Security Information and Event Management System (SIEM)

Prepared By:
Kazim Ali Obad

Supervisor:

Anmar Mohammed
MOHAMMED .B. HASSAN

Contents

1. What Is a SIEM?

SIEM is something that comes up constantly in interviews, in client meetings, everywhere. We are in the age of data, and no serious organization is running without a SIEM solution.

So what exactly is a SIEM? The abbreviation stands for:

- Security Information and Event Management System

Think of it this way. Inside a company you have workstations, servers, firewalls, switches, routers, cameras, electronic access control doors with fingerprint readers everything. Every single one of those devices is constantly writing log entries about what is happening: someone opened a file, someone tried to log in, a network connection was made, a process started. All of that generates what we call events or logs.

Before SIEM existed, if an incident happened you would have to walk to each machine one by one, open its event viewer or log files, and manually look through hundreds of events to figure out what went wrong. Imagine you have a hundred computers. That is not realistic.

The SIEM comes in and says: "I will collect ALL the logs from ALL your devices into one centralized place. Then I will normalize that raw, messy data, make it readable, correlate the events, and present it to you on a single dashboard so you can see everything that is happening across the entire organization in real time."

> ***Note: The SIEM is your eyes. It does not block attacks by itself it gives you visibility so you can act.***

2. Network Visibility Through SIEM

One of the most important things the SIEM gives you is what we call network visibility. Every network component can generate one or more log sources. We divide those log sources into two logical categories:

2.1 Host Centric Log Sources

These are logs generated from within a host machine itself a workstation, a server, an endpoint. Examples of tools that produce host centric logs:

- Windows Event Logs
- Sysmon

The kinds of activities captured in host centric logs include:

- A user accessing or modifying a file
- A user attempting to authenticate
- A process execution event
- A process adding, editing, or deleting a registry key
- PowerShell execution

2.2 Network Centric Log Sources

These are logs generated when hosts communicate with each other or with the outside world. Network based protocols like SSH, VPN, HTTP/S, FTP produce these. Examples:

- An SSH connection being established
- A file accessed over FTP
- Web traffic
- A user connecting to company resources through a VPN

3. Why Do We Need SIEM?

All of these devices generate hundreds of events per second. There is no human being on earth who can manually review every log entry. That is the problem SIEM solves. Here are the key capabilities it provides:

- Realtime log ingestion it is watching live, not after the fact
- Alerting against abnormal activities
- 24/7 monitoring and visibility
- Early detection of threats before damage is done
- Data insights and visualization through dashboards
- Ability to investigate past incidents going back in time through stored logs

There have been cases major banks, Disney, Saudi Aramco where attackers were inside the network for months or even years without being detected. Why? Because there was no proper visibility. If you have a SIEM with good correlation rules and proper log coverage, the story is completely different.

4. Log Sources and Log Ingestion

4.1 Windows Event Logs

In Windows, every event is recorded and can be viewed through the Event Viewer utility. Windows assigns a unique Event ID to each type of activity, which makes tracking and correlating events much easier. For example:

- Event ID 4688 A new process has been created
- Event ID 104 The event log was cleared
- Event ID 4624 Successful logon
- Event ID 4625 Failed login attempt

The Event Viewer is the built in Windows tool for looking at all these logs. The SIEM pulls those logs from every Windows endpoint in your environment and brings them together in one place so you can actually work with them.

4.2 Linux Log Locations

Linux stores all related logs events, errors, warnings in specific file paths. The main locations the SIEM reads from are:

```
/var/log/httpd   HTTP request, response, and error logs
/var/log/cron    Events related to cron jobs
/var/log/auth.log  or  /var/log/secure   Authentication logs
/var/log/kern    Kernelrelated events
```

4.3 Web Server Logs

For web servers, it is critical to monitor all requests and responses coming in and out for any potential web attack attempt. In Linux, Apache logs are typically found at /var/log/apache or /var/log/httpd.

4.4 How Logs Get Into the SIEM Log Ingestion Methods

Each SIEM solution has its own way of collecting logs. The four main methods are:

1. **Method 1: Agent / Forwarder** A lightweight software agent is installed on the endpoint. It is configured to capture the important logs and forward them to the SIEM server. In Splunk this is called the Universal Forwarder.
2. **Method 2: Syslog** A widely used protocol that allows devices like web servers, databases, and network appliances to send logs in real time to a centralized destination.
3. **Method 3: Manual** Upload Some SIEMs like Splunk and ELK allow you to upload an offline log file directly for quick analysis. Once uploaded, the data is normalized and made available for searching.
4. **Method 4: Port Forwarding** The SIEM is configured to listen on a specific port, and the endpoints forward their data directly to that port.

5. Dashboards

After logs are normalized and ingested, the SIEM presents everything in the form of actionable insights on multiple dashboards. Think of it as the cockpit from which you monitor the entire environment.

Every SIEM product comes with default dashboards and also lets you build custom ones. Information you typically find on a SIEM dashboard includes:

- Alert highlights and critical events
- System notifications and health alerts
- Failed login attempt lists
- Events ingested count over time
- Rules triggered
- Top domains visited
- Top source IPs, destination ports, active processes

6. Correlation Rules

This is one of the most important concepts to understand. Correlation rules are what transform raw logs into meaningful security alerts. They are logical expressions set to trigger when a specific condition or combination of conditions is met.

Let me give you real examples of correlation rules:

- **If a user gets 5 failed login attempts within 10 seconds raise an alert**: "Multiple Failed Login Attempts"
- **If login succeeds immediately after multiple failed attempts raise an alert:** "Successful Login After Multiple Failures"
- **If a USB device is plugged into any machine raise an alert** (especially if USB is restricted by company policy)
- **If outbound traffic exceeds 25 MB raise an alert:** "Potential Data Exfiltration Attempt"

> **Note: the SIEM only raises the alert. It is the security team, the firewall, or the IPS that takes blocking action. The SIEM is visibility, not enforcement.**

6.1 How to Build a Correlation Rule

When you build a correlation rule, you always need to identify two things first:

- **Log Source** which device or system is generating the event you want to watch? (e.g., Windows Event Log, a firewall, a web server)
- **Event ID** which specific Event ID corresponds to the activity you want to detect?

Here are two concrete use cases:

Use Case 1 Attacker Clears the Event Log

Attackers, after they compromise a system, almost always try to cover their tracks by clearing the event log. When someone clears the Windows event log, it generates Event ID 104. So the rule is:

> ***Rule: IF Log Source is WinEventLog AND EventID is 104 THEN trigger alert: "Event Log Cleared"***

Use Case 2 Whoami Execution After Exploitation

After gaining access to a system, attackers commonly run the whoami command to understand their privilege level. When any process runs on Windows, it generates Event ID 4688 with the new process name. So the rule is:

> ***Rule: IF Log Source is WinEventLog AND EventCode is 4688 AND NewProcessName contains "whoami" THEN trigger alert: "WHOAMI Command Execution Detected"***

if you have legitimate system administrators who run those same commands as part of their daily work, you do not want to trigger alerts on them. You use the log source filtering and exclusions to define that context.

7. Alert Investigation

When you are monitoring in the SOC and an alert fires, you do not immediately declare an incident. The analyst goes through an investigation process. The workflow is:

1. The alert triggers and shows up on the dashboard
2. The analyst opens the alert and examines the associated events and flows
3. The rule that fired is reviewed which conditions were met?
4. The analyst determines: is this a True Positive or a False Positive?

Depending on the outcome, actions include:

- False Alarm tune the rule to avoid similar false positives in the future
- True Positive perform further investigation
- Contact the asset owner to ask about the suspicious activity
- If confirmed malicious: isolate the infected host, block the suspicious IP

8. What Is Splunk?

Splunk is the dominant SIEM in the market. about 80 to 90 percent of organizations that have a SIEM are running Splunk. It is owned by Cisco now, which gives it heavy enterprise support and credibility. If you go for a SOC analyst interview, Splunk is almost certainly what they are running.

So what does Splunk actually do? It takes machine data that raw, unreadable stream of logs and events from all your devices and runs it through a pipeline:

- **Data Input** collects the raw data stream from any source, breaks it into 64K blocks, and adds metadata (hostname, source, encoding, destination index)
- **Data Parsing** breaks data into lines, identifies timestamps, creates individual events, annotates them with metadata keys
- **Data Indexing** stores events to disk and adds them to an index, enabling searchability
- **Data Search** enables users to query, view, and use event data; creates reports, dashboards, and alerts

9. Splunk Architecture

9.1 The Three Deployment Models

There are three ways to deploy Splunk depending on the size and needs of the organization:

Model 1: Standalone: one single machine running everything: searching, indexing, parsing, and data input. Good for testing or very small environments. The weakness is obvious: if that one machine fails, everything is gone.

Model 2: Basic Forwarding: you still have one central Splunk server, but you add Forwarders installed on endpoint machines that collect the logs and send them in. The server handles the rest. If one forwarder fails, you still have others.

Model 3: Distributed Deployment: the enterprise grade model. Forwarders, Indexers, and Search Heads each run on dedicated, separate machines. This gives you redundancy, load distribution, and high availability. If one indexer fails, others keep running.

> ***Important: All three deployment types are installed from the exact same Splunk package. It is the configuration that determines the role, not a different installer.***

9.2 Processing Components

Splunk has three core processing components:

- **Forwarder** the lightweight agent installed on endpoints to collect and forward logs to the indexer
- **Indexer receives** data, parses it into events, stores it to disk, and makes it searchable
- **Search Head** provides the user interface; users submit SPL (Search Processing Language) queries here, and it sends those queries to the indexers

9.3 Management Components

Supporting the processing components are several management layer components:

- Deployment Server the centralized configuration manager; pushes apps and config files (inputs.conf, outputs.conf, props.conf) to all forwarders
- License Server manages Splunk Enterprise licenses and license pools
- Cluster Master / Master Node manages and regulates indexer clusters, ensuring data replication and automatic failover
- Deployer distributes apps and configuration bundles to Search Head cluster members
- Monitoring Console provides detailed topology and performance information for your Splunk deployment

10. Splunk Forwarder Types

10.1 Universal Forwarder (UF)

The Universal Forwarder is the most common way to get data into Splunk. Its characteristics:

- Collects logs and sends them to a Splunk Enterprise indexer
- Very fast, minimal resource usage on the host
- Data is NOT parsed (raw/unparsed) the indexer does the parsing
- No web UI
- No Python support
- No license required
- Regarded as the best data collection method

10.2 Heavy Forwarder (HF)

The Heavy Forwarder is a full Splunk Enterprise instance that can also forward data. Use cases where you need one:

- Advanced routing sending different data to different indexers
- Masking sensitive data before indexing (e.g., credit card numbers)
- Running addons that receive data from external sources like databases (DBConnect) or HTTP (HTTP Event Collector)
- When a UI is needed Universal Forwarders have no UI

Key difference: the Heavy Forwarder PARSES data before sending it (cooked data). This means the indexer does NOT parse it again when it arrives. It requires a license.