Information Security

Task 3

Name: Sara Ayman Abdelbassir
ID: 2205129

Log File Analysis Report

Introduction

This report presents a detailed analysis of the access log file. The goal is to uncover key metrics including request types, IP activity, server errors, and usage patterns. The analysis was conducted using a Bash script that processes the log file line-by-line and extracts statistical insights.

1. General Statistics

- Total Requests: 10,000
- GET Requests: 9,952
- POST Requests: 5
- Unique IP Addresses: 1,753

2. Top 5 IPs by GET/POST Requests

Rank	IP Address	Request Type	Count
1	66.249.73.135	GET	482
2	46.105.14.53	GET	364
3	130.237.218.86	GET	357
4	75.97.9.59	GET	273
5	50.16.19.13	GET	113

3. Failed Requests

- Total Failures (4xx/5xx): 220
- Failure Percentage: 2.20%

4. Most Active IP

- IP Address: 66.249.73.135
- Requests: 482

5. Average Requests Per Day

- Average: 2,500 requests/day

6. Top Days with Failed Requests

Date	Failed Requests
19/May/2015	66
18/May/2015	66
20/May/2015	58
17/May/2015	30

7. Requests Per Hour

Hour	Requests
00:00	361
01:00	360
02:00	365
03:00	354
04:00	355
05:00	371
06:00	366

Hour	Requests
07:00	357
08:00	345
09:00	364
10:00	443
11:00	459
12:00	462
13:00	475
14:00	498
15:00	496
16:00	473
17:00	484
18:00	478
19:00	493
20:00	486
21:00	453
22:00	346
23:00	356

8.Status Code Breakdown

Status Code	Count
200	9,126
304	445
404	213
301	164
206	45
500	3

Status Code	Count
416	2
403	2

9. Top IPs by Request Type

Top 3 by GET Requests:

- 66.249.73.135 → 482
- 46.105.14.53 → 364
- 30.237.218.86 → 357

Top 3 by POST Requests:

- 78.173.140.106 → 3
- 91.236.74.121 → 1
- 37.115.186.244 → 1

10. Failure Patterns by Hour

Hour	Failures
09:00	18
05:00	15
06:00	14
17:00	12
13:00	12

Conclusion

This analysis reveals normal traffic with a high volume of GET requests and a small percentage of failed attempts (mostly 404 errors). The most active IP address is 66.249.73.135. There's a consistent request load across all hours, with slightly more failures occurring in the morning. These insights can help improve server performance and target monitoring efforts more effectively.

Analysis Suggestions:

Reducing the Number of Failures:

Failures Analysis: A significant number of requests are failing, particularly those with 4xx (client errors) and 5xx (server errors) status codes. Based on the analysis, 220 requests (2.20%) resulted in failure.

Recommendation: Investigate the root cause of these failures. For 4xx errors, check for issues like broken links, incorrect URLs, or permission errors. For 5xx errors, the server may be experiencing issues, such as internal server errors or overloads. Implementing better error handling and monitoring systems would help minimize these failures.

Days and Times Requiring Attention:

Top Days with Failures: The analysis shows that certain days, such as 19/May/2015, 18/May/2015, and 20/May/2015, had the highest number of failed requests.

Recommendation: These days should be reviewed further to understand if there was a server issue, heavy traffic, or a misconfiguration that led to increased failures. Identifying these trends will help anticipate future issues and allocate resources accordingly.

Security Concerns and Anomalies:

IP Analysis: A particular IP, 66.249.73.135, made a high number of requests (482 GET requests), which stands out as potentially suspicious. This could be a bot or crawler.

Recommendation: Implement rate limiting or security measures such as CAPTCHA or bot detection systems to mitigate excessive requests from a single IP. Furthermore, analyze any high-traffic IPs that generate large amounts of GET or POST requests to ensure no malicious activity is taking place.

Improving System or Service:

Request Patterns: The analysis shows that requests tend to peak at certain hours, with the highest volume occurring at 10:00 AM, 11:00 AM, and 12:00 PM.

Recommendation: During these peak times, consider scaling your server resources to handle the increased load and reduce the risk of server failures or slowdowns. Also, ensure your caching strategy is optimized for high-traffic hours to improve the system’s performance.

Improving User Experience:

GET vs. POST Requests: The data shows a dominant number of GET requests (9952) compared to POST requests (5).

Recommendation: This suggests that most of the activity involves retrieving data rather than sending data. You can optimize the website's content delivery for faster loading times, especially for GET requests. Additionally, ensure that POST requests are efficiently processed, particularly if they are used for form submissions or data updates.

Trends in Request Times:

Requests by Hour: Requests appear to peak at 00:00 and 01:00, which might indicate heavy usage during late hours.

Recommendation: If this trend persists, ensure that the server is adequately scaled to handle high traffic during these hours. Analyze the types of requests during these periods to determine if additional resources or optimizations are needed.

Status Code Breakdown:

Recommendation: The majority of the requests are returning a status code of 200 (successful), which indicates that most requests are being processed successfully. However, 404 (Not Found) and 301 (Moved Permanently) errors are notable. It would be good to review the server logs for any misconfigurations or broken links that could lead to 404 errors.