HTTP Log Analysis Using Splunk SIEM

Overview

This project demonstrates the process of ingesting and analyzing HTTP access logs using Splunk SIEM to identify web server activity, extract traffic patterns, and detect potentially malicious behavior. The dataset contained **over 1.9 million events**, providing rich telemetry for behavioral and threat analysis.

Objective

- Upload and parse raw HTTP log files in Splunk
- Extract and analyze HTTP methods, URIs, response codes, and user agents
- Identify suspicious patterns, user agents, and potential exploit attempts
- Generate actionable insights from threat activity

Data Ingestion Summary

- **Data Source:** Sample HTTP access logs
- **Log Volume:** 1,939,615 events
- **Timeframe:** Logs analyzed up to 5/30/2025 2:40:44 PM
- **Sourcetype:** `access_combined`
- **Index Used:** `http_index`

Core Analyses

1. Top Requested URLs

SPL Query:

```
index=http_index sourcetype=access_combined
| top limit=10 uri
```

Key Finding:

- The URI `/servlet/admin` accounted for **48%** of all requests — likely an admin interface target for scanning or brute-force attempts.

---

2. Top HTTP Methods

SPL Query:

```
index=http_index sourcetype=access_combined
| stats count by method
```

Method	Count	Percentage
HEAD	509,600	58.23%
GET	206,961	23.65%
POST	155,314	17.75%
OPTIONS	2,075	<1%

Insight: A disproportionate use of the `HEAD` method (typically used for probing) suggests automated tools or scanners.

---

3. HTTP Status Codes

Status Code	Count

404	646,241
200	163,629
400	8,668
303	5,276
403	3,236

Insight: The high volume of `404 Not Found` and `403 Forbidden` responses may indicate scanning or enumeration attempts.

---

4. Top Source IP

IP Address	Count
192.168.203.63	503,641

This internal IP generated over **500K requests**, indicating possible misconfiguration, scanning activity, or automation.

Suspicious User Agent Detection

SPL Query:

```
index=http_index sourcetype=access_combined
| search user_agent="*sqlmap*" OR user_agent="*nmap*"
```

Events Matched: 9,617

Sample Suspicious Log Entry

Field	Value
Method	GET
URI	`/cwhp/auditLog.do?file=.............Program FilesCSCOpxlibclasspathcomcisco`

User-Agent	`Mozilla/5.0 (compatible; Nmap Scripting Engine; http://nmap.org/book/nse.html)`
Source IP	`192.168.202.79`
Target IP	`192.168.229.101`
Port	80

Breakdown:

- **User Agent:** Identifies the **Nmap Scripting Engine (NSE)** — a common reconnaissance tool.
- **URI:** Includes a **directory traversal** attempt targeting `CSCOpx`, CiscoWorks' Java classpath — indicative of an **exploit or information disclosure** attempt.
- **Implication:** Automated reconnaissance and likely exploit probing against network management software.

Threat Classification

- **Reconnaissance Activity:** Nmap NSE user agent confirms scripted probing
- **Exploit Attempt:** Directory traversal string targeting CiscoWorks suggests known vulnerability exploitation
- **Insider Risk or Compromised Host:** Internal IP as attacker source raises red flags

---

Recommended Actions

1. **Isolate and Investigate**: Host `192.168.202.79` should be reviewed for compromise.
2. **Correlate Logs**: Cross-reference activity with firewall, endpoint, and authentication logs.
3. **Alert SOC/IR Team**: If this is production infrastructure, this activity should trigger alerts.

4. **Create Detection Rules**:
    - Match against `user_agent="*nmap*"` or known recon tools
    - Monitor excessive 404/403 errors from single IPs
    - Detect directory traversal strings (`../`, overlong URIs)

Conclusion

This project highlights the value of HTTP log analysis for identifying reconnaissance and exploit activity in web environments. Splunk's powerful search and visualization capabilities enabled effective detection of both anomalous behavior and targeted attacks. This analysis can be expanded further by integrating with network logs, endpoint alerts, and threat intelligence.