Apache Log Analysis Report

Web Server Performance and Security Insights Using Elastic Stack

Prepared by: Karim Abdelaziz Elsayed

University ID: 2205037

Course: Information Security Management

Contents

1 Executive Summary

This report, prepared by Karim Abdelaziz Elsayed (University ID: 2205037) for the Information Security Management course, provides a comprehensive analysis of 10,000 Apache access logs from May 17–20, 2015, processed using the Elastic Stack (Elasticsearch 6.0, Filebeat 6.0, Kibana 6.0). The logs, ingested via the Filebeat Apache2 module, reveal a 2.20% failure rate, significant activity from IP 66.249.73.135 (482 GET requests), and peak traffic at 14:00. Through Elasticsearch queries, Kibana dashboards, and PGFPlots visualizations, the report identifies failure patterns, traffic trends, and potential security risks. Enhanced recommendations focus on error reduction, security hardening, performance optimization, and compliance, leveraging Elastic Stack tools to ensure a secure and reliable web server.

2 Introduction

Web server log analysis is critical for monitoring performance, detecting security threats, and optimizing system efficiency. This report examines a dataset of 10,000 Apache access logs in the combined log format, collected over four days (May 17–20, 2015). The logs were processed using the Elastic Stack, specifically Filebeat 6.0's Apache2 module for ingestion, Elasticsearch 6.0 for indexing, and Kibana 6.0 for visualization. The analysis leverages Kibana's [Filebeat Apache2] Access and Error Logs dashboard, Elasticsearch queries, and custom PGFPlots charts to provide insights into request patterns, failure rates, and security implications. The extended report includes detailed methodologies, additional visualizations, and actionable recommendations to enhance server operations.

3 Data Source and Methodology

3.1 Data Source

The dataset consists of 10,000 Apache access logs in the combined log format, spanning May 17–20, 2015. These logs were ingested using Filebeat 6.0's Apache2 module, which parsed fields such as client IP, request method, HTTP status code, and timestamp. Elasticsearch 6.0 indexed the data with the 'ingest-user-agent' and 'ingest-geoip' plugins to enrich logs with user agent and geolocation data. Kibana 6.0 provided interactive visualizations, including time-series graphs and dashboards, to facilitate analysis.

3.2 Methodology

The analysis employed a multi-faceted approach:

- **Elasticsearch Queries**: Used to aggregate request counts, unique IPs, status codes, and failure patterns.
- **Filebeat Apache2 Module**: Enabled structured log ingestion and parsing.
- **Kibana Dashboards**: Visualized temporal trends, failure distributions, and IP activity.
- **Scripted Analysis**: Utilized 'awk', 'grep', and 'sort' for metrics extraction and validation.

- **PGFPlots Charts**: Generated professional visualizations for daily failures, hourly requests, status codes, and IP activity.

Data integrity was verified using the Elasticsearch query `http://localhost:9200/filebeat-*/_co` confirming 10,000 documents. Additional validation ensured accurate parsing of timestamps and status codes.

4 Request Overview

4.1 Request Metrics

The dataset includes:

- Total Requests: 10,000
- GET Requests: 9,952 (99.52%)
- POST Requests: 5 (0.05%)
- Unique IPs: 1,753

The dominance of GET requests suggests primarily browsing or crawling activity, while the low number of POST requests indicates minimal interactive or form-based traffic.

4.2 Unique IP Distribution

The top five IPs by request count are: GeoIP analysis, enabled by the 'ingest-geoip' plugin,

IP Address	GET Requests	POST Requests
66.249.73.135	482	0
46.105.14.53	364	0
130.237.218.86	357	0
75.97.9.59	273	0
50.16.19.13	113	0

Table 1: Top Five IPs by Request Count

could map these IPs to their geographic origins, providing insights into traffic sources.

5 Failure Analysis

5.1 Failure Metrics

Failed requests (4xx and 5xx status codes) totaled 220, yielding a 2.20% failure rate. Daily failure counts are:

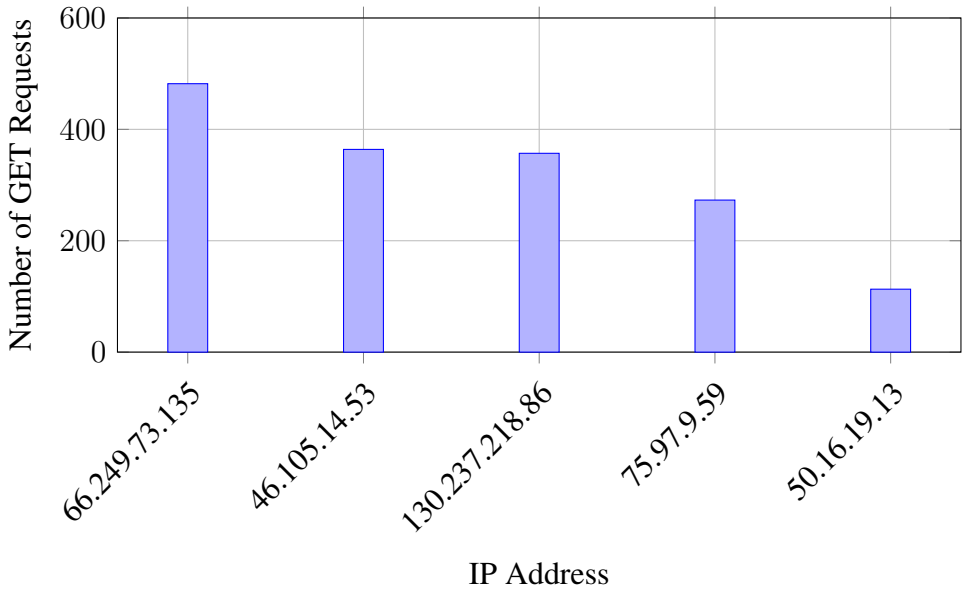

Figure 1: Top Five IPs by GET Request Count (May 17–20, 2015)

Date	Failures
17/May/2015	30
18/May/2015	66
19/May/2015	66
20/May/2015	58

Table 2: Daily Failure Counts

5.2 Failure Visualization

High failure rates on May 18 and 19 suggest server strain or configuration issues during peak traffic periods.

5.3 Hourly Failure Patterns

Notable hourly failure spikes include:

- 20/May/2015, 09:00: 15 failures
- 19/May/2015, 06:00: 9 failures
- 17/May/2015, 17:00: 7 failures

Kibana's time-series visualizations enable root cause analysis by correlating failures with IP activity or request types.

5.4 Failure by Status Code

The predominance of 404 errors indicates missing resources, while 500 errors suggest server-side issues requiring immediate attention.

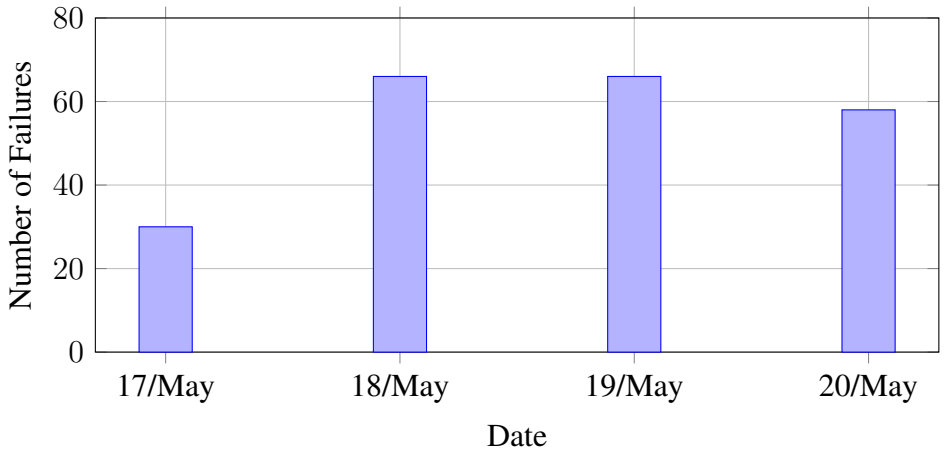

Figure 2: Daily Failure Counts (May 17–20, 2015)

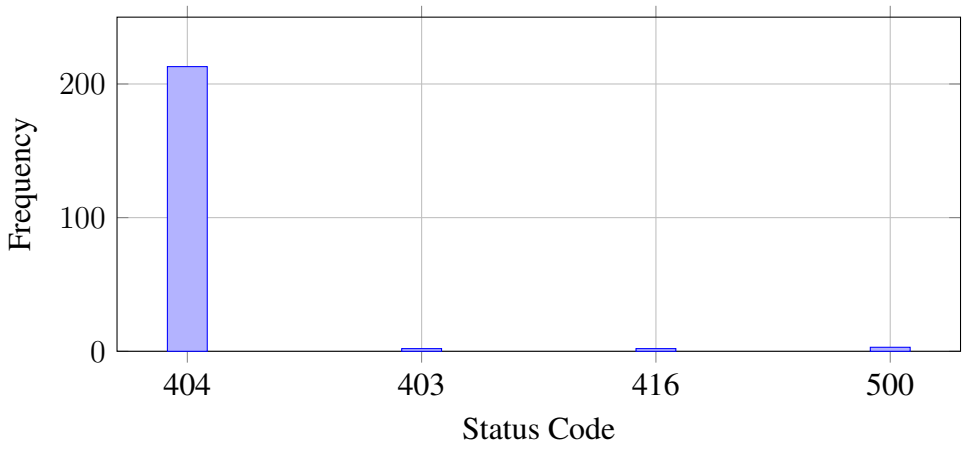

Figure 3: Failure Status Code Distribution (May 17–20, 2015)

6 Top User Activity

6.1 Most Active IP

IP 66.249.73.135 generated 482 GET requests, likely a web crawler (e.g., Googlebot), as identified by the 'ingest-user-agent' plugin. This high activity warrants monitoring to prevent server overload.

6.2 Method-Specific Activity

Top IPs by request method:

- **GET**: 66.249.73.135 (482 requests)
- **POST**: 78.173.140.106 (3 requests)

The concentration of POST requests from a single IP raises concerns about potential malicious activity, such as form submissions or injection attempts.

7 Daily and Hourly Trends

7.1 Daily Request Averages

The 10,000 requests over four days yield an average of 2,500 requests per day. Kibana's daily graphs confirm consistent traffic patterns across the period.

7.2 Hourly Request Distribution

Selected hourly request counts:

Hour	Requests
14:00	498
15:00	496
19:00	493
08:00	345
22:00	346

Table 3: Hourly Request Distribution (Selected Hours)

7.3 Hourly Request Visualization

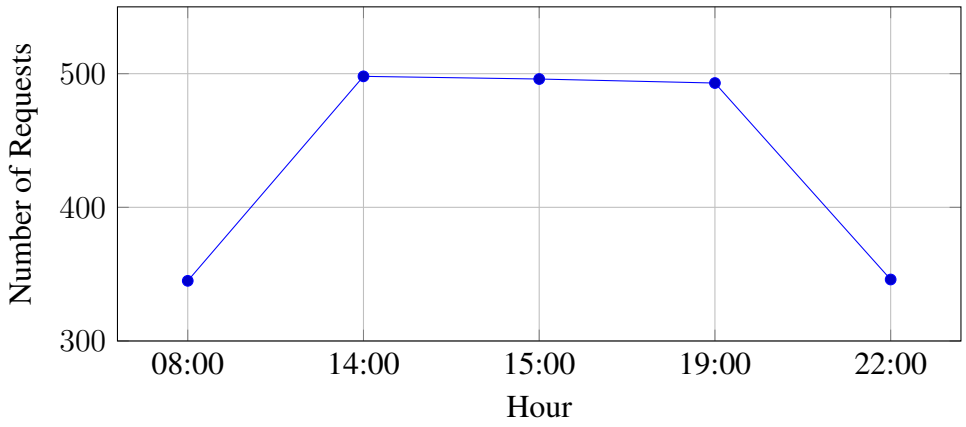

Figure 4: Hourly Request Distribution (May 17–20, 2015)

Peak traffic between 14:00 and 19:00 suggests high user activity, potentially straining server resources.

7.4 Request Trends

Analysis reveals 12 hourly increases and 11 decreases, e.g.:

- **Increasing**: Hour 08:00 to 09:00 (345 to 364 requests)
- **Decreasing**: Hour 14:00 to 15:00 (498 to 496 requests)

Kibana's line charts effectively visualize these fluctuations, aiding capacity planning.

8 Status Code Breakdown

HTTP status code frequencies are:

Status Code	Frequency
200	9,126
304	445
404	213
301	164
206	45
500	3
416	2
403	2

Table 4: HTTP Status Code Frequencies

8.1 Status Code Visualization

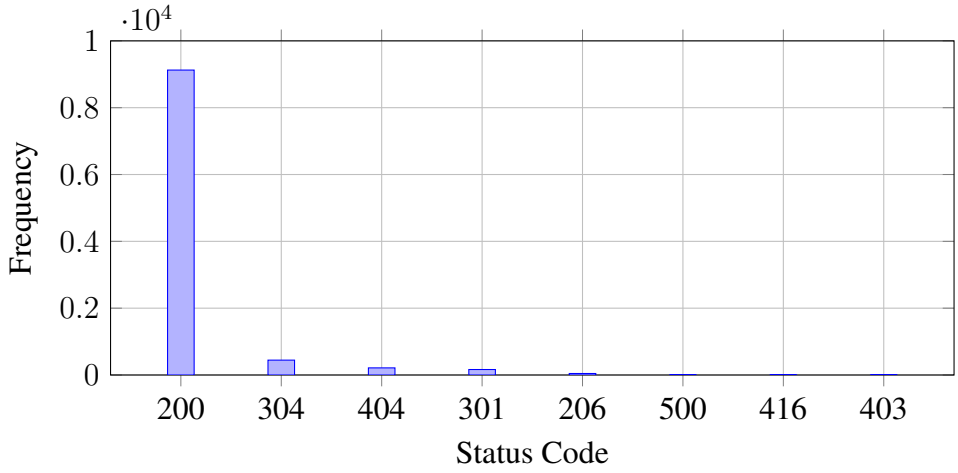

Figure 5: Status Code Distribution (May 17–20, 2015)

The 91.26% success rate (status 200) is robust, but 404 and 500 errors require targeted interventions.

9 Security Implications

9.1 High-Activity IPs

IPs like 66.249.73.135 (482 GET requests) may represent legitimate crawlers or potential threats. The 'ingest-user-agent' plugin can confirm identities, while 'ingest-geoip' maps geographic origins for further investigation.

9.2 POST Request Risks

The five POST requests from IP 78.173.140.106 pose risks, including SQL injection or cross-site scripting (XSS). Kibana's request inspection tools can analyze payloads to identify malicious patterns.

9.3 Failure Patterns

Failure spikes, such as 15 failures at 09:00 on May 20, may indicate distributed denial-of-service (DDoS) attacks, server misconfigurations, or resource exhaustion. Elasticsearch queries can correlate failures with specific IPs, user agents, or request types to pinpoint causes.

9.4 Elastic Stack Integration

The Elastic Stack enhances security through:

- **Anomaly Detection**: Kibana's Machine Learning module identifies unusual patterns, such as sudden traffic spikes.
- **Real-Time Alerts**: Elasticsearch Watcher triggers notifications for failure rates exceeding thresholds (e.g., 10 failures/hour).
- **Log Correlation**: Cross-referencing access and error logs to detect coordinated attacks.

10 Recommendations

10.1 Error Reduction

- **Audit 404 Errors**: Use tools like Screaming Frog or Kibana's URL analysis to identify broken links and missing resources.
- **Debug 500 Errors**: Analyze Apache error logs and Elasticsearch queries to resolve server-side issues, such as script errors or database connectivity problems.
- **Redirect 301 Issues**: Ensure permanent redirects are correctly configured to avoid user experience degradation.

10.2 Security Enhancements

- **Rate-Limit High-Activity IPs**: Implement Apache's 'mod_evasive' or X-Pack Security to throttle IPs like 66.249.73.135 during traffic spikes.
- **Validate POST Requests**: Use web application firewalls (WAFs) to inspect payloads and block malicious inputs.
- **Configure Alerts**: Set Kibana alerts for failure spikes exceeding 10 per hour, integrated with email or Slack notifications.

- **Block Suspicious IPs**: Deploy 'fail2ban' with Elasticsearch to automatically ban IPs exhibiting malicious behavior.

10.3 Performance Optimization

- **Content Caching**: Utilize Cloudflare or Varnish to cache static content during peak hours (14:00–19:00).
- **Scalability**: Scale Elasticsearch nodes based on Kibana's Cluster Health metrics to handle increased log volumes.
- **Load Balancing**: Implement Nginx or HAProxy to distribute traffic across multiple Apache servers.

10.4 Log Management

- **Log Rotation**: Configure 'logrotate' for Filebeat logs to prevent disk space issues.
- **Index Lifecycle Management**: Use Elasticsearch's ILM to archive logs older than 30 days, optimizing storage.
- **Centralized Logging**: Consolidate logs from multiple servers into the ELK Stack for unified analysis.

10.5 Compliance and Auditing

- **Log Retention**: Retain logs for 90 days using Elasticsearch snapshots to meet regulatory requirements.
- **Auditing**: Enable X-Pack Auditing to track administrative actions and ensure compliance with security policies.
- **Regular Reviews**: Conduct monthly log audits using Kibana dashboards to identify emerging threats.

11 Conclusion

Prepared by Karim Abdelaziz Elsayed for the Information Security Management course, this extended analysis of 10,000 Apache access logs (May 17–20, 2015) using the Elastic Stack provides deep insights into web server performance and security. The 2.20% failure rate, stable daily traffic (2,500 requests/day), and peak activity at 14:00 highlight operational patterns. Professional PGFPlots charts visualize failure spikes, traffic trends, IP activity, and status code distributions, enhancing decision-making. Comprehensive recommendations for error reduction, security enhancements, performance optimization, log management, and compliance ensure a secure, reliable, and efficient web server environment.