cs-sa07-24019
John_Mbithi_Mutave

Intro to Log Analysis

Introduction

Log analysis is a crucial aspect of cybersecurity, providing the means to monitor, detect, and respond to security events within an organization. This report offers an introduction to log analysis, covering the basics, best practices, and essential tools required for effective detection and response. The aim is to equip cybersecurity professionals with the knowledge and skills needed to analyze logs efficiently, identify potential threats, and take appropriate actions to mitigate risks.

Task 1: Introduction

Log analysis involves examining and interpreting data generated by various systems, applications, and devices within a network. These logs provide a detailed record of events, which can be analyzed to detect anomalies, identify security incidents, and investigate the root causes of issues. Effective log analysis is vital for maintaining the security and integrity of an organization’s IT infrastructure.

Task 2: Log Analysis Basics

Key Concepts

- **Logs**: Structured or unstructured data generated by operating systems, applications, network devices, and other sources. They record events such as user activities, system errors, and security alerts.
- **Log Sources**: Various components within an IT environment that generate logs, including servers, firewalls, routers, and applications.
- **Log Collection**: The process of gathering logs from multiple sources and storing them in a central repository for analysis.
- **Log Parsing**: Converting raw log data into a structured format that can be easily analyzed.
- **Log Normalization**: Standardizing log data to ensure consistency across different log sources.

Importance of Log Analysis

- **Security Monitoring**: Identifying potential security incidents and breaches.
- **Compliance**: Ensuring adherence to regulatory requirements by maintaining detailed logs.
- **Troubleshooting**: Diagnosing and resolving technical issues.
- **Operational Insights**: Gaining visibility into system performance and user behavior.

Task 3: Investigation Theory

Investigation Process

1. **Log Collection**: Aggregating logs from various sources.
2. **Log Parsing and Normalization**: Structuring and standardizing log data.
3. **Log Enrichment**: Adding context to logs by correlating with other data sources.
4. **Event Correlation**: Identifying patterns and relationships between different log events.
5. **Anomaly Detection**: Using statistical methods and machine learning to identify unusual behavior.
6. **Incident Response**: Taking appropriate actions to mitigate identified threats.

Key Principles

- **Accuracy**: Ensuring the integrity and reliability of log data.
- **Timeliness**: Analyzing logs in real-time to detect and respond to incidents promptly.
- **Context**: Understanding the broader context of log events to accurately interpret their significance.

Task 4: Detection Engineering

Definition

Detection engineering involves designing and implementing detection mechanisms to identify security threats within log data.

Components

- **Indicators of Compromise (IOCs)**: Artifacts observed in logs that indicate potential malicious activity.
- **Detection Rules**: Predefined patterns and conditions used to identify IOCs.
- **Threat Intelligence**: Information about current threats and attack techniques used to inform detection rules.
- **Automation**: Leveraging automated tools and scripts to enhance detection capabilities.

Best Practices

- **Regular Updates**: Continuously updating detection rules based on emerging threats.
- **Customization**: Tailoring detection mechanisms to the specific environment and threat landscape.
- **Testing and Validation**: Regularly testing detection rules to ensure their effectiveness.

Task 5: Automated vs. Manual Analysis

Automated Analysis

- **Advantages**: Speed, scalability, and the ability to handle large volumes of data.
- **Tools**: Security Information and Event Management (SIEM) systems, automated scripts, and machine learning algorithms.

- **Limitations**: Potential for false positives and the need for periodic tuning and maintenance.

Manual Analysis

- **Advantages**: Human intuition, contextual understanding, and the ability to investigate complex incidents.
- **Techniques**: Manual log review, use of specialized analysis tools, and correlation with other data sources.
- **Challenges**: Time-consuming and resource-intensive.

Hybrid Approach

Combining automated and manual analysis to leverage the strengths of both methods. Automated tools can handle routine tasks and large-scale data processing, while human analysts focus on complex investigations and contextual interpretation.

Task 6: Log Analysis Tools: Command Line

Common Command-Line Tools

- **grep**: A powerful search tool for finding patterns in log files.
- **awk**: A programming language for text processing and data extraction.
- **sed**: A stream editor for parsing and transforming text.
- **tail**: A utility for viewing the end of log files in real-time.
- **cut**: A tool for extracting specific fields from log data.

Usage Examples

- **grep**: grep "error" /var/log/syslog - Searches for the keyword "error" in the syslog file.
- **awk**: awk '{print $1, $3, $5}' /var/log/syslog - Extracts the first, third, and fifth columns from the syslog file.

- **sed**: sed 's/error/ERROR/' /var/log/syslog - Replaces the word "error" with "ERROR" in the syslog file.

Task 7: Log Analysis Tools: Regular Expressions

Overview

Regular expressions (regex) are sequences of characters that define search patterns. They are used to identify specific patterns within log data.

Common Regex Patterns

- **IP Address**: \b\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}\b
- **Email Address**: [a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}
- **Date and Time**: \d{4}-\d{2}-\d{2} \d{2}:\d{2}:\d{2}
- **URLs**: https?://[^\s/$.?#].[^\s]*

Applications

- **Pattern Matching**: Identifying specific log entries based on predefined patterns.
- **Data Extraction**: Extracting relevant information from log data for further analysis.
- **Filtering**: Isolating log entries that match specific criteria.

Task 8: Log Analysis Tools: CyberChef

Introduction

CyberChef is a web-based tool that provides a wide range of data analysis and manipulation functions. It is designed to simplify complex data transformations and make log analysis more efficient.

Key Features

- **Data Conversion**: Converting data between different formats (e.g., hex, base64, binary).
- **Encryption/Decryption**: Applying cryptographic functions to data.
- **Text Analysis**: Performing text manipulation and pattern matching.
- **Data Extraction**: Extracting and parsing specific elements from log data.

Usage Examples

- **Base64 Decoding**: Decoding a base64-encoded log entry to reveal its original content.
- **Regex Match**: Using regex to extract IP addresses from log data.
- **Hashing**: Generating hash values (e.g., MD5, SHA-256) for log entries.

Task 9: Log Analysis Tools: Yara and Sigma

Yara

Yara is a tool used for creating and executing rules to identify specific patterns within files and processes. It is widely used in malware analysis and threat hunting.

Sigma

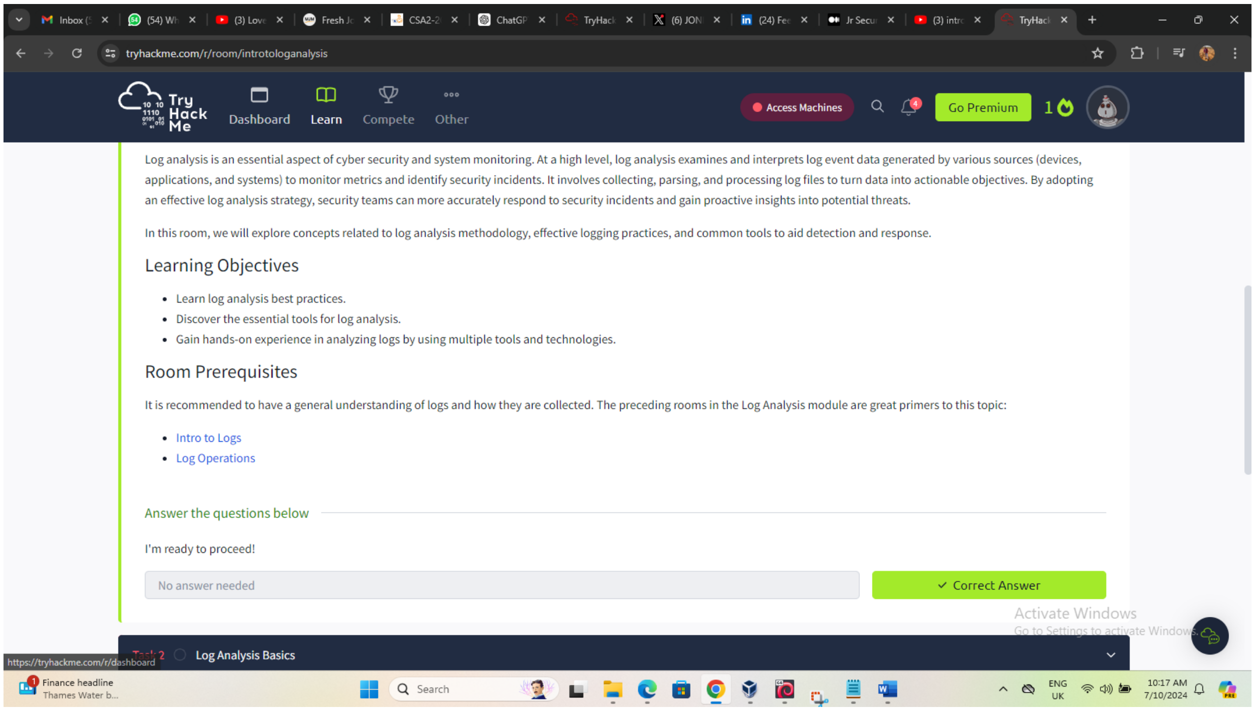
Sigma is an open standard for writing rules to detect suspicious activity in log data. Sigma rules can be converted to various formats compatible with different SIEM systems.

Applications

- **Threat Detection**: Using Yara and Sigma rules to detect known indicators of compromise.
- **Custom Rules**: Writing custom detection rules tailored to the specific environment and threat landscape.

- **Integration**: Integrating Yara and Sigma with existing security tools and workflows.

Screenshot overview of the task



Log analysis is an essential aspect of cyber security and system monitoring. At a high level, log analysis examines and interprets log event data generated by various sources (devices, applications, and systems) to monitor metrics and identify security incidents. It involves collecting, parsing, and processing log files to turn data into actionable objectives. By adopting an effective log analysis strategy, security teams can more accurately respond to security incidents and gain proactive insights into potential threats.

In this room, we will explore concepts related to log analysis methodology, effective logging practices, and common tools to aid detection and response.

Learning Objectives

- Learn log analysis best practices.
- Discover the essential tools for log analysis.
- Gain hands-on experience in analyzing logs by using multiple tools and technologies.

Room Prerequisites

It is recommended to have a general understanding of logs and how they are collected. The preceding rooms in the Log Analysis module are great primers to this topic:

- Intro to Logs
- Log Operations

Answer the questions below

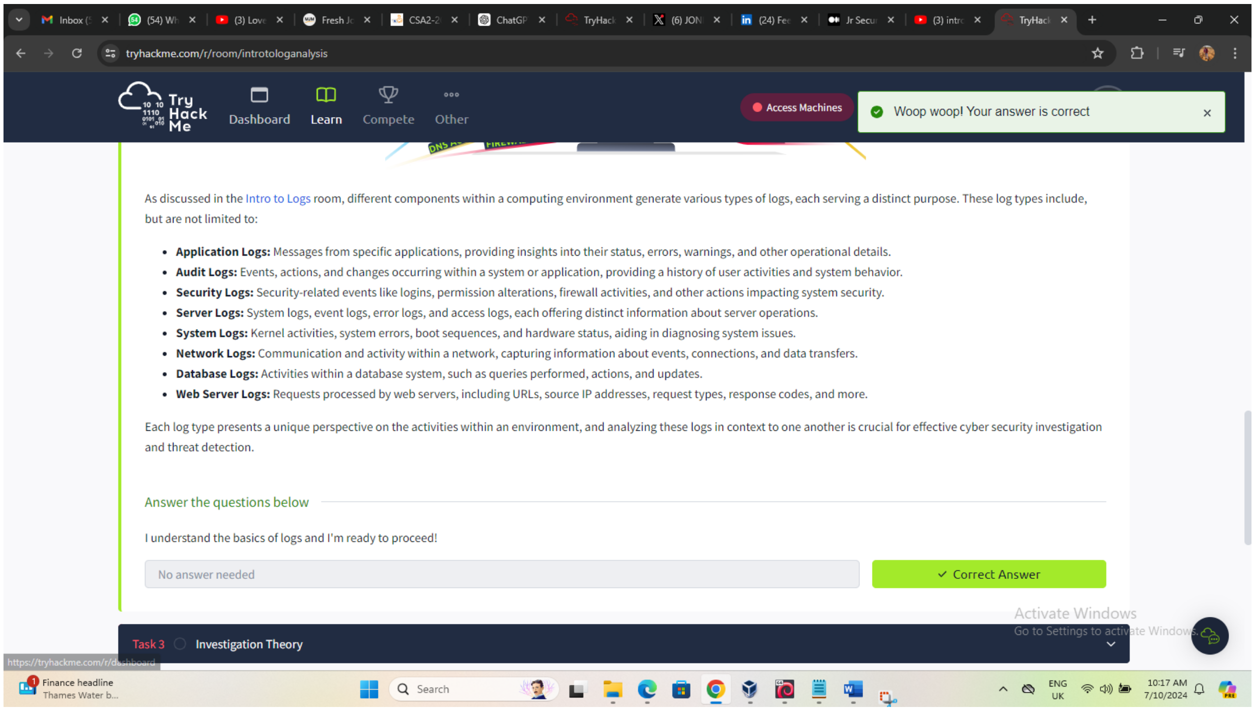
I'm ready to proceed!

No answer needed

✓ Correct Answer

cs-sa07-24019

John_Mbithi_Mutave



tryhackme.com/r/room/introtologanalysis

Woop woop! Your answer is correct

As discussed in the Intro to Logs room, different components within a computing environment generate various types of logs, each serving a distinct purpose. These log types include, but are not limited to:

- **Application Logs:** Messages from specific applications, providing insights into their status, errors, warnings, and other operational details.
- **Audit Logs:** Events, actions, and changes occurring within a system or application, providing a history of user activities and system behavior.
- **Security Logs:** Security-related events like logins, permission alterations, firewall activities, and other actions impacting system security.
- **Server Logs:** System logs, event logs, error logs, and access logs, each offering distinct information about server operations.
- **System Logs:** Kernel activities, system errors, boot sequences, and hardware status, aiding in diagnosing system issues.
- **Network Logs:** Communication and activity within a network, capturing information about events, connections, and data transfers.
- **Database Logs:** Activities within a database system, such as queries performed, actions, and updates.
- **Web Server Logs:** Requests processed by web servers, including URLs, source IP addresses, request types, response codes, and more.

Each log type presents a unique perspective on the activities within an environment, and analyzing these logs in context to one another is crucial for effective cyber security investigation and threat detection.

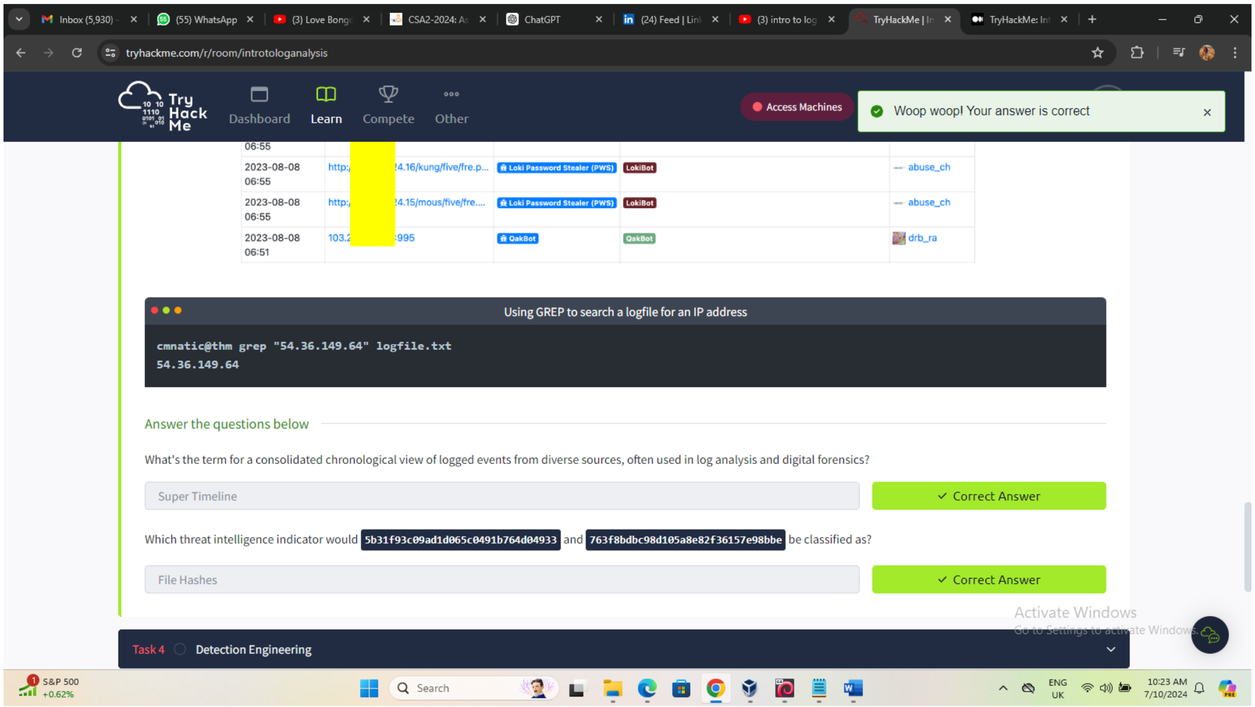
Answer the questions below

I understand the basics of logs and I'm ready to proceed!

No answer needed

✓ Correct Answer

Task 3 Investigation Theory



tryhackme.com/r/room/introtologanalysis

Woop woop! Your answer is correct

06:55				
2023-08-08 06:55	http:/[redacted]24.16/kung/five/fre.p...	Loki Password Stealer (PWS)	LokiBot	abuse_ch
2023-08-08 06:55	http:/[redacted]24.15/mous/five/fre....	Loki Password Stealer (PWS)	LokiBot	abuse_ch
2023-08-08 06:51	103.2[redacted]:995	QakBot	QakBot	drb_ra

Using GREP to search a logfile for an IP address

```
cmnatic@thm grep "54.36.149.64" logfile.txt
54.36.149.64
```

Answer the questions below

What's the term for a consolidated chronological view of logged events from diverse sources, often used in log analysis and digital forensics?

Super Timeline

✓ Correct Answer

Which threat intelligence indicator would `5b31f93c09ad1d065c0491b764d04933` and `763f8bdbc98d105a8e82f36157e98bbe` be classified as?

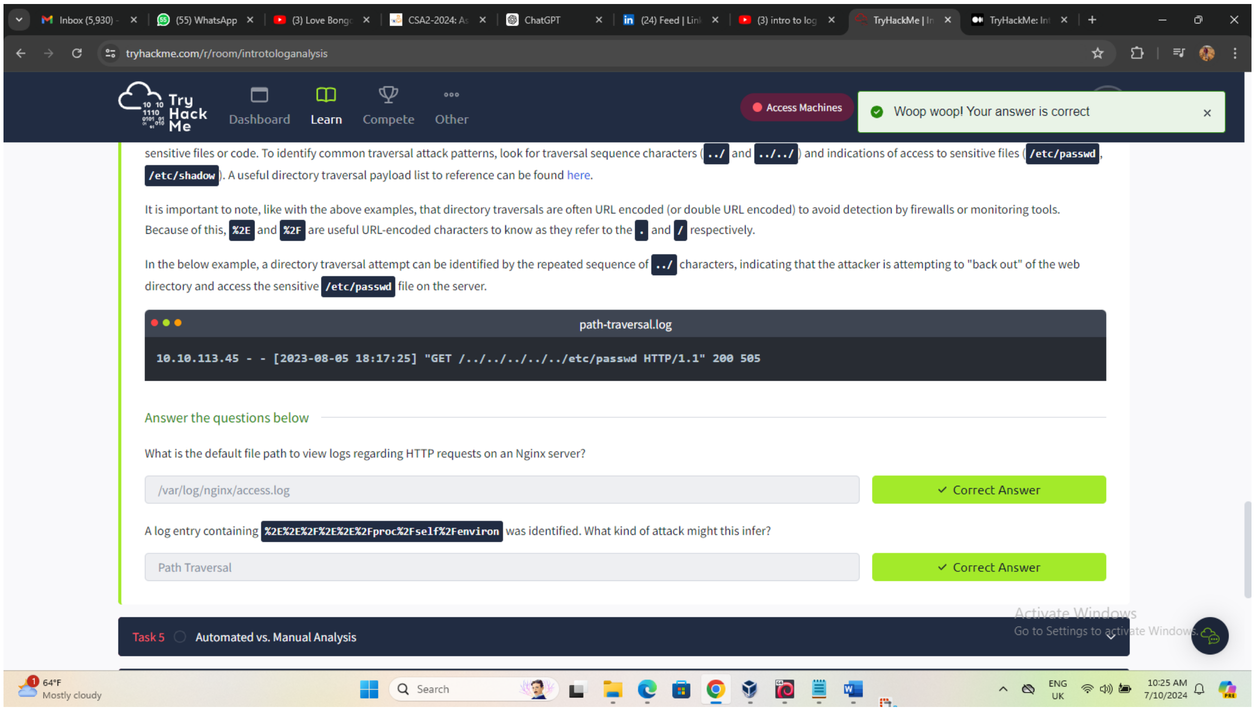
File Hashes

✓ Correct Answer

Task 4 Detection Engineering

cs-sa07-24019

John_Mbithi_Mutave



tryhackme.com/r/room/introtologanalysis
Dashboard
Learn
Compete
Other
Access Machines
Woop woop! Your answer is correct

sensitive files or code. To identify common traversal attack patterns, look for traversal sequence characters (`../` and `../../`) and indications of access to sensitive files (`/etc/passwd`, `/etc/shadow`). A useful directory traversal payload list to reference can be found here.

It is important to note, like with the above examples, that directory traversals are often URL encoded (or double URL encoded) to avoid detection by firewalls or monitoring tools. Because of this, `%2E` and `%2F` are useful URL-encoded characters to know as they refer to the `.` and `/` respectively.

In the below example, a directory traversal attempt can be identified by the repeated sequence of `../` characters, indicating that the attacker is attempting to "back out" of the web directory and access the sensitive `/etc/passwd` file on the server.

path-traversal.log

```
10.10.113.45 - - [2023-08-05 18:17:25] "GET /../../../../../etc/passwd HTTP/1.1" 200 505
```

Answer the questions below

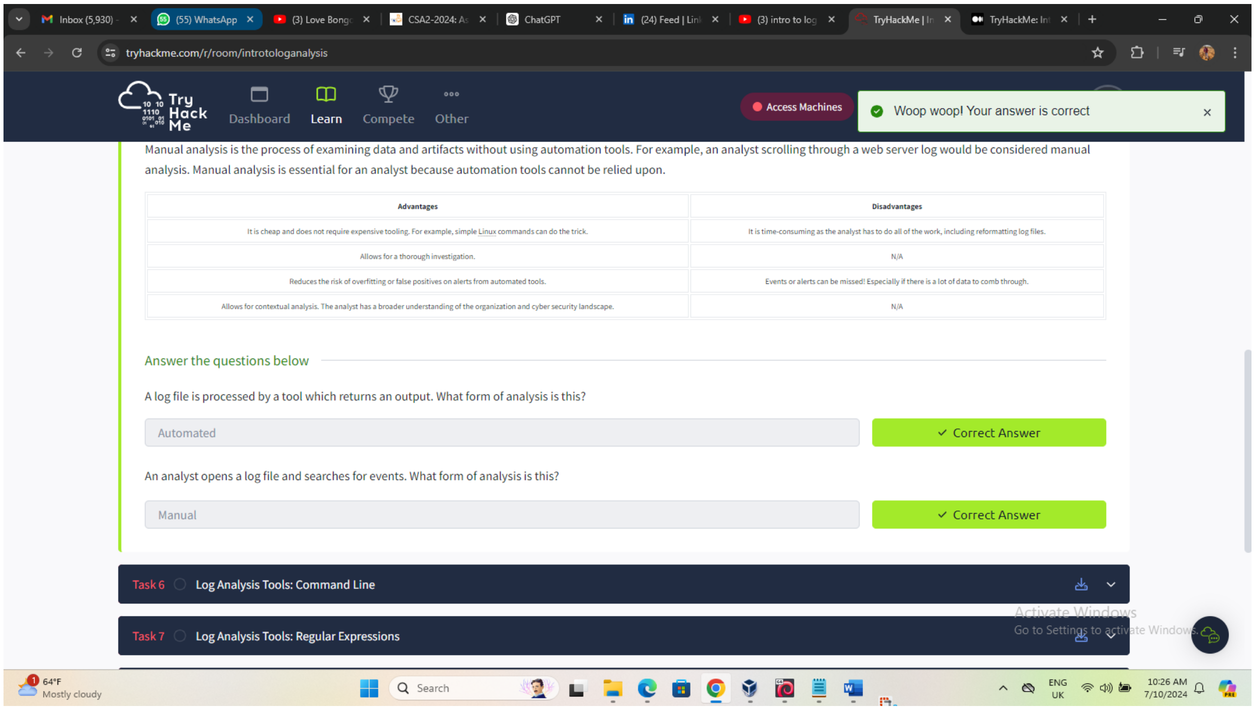
What is the default file path to view logs regarding HTTP requests on an Nginx server?

/var/log/nginx/access.log ✓ Correct Answer

A log entry containing `%2E%2E%2F%2E%2E%2Fproc%2Fself%2Fenviron` was identified. What kind of attack might this infer?

Path Traversal ✓ Correct Answer

Task 5 Automated vs. Manual Analysis



tryhackme.com/r/room/introtologanalysis
Dashboard
Learn
Compete
Other
Access Machines
Woop woop! Your answer is correct

Manual analysis is the process of examining data and artifacts without using automation tools. For example, an analyst scrolling through a web server log would be considered manual analysis. Manual analysis is essential for an analyst because automation tools cannot be relied upon.

Advantages	Disadvantages
It is cheap and does not require expensive tooling. For example, simple Linux commands can do the trick.	It is time-consuming as the analyst has to do all of the work, including reformatting log files.
Allows for a thorough investigation.	N/A
Reduces the risk of overfitting or false positives on alerts from automated tools.	Events or alerts can be missed! Especially if there is a lot of data to comb through.
Allows for contextual analysis. The analyst has a broader understanding of the organization and cyber security landscape.	N/A

Answer the questions below

A log file is processed by a tool which returns an output. What form of analysis is this?

Automated ✓ Correct Answer

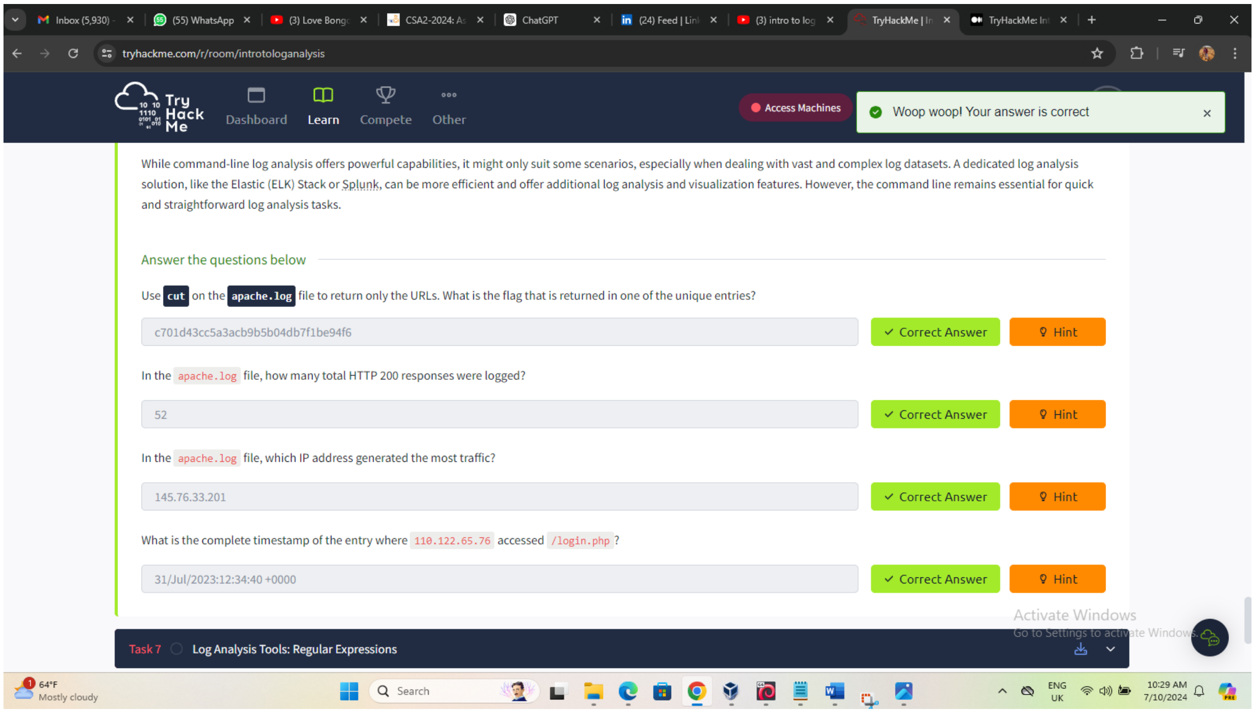
An analyst opens a log file and searches for events. What form of analysis is this?

Manual ✓ Correct Answer

Task 6 Log Analysis Tools: Command Line

Task 7 Log Analysis Tools: Regular Expressions

cs-sa07-24019
John_Mbithi_Mutave



Woop woop! Your answer is correct

While command-line log analysis offers powerful capabilities, it might only suit some scenarios, especially when dealing with vast and complex log datasets. A dedicated log analysis solution, like the Elastic (ELK) Stack or Splunk, can be more efficient and offer additional log analysis and visualization features. However, the command line remains essential for quick and straightforward log analysis tasks.

Answer the questions below

Use `cut` on the `apache.log` file to return only the URLs. What is the flag that is returned in one of the unique entries?

c701d43cc5a3acb9b5b04db7f1be94f6

✓ Correct Answer Hint

In the `apache.log` file, how many total HTTP 200 responses were logged?

52

✓ Correct Answer Hint

In the `apache.log` file, which IP address generated the most traffic?

145.76.33.201

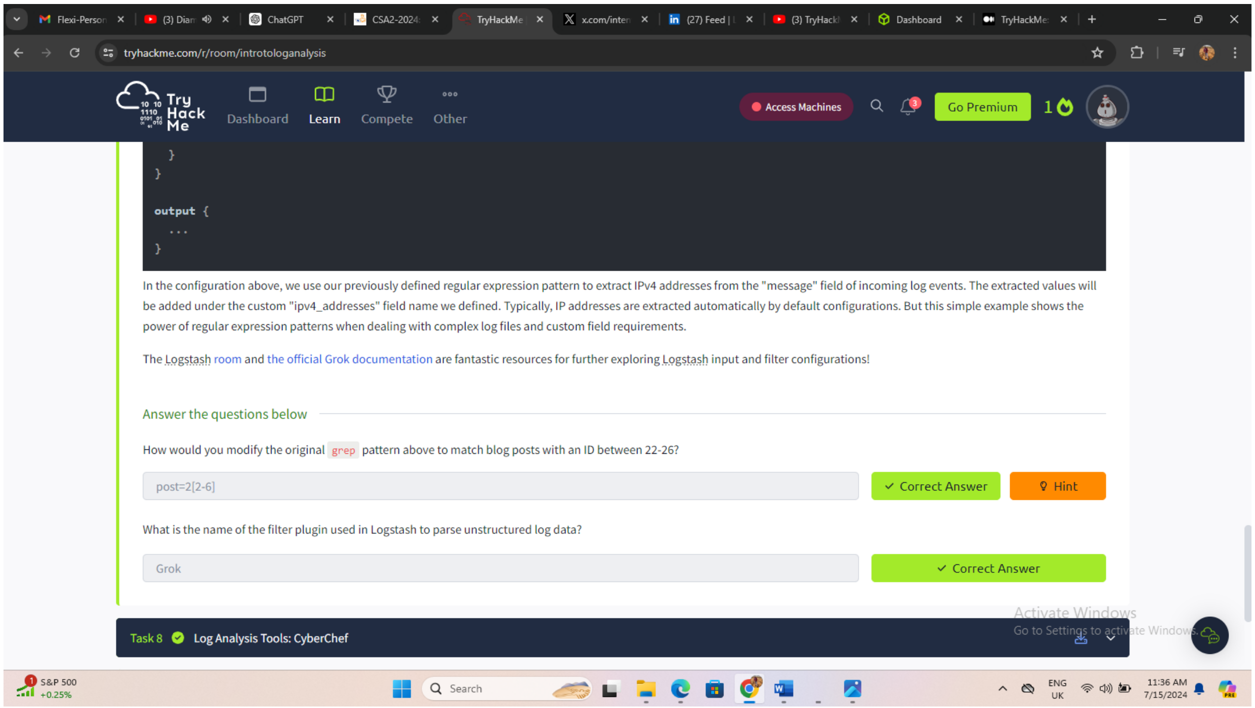
✓ Correct Answer Hint

What is the complete timestamp of the entry where `110.122.65.76` accessed `/login.php`?

31/Jul/2023:12:34:40 +0000

✓ Correct Answer Hint

Task 7 ○ Log Analysis Tools: Regular Expressions



```
    }
}

output {
  ...
}
```

In the configuration above, we use our previously defined regular expression pattern to extract IPv4 addresses from the "message" field of incoming log events. The extracted values will be added under the custom "ipv4_addresses" field name we defined. Typically, IP addresses are extracted automatically by default configurations. But this simple example shows the power of regular expression patterns when dealing with complex log files and custom field requirements.

The Logstash room and the official Grok documentation are fantastic resources for further exploring Logstash input and filter configurations!

Answer the questions below

How would you modify the original `grep` pattern above to match blog posts with an ID between 22-26?

post=2[2-6]

✓ Correct Answer Hint

What is the name of the filter plugin used in Logstash to parse unstructured log data?

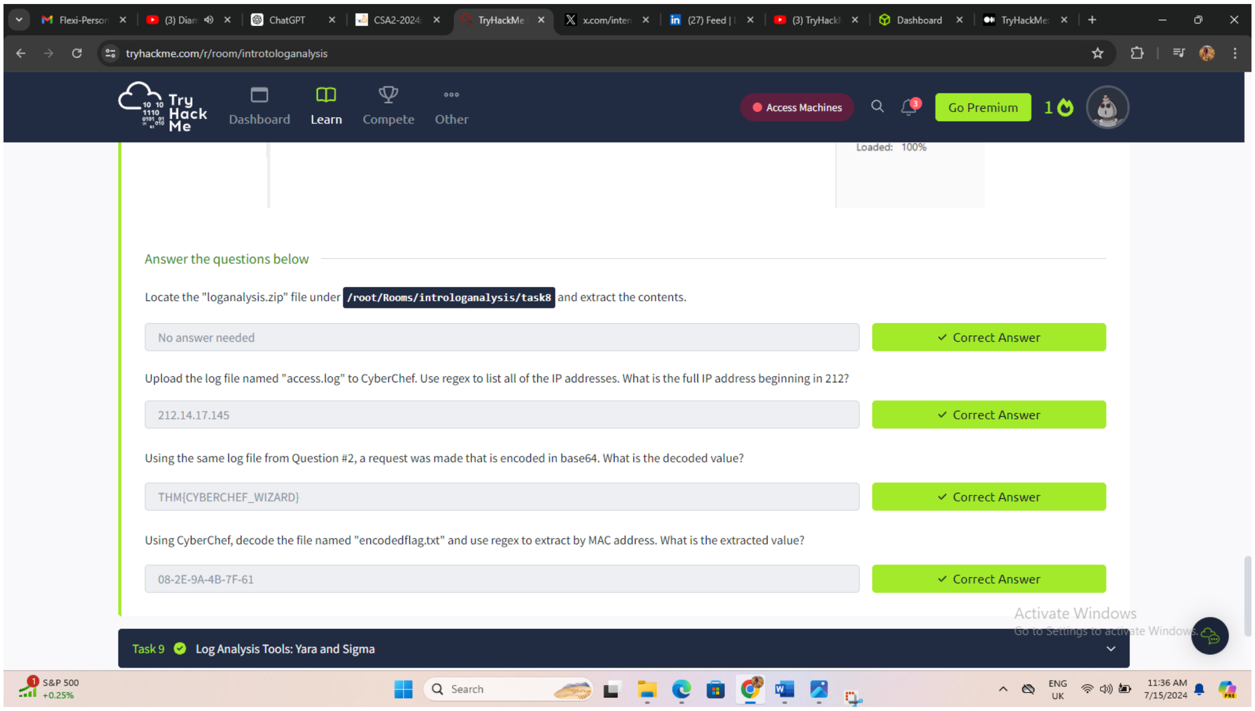
Grok

✓ Correct Answer

Task 8 ✓ Log Analysis Tools: CyberChef

cs-sa07-24019

John_Mbithi_Mutave



Answer the questions below

Locate the "loganalysis.zip" file under `/root/Rooms/introloganalysis/task8` and extract the contents.

No answer needed ✓ Correct Answer

Upload the log file named "access.log" to CyberChef. Use regex to list all of the IP addresses. What is the full IP address beginning in 212?

212.14.17.145 ✓ Correct Answer

Using the same log file from Question #2, a request was made that is encoded in base64. What is the decoded value?

THM{CYBERCHEF_WIZARD} ✓ Correct Answer

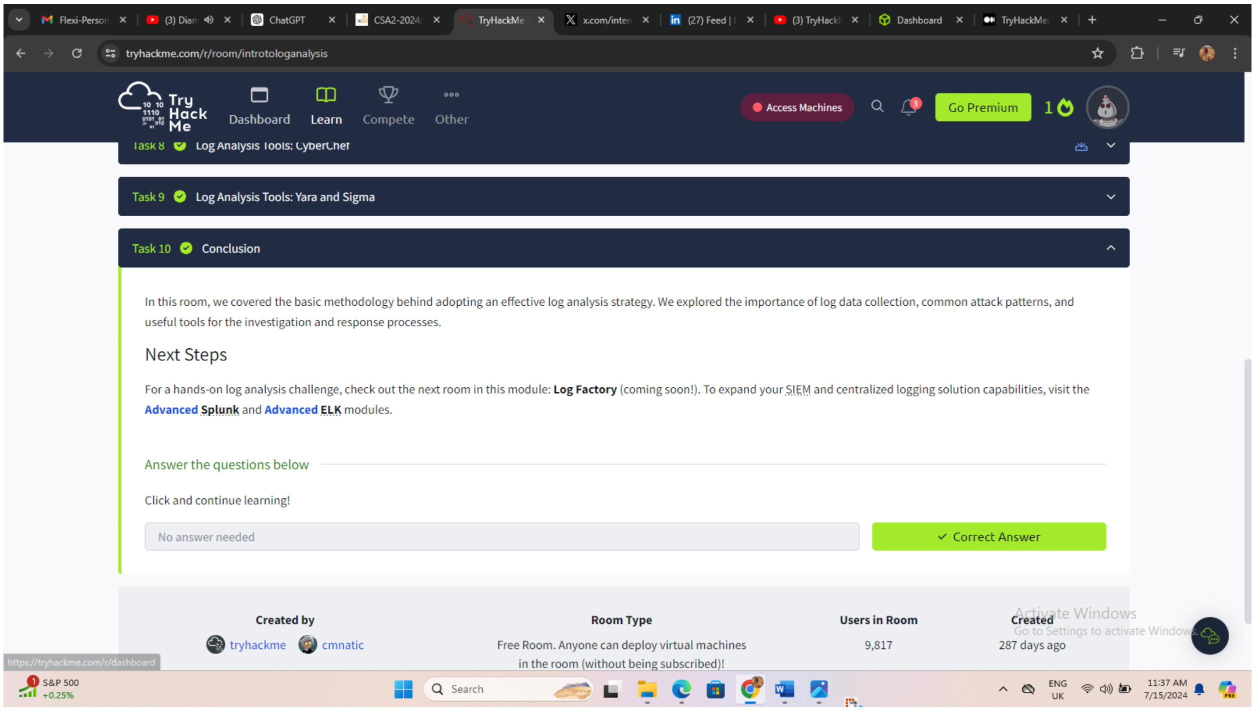
Using CyberChef, decode the file named "encodedflag.txt" and use regex to extract by MAC address. What is the extracted value?

08-2E-9A-4B-7F-61 ✓ Correct Answer

Task 9 Log Analysis Tools: Yara and Sigma

cs-sa07-24019

John_Mbithi_Mutave



Task 8 Log Analysis Tools: CyberChef

Task 9 Log Analysis Tools: Yara and Sigma

Task 10 Conclusion

In this room, we covered the basic methodology behind adopting an effective log analysis strategy. We explored the importance of log data collection, common attack patterns, and useful tools for the investigation and response processes.

Next Steps

For a hands-on log analysis challenge, check out the next room in this module: **Log Factory** (coming soon!). To expand your SIEM and centralized logging solution capabilities, visit the **Advanced Splunk** and **Advanced ELK** modules.

Answer the questions below

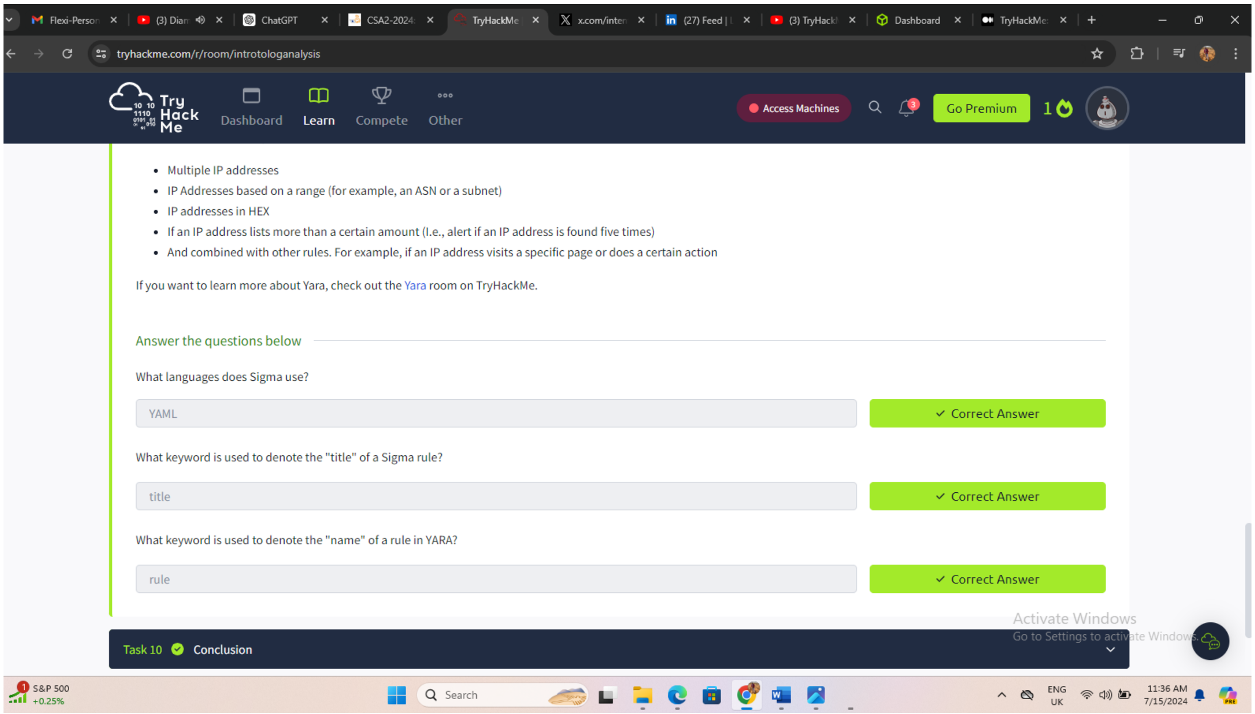
Click and continue learning!

No answer needed

✓ Correct Answer

Created by	Room Type	Users in Room	Created
tryhackme cmnatic	Free Room. Anyone can deploy virtual machines in the room (without being subscribed)!	9,817	287 days ago

https://tryhackme.com/r/dashboard
S&P 500 +0.25%
Search
ENG UK
11:37 AM 7/15/2024
Activate Windows
Go to Settings to activate Windows.



- Multiple IP addresses
- IP Addresses based on a range (for example, an ASN or a subnet)
- IP addresses in HEX
- If an IP address lists more than a certain amount (I.e., alert if an IP address is found five times)
- And combined with other rules. For example, if an IP address visits a specific page or does a certain action

If you want to learn more about Yara, check out the Yara room on TryHackMe.

Answer the questions below

What languages does Sigma use?

YAML

✓ Correct Answer

What keyword is used to denote the "title" of a Sigma rule?

title

✓ Correct Answer

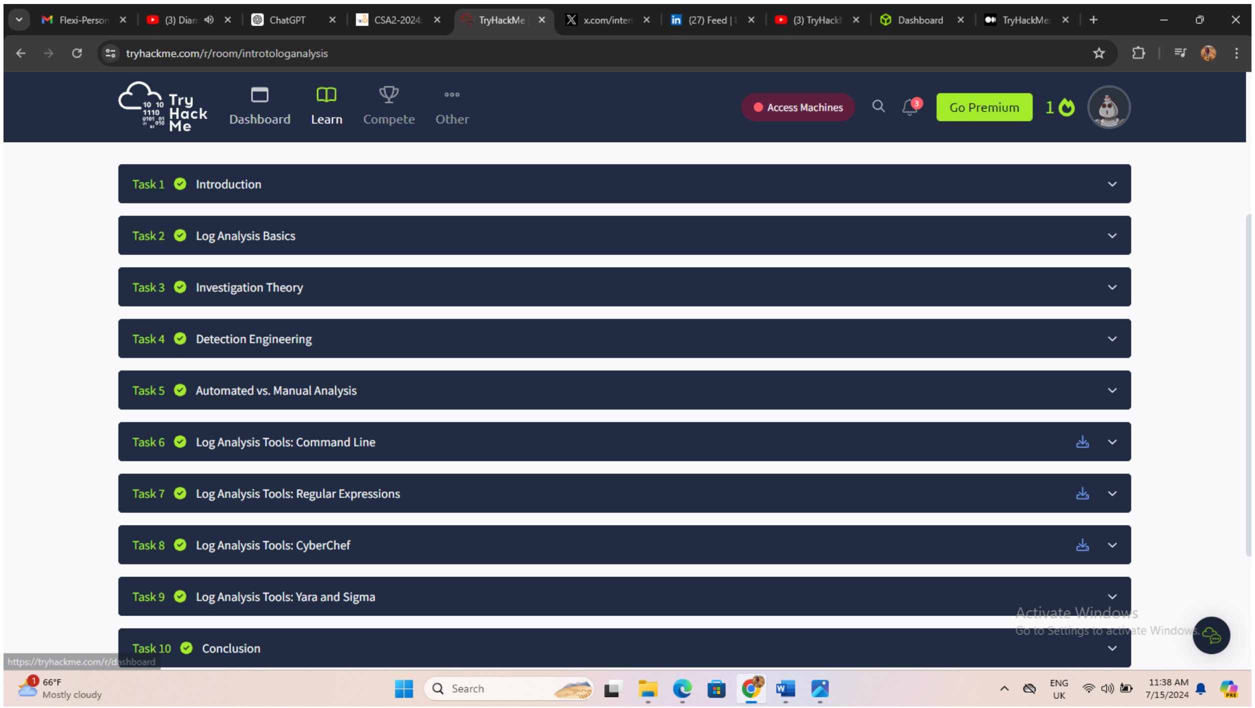
What keyword is used to denote the "name" of a rule in YARA?

rule

✓ Correct Answer

Task 10 Conclusion

cs-sa07-24019
John_Mbithi_Mutave



Task 1 ✓ Introduction

Task 2 ✓ Log Analysis Basics

Task 3 ✓ Investigation Theory

Task 4 ✓ Detection Engineering

Task 5 ✓ Automated vs. Manual Analysis

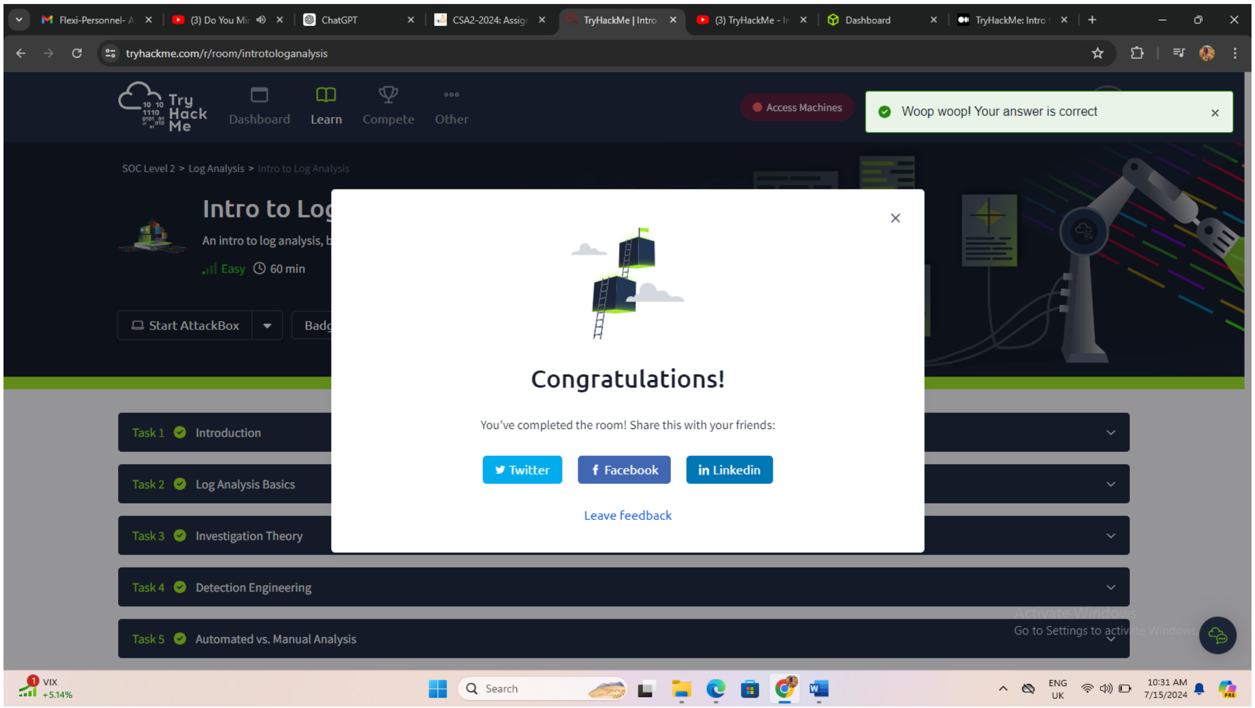
Task 6 ✓ Log Analysis Tools: Command Line

Task 7 ✓ Log Analysis Tools: Regular Expressions

Task 8 ✓ Log Analysis Tools: CyberChef

Task 9 ✓ Log Analysis Tools: Yara and Sigma

Task 10 ✓ Conclusion



Woop woop! Your answer is correct

Intro to Log

An intro to log analysis,

Easy 60 min

Congratulations!

You've completed the room! Share this with your friends:

Task 1 ✓ Introduction

Task 2 ✓ Log Analysis Basics

Task 3 ✓ Investigation Theory

Task 4 ✓ Detection Engineering

Task 5 ✓ Automated vs. Manual Analysis

Shareable link - https://tryhackme.com/r/room/introtologanalysis

Task 10: Conclusion

Log analysis is a vital component of cybersecurity, enabling organizations to monitor, detect, and respond to security incidents effectively. By understanding the basics of log analysis, adhering to best practices, and leveraging essential tools, security professionals can enhance their detection and response capabilities. The continuous evolution of threats necessitates ongoing learning and adaptation to ensure robust log analysis practices and maintain a secure IT environment. This introductory course provides a solid foundation for further exploration and mastery of log analysis techniques and tools.