Log File Analysis Report

Sarah Ibrahim (2205223)

May 10, 2025

1 Introduction

This report details the analysis of a web server log file using a Bash script, as required for the Log File Analysis assignment. The goal was to extract key statistics, identify request and failure patterns, and provide actionable suggestions to improve system performance, reduce failures, and address potential security concerns.

2 Analysis Results

The Bash script processed the log file and produced the following statistics and insights:

2.1 Request Counts

- **Total Requests**: 10,000
- **GET Requests**: 9,952
- **POST Requests**: 5

2.2 Unique IP Addresses

- **Total Unique IPs**: 1,753
- **Most Active IP (Overall and GET Requests)**: 66.249.73.135 (482 requests)
- **Most Active IP (POST Requests)**: 378.173.140.106

2.3 Failure Requests

- **Total Failed Requests**: 220
- **Failure Percentage**: 2.20%
- **Status Code Breakdown**:
  - 200: 9,126,200
  - 404: 213,404

  - 403: 164,301
  - 500: 45,206
  - Others: 7,319

2.4 Daily Request Averages

- **Average Requests per Day**: 2,500

2.5 Failure Analysis

- **Top 5 Days with Most Failures**:
  1. 19/May/2015: 66 failures
  2. 18/May/2015: 66 failures
  3. 20/May/2015: 58 failures
  4. 17/May/2015: 30 failures

2.6 Request Trends

- **Requests per Hour**: Ranged from 34,508 to 49,814, with significant increases during 14:00–20:00.
- **Failures by Hour**: Peaked at 1,809, with notable occurrences in the afternoon.

3 Suggestions Based on Analysis

Based on the analysis, the following suggestions are proposed to enhance system reliability, optimize performance, and mitigate potential issues:

1. **Reducing Failures (404, 403, and 500 Errors)**:
   - Implement regular link validation to address the high volume of 404 errors (213,404 occurrences), ensuring broken links are fixed promptly.
   - Investigate root causes of 403 (164,301) and 500 (45,206) errors, such as permission issues or server misconfigurations, and apply targeted fixes.
   - Set up automated monitoring to detect and alert administrators of failure spikes, particularly on high-failure days like 18–20 May 2015.
2. **Managing High-Traffic Days and Times**:
   - Scale server resources during afternoon hours (14:00–20:00) to accommodate traffic spikes (up to 49,814 requests) and reduce failures (e.g., 1,809 in one hour).

   - Optimize caching mechanisms to improve response times during peak periods.
   - Analyze high-failure days (e.g., 19/May/2015 and 18/May/2015, both with 66 failures) to identify specific issues like server overload or configuration errors.

3. **Addressing Security Concerns and Anomalies**:
   - Monitor the high activity from IP 66.249.73.135 (482 requests), likely a bot (e.g., Googlebot). If this traffic is undesired, implement rate limiting to reduce server load.
   - The low number of POST requests (5) suggests limited form submissions. Establish anomaly detection to flag unexpected increases in POST activity, which could indicate malicious behavior.
   - Regularly audit the 1,753 unique IPs to detect unusual patterns, such as rapid requests from a single IP, which may signal potential security threats.
4. **Improving System and User Experience**:
   - Develop custom 404 pages to guide users back to valid content, improving navigation and user satisfaction.
   - Enhance server capacity planning based on the average daily request rate (2,500 requests/day) and peak hourly trends.
   - Continuously monitor request and failure patterns to proactively address emerging issues, ensuring long-term system reliability.

4 Conclusion

The log file analysis provided valuable insights into request patterns, failure trends, and IP activity. High 404 errors, concentrated bot activity, and afternoon traffic spikes indicate areas for improvement. By implementing the proposed suggestions—such as link validation, resource scaling, rate limiting, and anomaly detection—the system can achieve improved performance, reduced failures, and enhanced security. Ongoing monitoring and analysis are recommended to assess the impact of these changes and identify new opportunities for optimization.