
  - ✓Efficiency with pentesting data
  - ✓Highlight indicators of vulnerabilities
  - ✓Offer mitigation suggestions and strategies
- Limitations
  - o Answers limited to existing data; creativity?
  - o Lack of details on the use of Honeypot
  - o Limits in processing large data files
- Further research
- Questions/suggestions
- Thank you!

References

2025 CAE Community Symposium

Wang, P., & D'Cruze, H. (2024). AI-Assisted pentesting using ChatGPT-4. In Latifi, S. (eds) *ITNG 2024. Advances in Intelligent Systems and Computing, vol 1456 (pp.63-71)*. Springer, Cham.

Wang, P, & Johnson, C. (2024). The impacts of generative artificial intelligence (AI) in knowledge discovery and generation for cyber defense. *Issues in Information Systems, 25*(2), 230-243.

Wilson, R. (2018). Sun Tzu and the art of cyberwar. *Defense AT&L, January-February 2018,* 30-34.