12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

EVALUATING THE EFFECTIVENESS OF THE LOCALLY DEPLOYED GPT-OSS MODEL IN DETECTING SMB AND RDP BRUTE-FORCE ATTACKS FROM WINDOWS SECURITY LOGS

IGOR PETROVIĆ
Ministry of Defence, Belgrade, igor.petrovic@mod.gov.rs

MILOMIR JOVIĆEVIĆ
Ministry of Defence, Belgrade, milomir.jovicevic@mod.gov.rs

MARKO PREMOVIĆ
Serbian Armed Forces, Belgrade, marko.premovic@vs.rs

MLADEN VEINOVIĆ
Singidunum University, Belgrade, mveinovic@singidunum.ac.rs

IVAN TOT
University of Defence, Belgrade, ivan.tot@mod.gov.rs

Abstract: *Brute-force attacks targeting Server Message Block (SMB) and Remote Desktop Protocol (RDP) services remain among the most common techniques used to gain unauthorized access to Windows systems. Traditional detection approaches are typically based on manually defined rules and event correlation mechanisms, which may be difficult to maintain and adapt to evolving attack patterns. Recent advances in Large Language Models (LLMs) have demonstrated promising capabilities in log analysis and anomaly detection; however, their applicability to authentication-based attack detection in Windows environments remains insufficiently explored. This paper evaluates the effectiveness of a locally deployed GPT-OSS model in detecting SMB and RDP brute-force attacks from Windows Security logs. A controlled experimental environment was established consisting of a Windows host with SMB and RDP services enabled, a Kali Linux attack host executing brute-force attacks using Hydra, and an NXLog-based log collection pipeline. Authentication-related security events were preprocessed and analyzed using a context-aware approach in which the model received a sliding window of recent log entries and classified the latest event as benign or malicious. The experimental results demonstrate that the proposed approach achieves high detection performance, obtaining F1 scores of 0.999 for SMB and RDP brute-force attacks in contextual judgement. The findings indicate that combining targeted preprocessing of authentication logs with context-aware reasoning performed by a Mixture-of-Experts (MoE) LLM enables accurate identification of attack patterns without relying on handcrafted detection rules. The main contributions of this work include the development of a reproducible framework for LLM-based analysis of Windows Security logs, the creation of a labeled dataset of SMB and RDP brute-force events, and the experimental validation of GPT-OSS as an effective locally deployable solution for real-time security monitoring.*

Keywords: *RDP, SMB, brute-force attacks, GPT-OSS, LLM log analysis.*

1. INTRODUCTION

Remote administration services such as Server Message Block and Remote Desktop Protocol are widely used in enterprise and personal computing environments. The massive use of these services has opened up many opportunities for launching attacks, among which brute force attacks are very common way for gaining unauthorized access to Windows systems [1]. By repeatedly attempting different username and password combinations, attackers seek to compromise valid accounts and establish an initial foothold within a target network.

Traditional detecting approaches are primarily based on predefined rules, signatures, or threshold-based correlation mechanisms implemented within Security Information and Event Management platforms, but modern use of technology demands novel approach and automatization of detection process. In addition, the increasing volume and complexity of security logs generated by modern information systems create important challenges for security analysts responsible for monitoring and incident response. Recent advances in Large Language Models (LLMs) have demonstrated promising capabilities in understanding, analyzing, and reasoning over semi-structured and unstructured textual data [2]. Unlike conventional machine learning approaches that typically require extensive feature engineering and model training, LLMs can leverage

Corresponding author: I. Petrović;
E-mail: igor.petrovic@mod.gov.rs
DOI: 10.5937/OTEH26092P

contextual information contained within log sequences and provide interpretable explanations for their decisions. Despite these developments, the effectiveness locally deployable LLMs for detecting authentication-based attacks from Windows Security logs remains insufficiently explored, especially when it comes to mixture of experts models that more rationally use the hardware resources of the servers on which they are run.

This paper presents a proof-of-concept framework for detecting SMB and RDP brute-force attacks using a locally deployed GPT-OSS model. The experimental environment consists of a Windows host with SMB and RDP services enabled, a Kali Linux attack machine executing brute-force attacks using Hydra, and an NXLog-based log collection pipeline. Windows Security events are preprocessed and provided to the GPT-OSS model in the form of a sliding context window containing the most recent log entries. The model is then tasked with determining whether the latest event corresponds to benign activity or to an ongoing brute-force attack.

The main contributions of this work are:

(1) the development of a practical framework for LLM-based real-time analysis of Windows Security logs,

(2) the experimental evaluation of a locally deployed GPT-OSS model in detecting SMB and RDP brute-force attacks,

(3) novel dataset of attack-driven and processed log messages, generated through controlled execution of brute-force attack scenarios. and

(4) an assessment of the feasibility of employing LLMs as components of future security monitoring and threat detection systems.

2. RELATED WORKS

Brute-force attacks targeting Windows authentication services remain one of the most prevalent techniques used to gain unauthorized access to enterprise systems. In particular, SMB and RDP services are frequently exposed within organizational networks and are therefore common targets for password-guessing attacks. RDP has become one of the primary attack vectors used by adversaries to obtain initial access [4], while repeated authentication attempts against SMB services continue to represent a significant security concern in Windows environments. The large volume of authentication-related events generated during such attacks makes timely detection difficult and motivates the development of more intelligent log analysis approaches.

Recent advances in Large Language Models (LLMs) have opened new possibilities for automated security log analysis. LogPrécis demonstrated the applicability of LLMs in cybersecurity by analyzing Unix shell attacks captured in honeypot environments and generating concise yet informative interpretations of security events [5]. Similarly, Bokkena investigated the use of LLMs for predictive threat detection in IT security and reported experimental results exceeding 95% accuracy, outperforming several conventional machine learning approaches while simultaneously reducing false positive detections [6]. The growing success of LLMs in complex reasoning tasks has also encouraged their adoption for higher-level behavioral analysis, where contextual understanding is essential for identifying malicious activities [3].

The potential of LLMs for anomaly detection and semantic interpretation of system behavior has been further demonstrated by different frameworks [7,8]. Furthermore, the survey presented in [9] highlights the increasing importance of LLM-based anomaly detection and Retrieval-Augmented Generation techniques for cybersecurity applications, while also identifying a lack of research focused on real-time log analysis scenarios.

For locally deployed LLM solutions, Mixture-of-Experts (MoE) architectures have shown particular promise due to their ability to utilize hardware resources more efficiently while maintaining strong inference performance. Among these models, the locally deployed GPT-OSS model has demonstrated favorable results on systems equipped with 32 GB of VRAM, making it a suitable candidate for real-time cybersecurity applications [2]. For this reason, GPT-OSS was selected as the foundation of the detection framework evaluated in this work.

3. EXPERIMENTAL SETUP

The experimental environment was designed to evaluate the capability of a locally deployed GPT-OSS model to detect SMB and RDP brute-force attacks from Windows Security logs. The overall architecture enabled the generation of authentication-related security events, collection and preprocessing of Windows logs, and automated classification of security events using the GPT-OSS model.

3.1. Test Environment

Figure 1 illustrates the architecture of the proposed experimental framework.

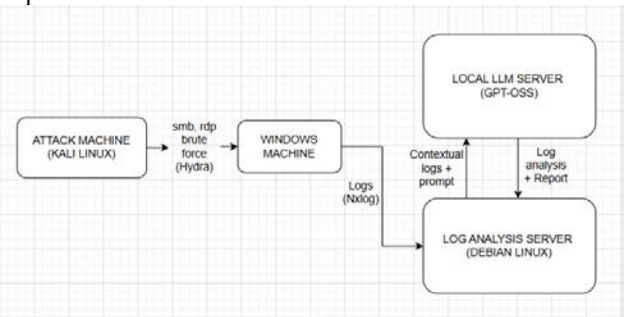


Figure 1. Experimental architecture

Figure 1 shows the architecture of the implemented framework. In order to achieve objectivity in evaluating the performance of the locally deployed GPT-OSS model, an experimental data set was created through SMB and RDP brute force attacks directed at a Windows machine. Log messages were collected and processed using the NXLOG event collection system on Windows operating

systems. The collected log messages are taken over by a specially designed application for real-time log analysis installed on a Debian Linux server and forwarded to GPT-OSS, which assigns a "malicious" or "benign" rating to the latest log message in correlation with a set of 50 previous log messages. Table 1 summarizes the main components of the experimental environment.

Table 1. Experimental environment components

Component	Description
Attacker Host	Kali Linux
Attack Tool	Hydra
Target Host	Windows system with SMB and RDP enabled
Log Collection	NXLog
AI Analysis Host	Local GPT-OSS deployment
Monitored Services	SMB and RDP

3.2. Attack Scenarios

In order to confirm the effectiveness of the locally deployed LLM GPT-OSS in assessing the maliciousness of log messages received from a Windows machine, it is necessary to establish a system in which it would be precisely known whether a log message is malicious or not. For this reason, 2 brute force attack scenarios were created that are directed at services opened on a Windows machine (SMB and RDP) with pre-prepared wordlists containing the real credentials of the machine. By launching the attack, every log message related to activity directed at the SMB and RDP services would be malicious in nature, which allows an objective assessment of the effectiveness of the observed framework. The attacks were implemented using the Hydra tool on a Kali Linux attack machine, and the attack flow is demonstrated in Figure 2.

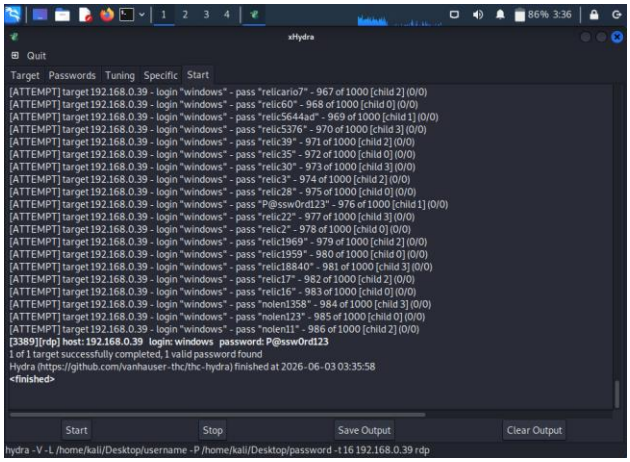


Figure 2. Generating log dataset by RDP brute-force attack using Hydra tool

3.3. Log Collection and Context Generation

Windows Security logs generated during the experiment were collected using NXLog and transformed into a structured representation containing the most relevant authentication-related attributes. Examples of extracted fields included Event ID, timestamp, username, source IP address, logon type, and authentication status.

Due to the simplicity of processing, the NXlog configuration allows for the collection of only log messages related to the SMB and RDP services, which are processed in such a way that each event is recorded as a single line in the log message, as presented:

2026-06-01 23:46:55 DESKTOP-8KVG4F0 An account failed to log on. Subject: Security ID: S-1-0-0 Account Name: - Account Domain: - Logon ID: 0x0 Logon Type: 3 Account For Which Logon Failed: Security ID: S-1-0-0 Account Name: admin Account Domain: Failure Information: Failure Reason: Unknown user name or bad password. Status: 0xC000006D Sub Status: 0xC0000064 Process Information: Caller Process ID: 0x0 Caller Process Name: - Network Information: Workstation Name: \\192.168.0.46 Source Network Address: 192.168.0.46 Source Port: 63342 Detailed Authentication Information: Logon Process: NtLmSsp Authentication Package: NTLM Transited Services: - Package Name (NTLM only): - Key Length: 0

Instead of analyzing individual events in isolation, the proposed framework used a sliding context window approach, which has proven effective in previous research using the same locally deployed model [2]. For each classification request, the GPT-OSS model received the most recent 50 log entries, including the event under investigation. This design was intended to provide temporal and contextual information that could help the model recognize repetitive authentication failures characteristic of brute-force attacks. The GPT-OSS model was instructed to analyze the supplied log sequence and determine whether the latest event represented normal activity, an SMB brute-force attack, or an RDP brute-force attack. The generated classification was subsequently compared with the ground-truth label associated with the experimental scenario.

4. GPT-OSS-BASED DETECTION APPROACH

The proposed detection framework performs real-time analysis of Windows Security logs using a locally deployed GPT-OSS model. After collection and preprocessing by NXLog, each log event is forwarded to an analysis pipeline that evaluates both the individual event and its surrounding context. To provide temporal awareness, a sliding window containing the most recent security events is maintained in memory and continuously updated as new logs arrive. For every newly received log entry, system evaluates the same log together with a context window consisting of previously observed events. This approach enables the model to identify behavioral patterns that are characteristic of brute-force attacks, such

as repeated authentication failures originating from the same source host or targeting the same user account.

The GPT-OSS model receives the log data through a structured prompt that instructs it to classify the observed activity as either benign or malicious. When malicious behavior is detected, the model is additionally required to identify the attack category and provide a short explanation supporting its decision. The generated output is subsequently stored together with the original log entry and metadata for further evaluation.

To support future analysis and reproducibility, all processed events and corresponding model decisions are archived in a vector database and exported to a structured dataset. This design enables direct comparison between classifications obtained from single-event analysis and those produced using contextual information, allowing the effectiveness of context-aware LLM-based detection to be systematically evaluated.

5. RESULTS AND DISCUSSION

As the technical feasibility of the system was presented in the previous chapter, it was necessary to confirm the efficiency of GPT-OSS in log analysis. The metric used to evaluate model performance, as well as statistical analysis, is the F1 score. The results important for calculating the F1 score (for RDP and SMB brute-force attacks) are given in Table 2.

Table 2. Confusion matrix results for single-log and contextual log analysis of logs

Attack type	GPT-OSS 20B single log judgement				GPT-OSS 20B contextual log judgement			
	TP	FP	TN	FN	TP	FP	TN	FN
RDP brute-force	975	0	0	17	991	0	0	1
SMB brute-force	761	0	0	217	976	0	0	2

The presented results represent the evaluation of the GPT-OSS 20B model listed in the table in classifying log messages as malicious or benign for two specific attack scenarios - RDP and SMB brute force attacks on Windows 10 machine. All log messages generated as a system response to attacks were considered malicious. By using parameters given in Table 2, Precision, Recall and F1 score are calculated and presented in Table 3.

Table 3. Confusion matrix results for single-log and contextual log analysis of logs

Attack type	GPT-OSS 20B single log judgement			GPT-OSS 20B contextual log judgement		
	Precision	Recall	F1	Precision	Recall	F1
RDP brute-force	1	0.983	0.991	1	0.999	0.999
SMB brute-force	1	0.778	0.875	1	0.998	0.999

Given results show that GPT-OSS 20B achieves very high accuracy in detecting both RDP and SMB brute-force attacks, with very high F1 score in all experiments, indicating a complete absence of false positive detections. However, analysis of individual logs shows a significantly lower recall, especially for SMB brute-force attacks (0.778), suggesting that the model misses a large number of malicious events when it lacks additional context. By introducing context from 50 previous logs, the performance improves drastically, with Recall reaching 0.999 for RDP and 0.998 for SMB brute-force attacks. Consequently, the F1 measure increases from 0.991 to 0.999 for RDP and from 0.875 to 0.999 for SMB, confirming that contextual information plays a key role in correctly recognizing attack patterns. The results indicate that such a system has significant potential for application as an additional layer of protection within security operations centers and incident detection and response systems. It is particularly suitable for continuous analysis of large volumes of logs and early detection of attack patterns through event correlation, which can improve the overall effectiveness of cyber protection and reduce the likelihood of missing malicious activities.

5. CONCLUSION

This paper presented an approach for applying a locally deployed GPT-OSS large language model to the detection of SMB and RDP brute-force attacks using Windows Security logs. The motivation for this research originates from the continued prevalence of authentication-based attacks against critical Windows services and the limitations of traditional rule-based detection mechanisms when dealing with large volumes of security events. To address these challenges, a proof-of-concept framework was developed that combines Windows log collection, NXLog-based preprocessing, context-aware event correlation, and LLM-driven classification.

The experimental evaluation was conducted in a controlled environment consisting of a Windows target host with SMB and RDP services enabled and a Kali Linux attack host executing brute-force attacks using Hydra. A dedicated dataset of Windows Security events was generated and labeled based on the performed attack scenarios. To improve the relevance of the information provided to the model, the preprocessing stage focused on authentication-related events and attributes associated with the monitored services, reducing noise while preserving security-critical context.

The obtained results demonstrate that the GPT-OSS model is highly effective in identifying brute-force activity from Windows Security logs. The model achieved an F1 score of 0.999 for RDP brute-force detection and SMB brute-force detection, indicating excellent classification performance across both attack scenarios in contextual judgement. These results confirm that a context-aware analysis approach enables the model to recognize repetitive authentication patterns and distinguish malicious activity from legitimate user behavior with a high degree of accuracy.

An additional contribution of this work is the demonstration that Mixture-of-Experts (MoE) architectures can be effectively deployed on commodity server hardware equipped with 32 GB of VRAM while

still delivering high detection performance. The GPT-OSS model was selected based on its previously demonstrated efficiency under these hardware constraints, and the results obtained in this study further validate its suitability for real-time cybersecurity applications. The proposed framework demonstrates that locally deployed open-weight language models can serve as practical components of modern security monitoring systems without requiring extensive feature engineering or manually crafted detection rules.

References

[1] CYBERSECURITY AND INFRASTRUCTURE SECURITY AGENCY: Weak Security Controls and Practices Routinely Exploited for Initial Access, CISA, 2022.

[2] PETROVIĆ, I., VEINOVIĆ, M., ILIĆ, S., JOVIĆEVIĆ, M.: *Framework for Semantic Threat Detection in Docker Container Environments with Local MoE LLMs*, Electronics, 15(8):1664 (2026), DOI: https://doi.org/10.3390/electronics15081664

[3] CUI, M., et al.: AGLHunter: Automated Threat Hunting Using In-Context Learning-Enhanced LLM, 2025 28th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Compiegne, France, 2025, pp. 1812–1819, DOI: https://doi.org/10.1109/CSCWD64889.2025.11033364.

[4] VITLA, S.: *Unsecured Remote Desktop Protocol (RDP) Access: A Gateway for Ransomware Attacks and Corporate Extortion*, Journal of Computer Science and Technology Studies, 6(2) (2024) 150–165. DOI: 10.32996/jcsts.2024.6.2.17.

[5] BOFFA, M., DRAGO, I., MELLIA, M., VASSIO, L., GIORDANO, D., VALENTIM, R., BEN HOUIDI, Z.: LogPrécis: Unleashing Language Models for Automated Malicious Log Analysis: Précis: A Concise Summary of Essential Points, Statements, or Facts, Computers & Security, 141 (2024) 103805, DOI: https://doi.org/10.1016/j.cose.2024.103805

[6] BOKKENA, B.: *Enhancing IT Security with LLM-Powered Predictive Threat Intelligence*, Proceedings of the 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2024, 751–756, DOI: 10.1109/ICOSEC61587.2024.10722712

[7] ZHANG, H., SHAO, S., LI, S., ZHONG, Z.: SmartGuard: Leveraging Large Language Models for Network Attack Detection Through Audit Log Analysis and Summarization, arXiv:2506.16981, 2025.

[8] LI, C., ZHU, Z., ZHANG, X.: RedChronos: A Large Language Model-Based Log Analysis System for Insider Threat Detection in Enterprises, arXiv:2503.02702v2, 2025.

[9] AKHTAT, S., KHAN, S., PARKINSON, S.: *LLM-Based Event Log Analysis Techniques: A Survey*, arXiv:2502.00677v1, 2025.