**State of the Art**

The creation of AI and machine learning tools has been essential in identifying zero-day threats in the face of hostile actors' constantly developing evasion tactics. Organizations frequently face difficulties implementing such tools due to their cost and complexity. This is where YARA's rule set comes in handy, providing a cost-efficient and efficient method of detecting fresh malware without only depending on AI-driven solutions. To identify encrypted payloads, encrypted JavaScript, or encrypted files inside PDFs, ARA can be used to develop rules. This proactive approach ensures that encrypted malware is detected and thwarted, safeguarding your system from potential risks. A 2022 project by Reversing Labs (Labs, 2022) project gives a demonstration to utilize YARA to detect the 2021 GwisinLocker zero-day encryption malware, which was targeted at pharmaceutical companies in South Korea. Reversing Labs in 2023, have an effective YARA project to detect Lorenz ransomware which surfaced in 2021. Several months after its inception, the Lorenz operation had found a new attack avenue to exploit. Based on research from Arctic Wolf Labs (Wolf, 2022), the operation used a critical vulnerability in Mitel MiVoice VOIP appliances to breach enterprises. The vulnerability, [CVE-2022-29499](https://www.mitel.com/en-ca/support/security-advisories/mitel-product-security-advisory-22-0002), is a remote code execution vulnerability that allows the threat actor to obtain a reverse shell, followed by using Chisel as a tunneling tool to pivot into the victim’s environment.

YARA can be leveraged to create rules that identify encrypted payloads, encrypted JavaScript, or encrypted files within PDFs. This proactive approach ensures that encrypted malware is detected and thwarted.

In their 2022 demonstration (Abrams, 2022), Abrams breaks down the Windows MoTW bypass zero-day flaw which was first reported by Hewlett-Packard. The zero-day vulnerability leverages a flaw in the Windows Shell, enabling threat actors to execute arbitrary JavaScript code without raising the usual security prompts or warnings. By exploiting this vulnerability, attackers can evade detection mechanisms and potentially compromise vulnerable systems. This loophole poses a significant risk as it bypasses the built-in security measures designed to protect users from executing potentially harmful scripts. YARA can effectively identify patterns which include specific function names, variable assignments e.g *wget* or *curl.* A rule can focus on specific functionality triggers that are commonly associated with the exploit of similar nature hence covering all related attacks based on a behavior match.

Despite the critical role that AI and machine learning tools play in identifying and reducing dangers, many organizations may find it difficult to afford and use it effectively. With the help of the YARA, enterprises may identify zero-day risks without having to exclusively rely on expensive AI-driven technologies. Large amounts of data can be handled effectively with YARA. It can quickly process files, making it suited for batch mode or real-time scanning of many PDF files. Additionally, the scanning process can be fine-tuned with its programmable rules, which minimizes false positives. To ensure precise detection, you can modify the criteria to match traits of malicious patterns.

Finally, YARA benefits from a collaborative community of security professionals who continuously contribute to its rule set. When a new exploit or vulnerability like the one described arises, security experts promptly develop and share YARA rules to detect and mitigate it. This collaborative approach ensures that organizations have access to the latest rules to defend against emerging threats.

ClamAV is an open source (ClamAV, 2023) (GPLv2) anti-virus toolkit, designed especially for e-mail scanning on mail gateways. It provides several utilities including a flexible and scalable multi-threaded daemon, a command line scanner and advanced tool for automatic database updates. The core of the package is an anti-virus engine available in a form of shared library. Real time protection (Linux only). The ClamOnAcc client for the ClamD scanning daemon provides on-access scanning on modern versions of Linux. This includes an optional capability to block file access until a file has been scanned on-access prevention. ClamAV's strength lies not only in its core capabilities but also in its timely signature updates. The ClamAV community, comprising security professionals and researchers worldwide, actively contributes to maintaining a vast database of malware signatures. This collaborative effort ensures that ClamAV users receive regular updates containing the latest detection patterns, effectively combating emerging threats, including obfuscated, JavaScript-based, and encrypted malware.

The goal of the project Sweeper is to create an effective and accurate pdf scanner and provide a functional user interface on which strong technical controls to further reduce attack vectors. Sweepers front end allows a file submission form which checks the validates the file type, size and most importantly sanitizes the file name to prevent injection which could lead to path traversal on the host server. The Sweeper web application uses the magic module (Daniel Mendler, 2023) to determine the MIME type of the file content. By checking that the MIME type is "application/pdf," it ensures that the uploaded file truly contains PDF data. The application also to extracts the file name without any directory components. This validation helps guard against potential attacks where an attacker may attempt to upload a file with a misleading extension and path traversal attacks respectively.

In conclusion, ClamAV is highly customizable and provides granular control offering good visibility to scanning operations. Its combination with YARA alone offers a strong, automated, and lightweight anti-virus protection tool that is perfectly suited for scanning PDF files in real-time on Windows and Linux servers.

# Bibliography

Abrams, L. (2022). *Exploited Windows zero-day lets JavaScript files bypass security warnings*. Retrieved from Bleeping Computer: https://www.bleepingcomputer.com/news/security/exploited-windows-zero-day-lets-javascript-files-bypass-security-warnings/

ClamAV. (2023). Retrieved from ClamAV: https://docs.clamav.net/Introduction.html

Daniel Mendler, J. W. (2023). Retrieved from https://github.com/mimemagicrb/mimemagic#

Labs, R. (2022, 11). Retrieved from From the Labs: YARA Rule for Detecting GwisinLocker: https://www.reversinglabs.com/from-the-labs/yara-rule-for-detecting-gwisinlocker

Wolf, A. (2022, 09). *Chiseling In: Lorenz Ransomware Group Cracks MiVoice And Calls Back For Free*. Retrieved from Chiseling In: Lorenz Ransomware Group Cracks MiVoice And Calls Back For Free: https://arcticwolf.com/resources/blog/lorenz-ransomware-chiseling-in/