

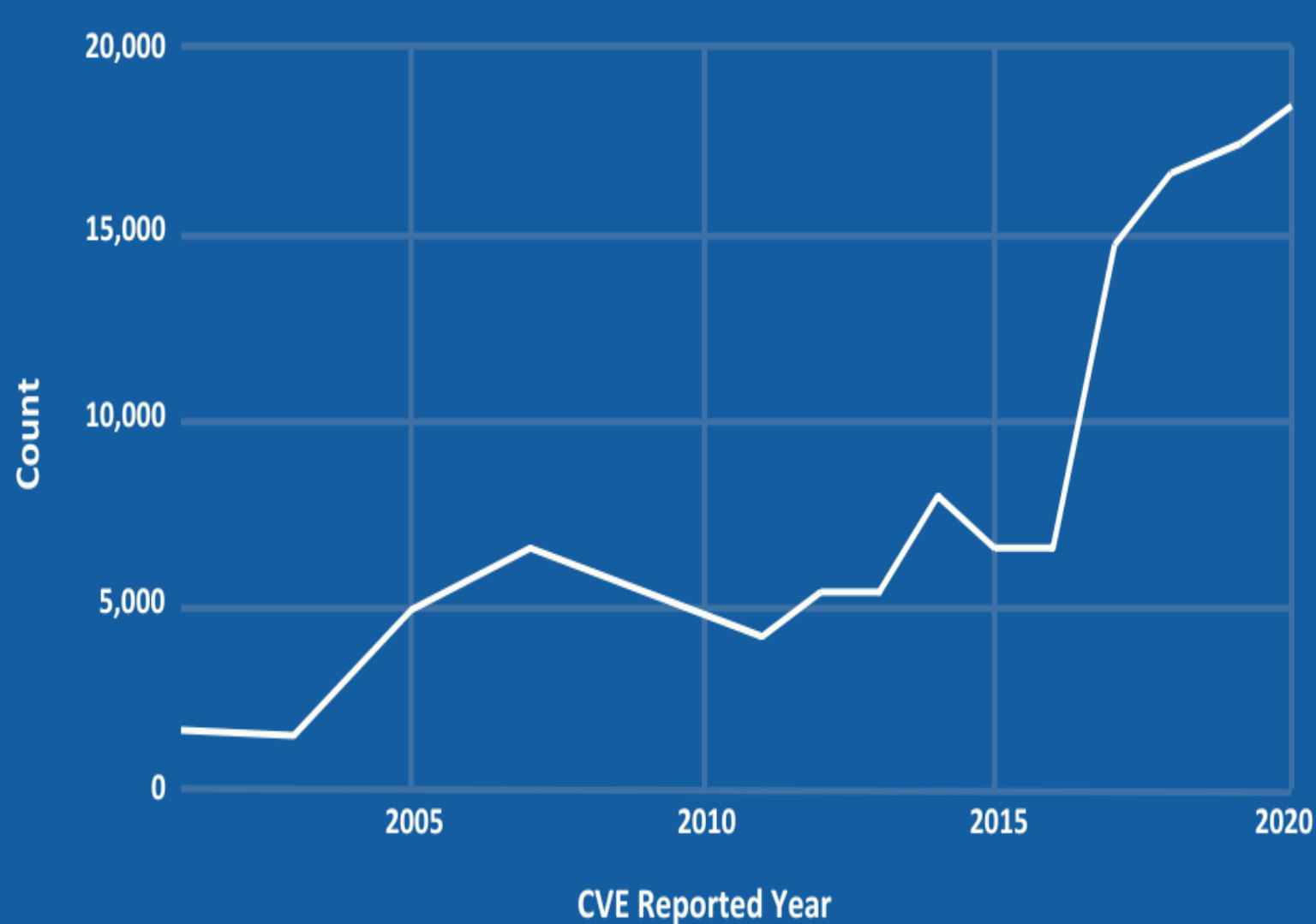
Introduction

- **Issue:** Cybersecurity reporting is slow and requires a lot of manual labor to sift through data to generate a report
- **Importance:** Timely and insightful threat reports are crucial for understanding and addressing cyber threats
- **Project Goal:** Explore using large language models to automate the generation of cyber threat reports

CVE's, what are they?

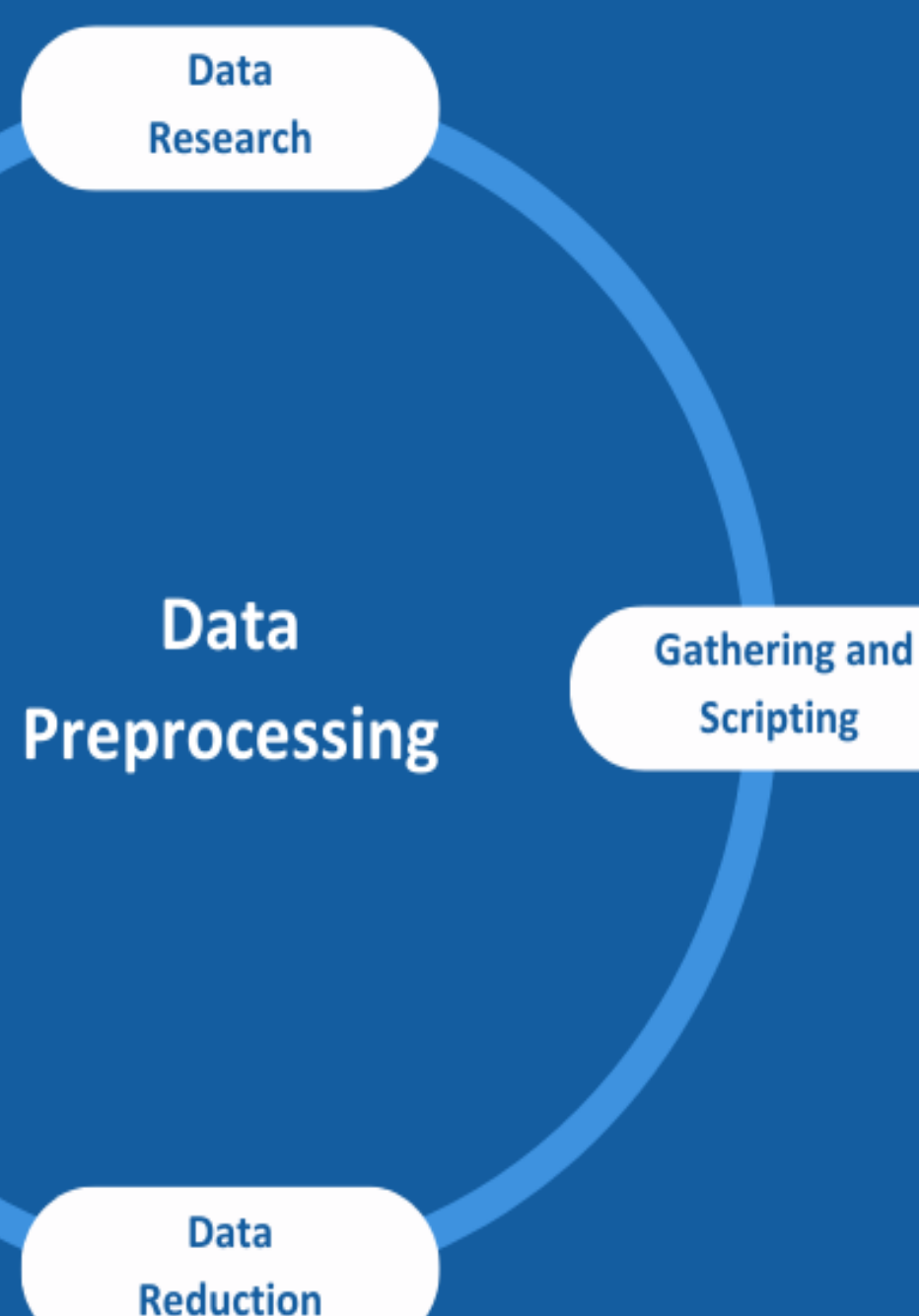
- A CVE (or Common Vulnerabilities and Exposures) is a unique identifier assigned to a security vulnerability in software or hardware.
- Since the early 2000's, we have seen a monumental rise in these vulnerabilities being found, and exploited everyday.
- CVE's are extremely important, as it tells security professionals what to update next, to keep your info secure!

Count of CVE's Reported by Year



Data

We acquired data from the Cybersecurity and Infrastructure Security Agency's repository of cyber threat reports. We used PostgreSQL format with Docker for secure data storage and easier processing.



Acknowledgments

Chad Zimmerman CISSP LM Associate Fellow, Dr. Daniel E. Hirleman, Ms. Gretchen Bliss, Prof. Bradley Warner, TAs: Grant Parker & Gregg Puttkammer, Kali Lacy, Lauren Dalder & The Data Mine Staff

LOCKHEED MARTIN



Current Threat Report

Prompt for Cybersecurity Threat Report

"Design a cybersecurity threat report for the vulnerability listed as CVE-2023-34362, provide all referencing documentation on the vulnerability and any remediation steps or mitigation steps"

Cybersecurity Threat Report

- Vulnerability Details:**
 - CVE ID: CVE-2023-34362
 - Vendor/Project: Progress
 - Product: MOVEit Transfer
 - Vulnerability Name: Progress MOVEit Transfer SQL Injection Vulnerability
 - Date Added: June 2, 2023
 - Short Description: Progress MOVEit Transfer contains a SQL injection vulnerability that could allow an unauthenticated attacker to gain unauthorized access to MOVEit Transfer's database. The attacker may be able to infer information about the database structure and contents and execute SQL statements for alterations or deletions.
- Referencing Documentation:**
 - CISA Alert: [CISA ALERT](#)
 - Additional Information: [MOVEit Transfer Critical Vulnerability](#)
- Remediation Steps:**
 - **Required Action:** Apply updates per vendor instructions.
 - **Due Date:** June 23, 2023
 - **Known Ransomware Campaign Use:** Known
 - **Notes:**
 - Review the CISA AA for associated IOCs.
 - Follow vendor instructions for updates and patches.
 - Monitor for any unauthorized access or unusual activities in the MOVEit Transfer system.
 - Implement security best practices to prevent SQL injection attacks.

It is crucial for organizations using the Progress MOVEit Transfer product to promptly apply the necessary updates and follow the mitigation steps outlined to protect their systems from potential exploitation of this SQL injection vulnerability. Stay vigilant and prioritize security measures to safeguard against cyber threats.

"Ideal" Threat Report

XZ-SSH Cyber Threat Report

Introduction

Firmware security firm Binarly has released a free online scanner to detect Linux executables impacted by the XZ Utils supply chain attack, tracked as CVE-2024-3094.

CVE-2024-3094 is a supply chain compromise in XZ Utils, a set of data compression tools and libraries used in many major Linux distributions.

Late last month, Microsoft engineer Andres Freud discovered the backdoor in the latest version of the XZ Utils package while investigating unusually slow SSH logins on Debian Sid, a rolling release of the Linux distribution.

Description

The XZ scanner Binarly says the approach taken so far in the threat mitigation efforts relies on simple checks such as byte string matching, file hash blocklisting, and YARA rules, which could lead to false positives.

This approach can trigger significant alert fatigue and doesn't help detect similar backdoors on other projects. To address this problem, Binarly developed a dedicated scanner that would work for the particular library and any file carrying the same backdoor.

"Such a complex and professionally designed comprehensive implantation framework is not developed for a one-shot operation.

It could already be deployed elsewhere or partially reused in other operations. That's exactly why we started focusing on more generic detection for this complex backdoor." - Binarly
Binarly's detection method employs static analysis of binaries to identify tampering of transitions in GNU Indirect Function (IFUNC).
"One of the core techniques used by the XZ backdoor to gain initial control during execution is the GNU Indirect Function (ifunc) attribute for the GCC compiler to resolve indirect function calls in runtime," explains Binarly.
"The implanted backdoor code initially intercepts or hooks execution.
"It modifies ifunc calls to replace a check "is_arch_extension_supported" which should simply invoke "cpuid" which inserts a call to "_get_cpuid" which is exported by the payload object file (i.e., liblzma_la-crc64-fast.o) and which calls malformed _get_cpuid() which is implanted into the code shown in the figure below."

Findings

"This detection is based on behavioral analysis and can detect any variants automatically if a similar backdoor is implanted somewhere else," Binarly's lead security researcher and CEO, Alex Matrosov, told BleepingComputer.

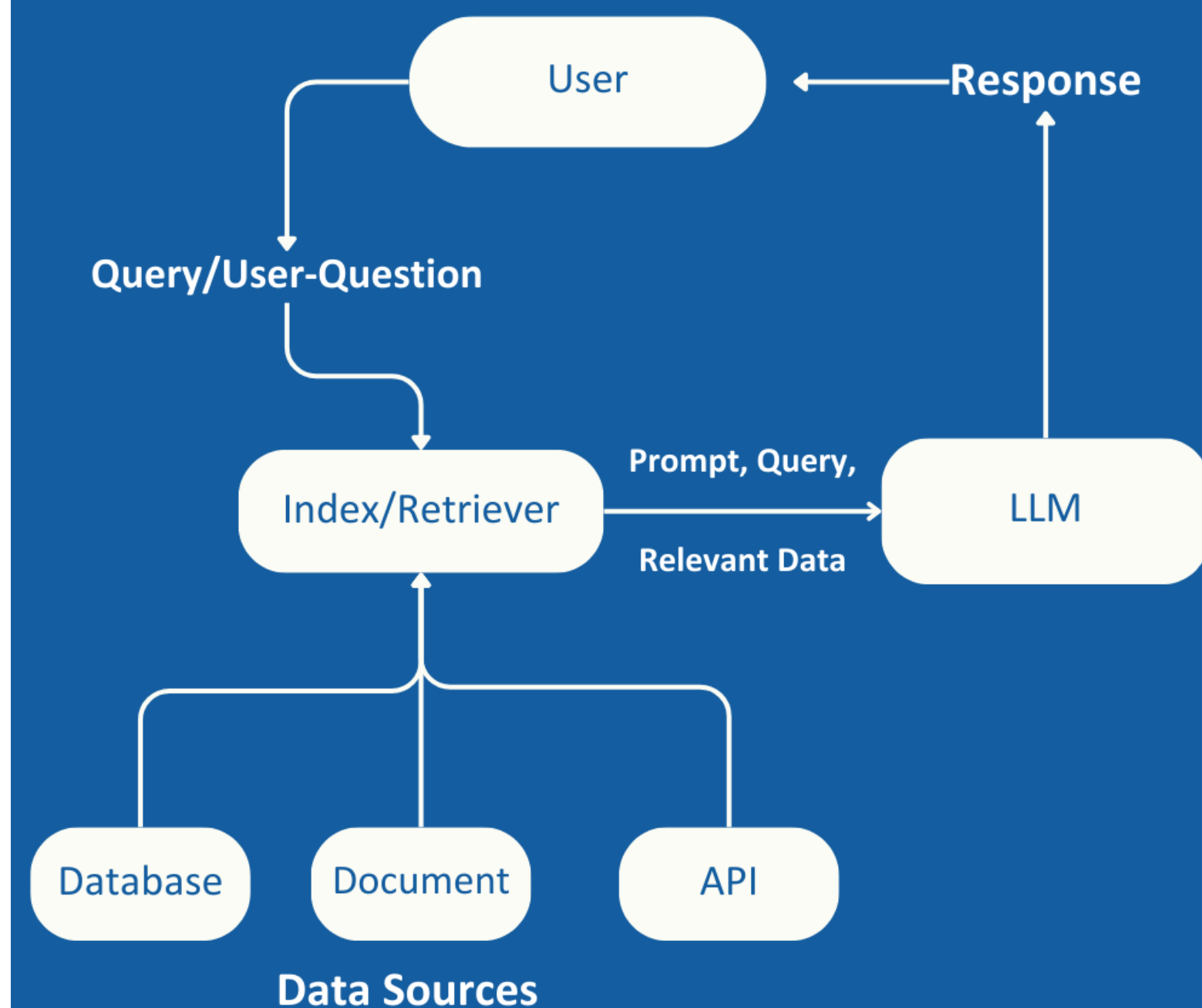
The backdoor scanner is available online at xz.fail, where people can upload their binary files for unlimited free checks.

Update 4/2 - Binarly has made a free API available to accommodate bulk scans for those who need it.

LLM

- Chose BLOOM for its smaller LLM model with 560 million parameters
- Despite its size, localizing BLOOM for training and testing on our personal devices was challenging and time-consuming
- Tested other models to compare and contrast how different models react to our input
- In a professional setting with better computational resources, opting for a larger model would expedite product development

User-LLM Interaction



Evaluation Criteria

When evaluating our model, we decided upon the following evaluation metrics:

- Accuracy of data provided
- Relevance prompted information
- Coherence and clarity of report
- Completeness of details

These metrics are important as they ensure the model generates credible, relevant information

Conclusion

We were successfully able to train our LLM through fine-tuning and RAG with our data, however, limited computer power stunted our results. Our LLM would give us an output, but we were never satisfied with the result. With more resources and time we would continue our process of training, testing, and evaluating until we had a worthy product.

Future Goals

- Implement real-time threat reporting and fully automate the threat analysis process
- Expand product to include in-depth user interface

References

"High-Level Concepts#." *LlamaIndex*, docs.llamaindex.ai/en/stable/getting_started/concepts/