Analysis of NASA HTTP Access Log

Assignment Report

Computer Science
Alexandria University
Information Security

Submitted by

Mohammed Abdulrahman Awad Khaled
ID: 2205114

May 2025
Academic Year: 2024 - 2025

1. Introduction

This report analyzes the NASA HTTP access logs from July 1995 using a custom Bash script. The goal is to explore traffic trends, identify failure patterns, and suggest optimization and security improvements.

2. Key Metrics

- **Total Requests:** 1,891,714

- **GET Requests:** 1,887,646 (99.78%)

- **POST Requests:** 111 (0.02%)

- **Unique IPs:** 81,983

- **Most Active IP:** `piweba3y.prodigy.com` (17,572 GETs)

- **Failed Requests:** 44,234 (2.34%)

3. Request Trends

Hourly request counts reveal periods of increasing or decreasing traffic. Notable patterns include:

- Increasing trends at hours 04, 08, 11, 14, 16, 18, 21, and 23.

- Decreasing trends around 01–03, 05–07, and 12–13.

- Highest activity observed in late afternoon and evening.

This helps identify peak usage hours for scaling and optimization.

4. Failure Patterns by Hour

Failure requests (4xx/5xx) begin appearing from hour 10 onwards:

- **Hour 15:** 837 failures

- **Hour 14:** 750 failures

- **Hour 11:** 731 failures

- Hours before 10 had 0 failures.

Failures rise during peak hours, indicating possible overload or misconfigurations.

5. Interpretation

- The system is highly read-oriented (almost all requests are GET).

- A few IPs dominate traffic, which could indicate scraping or bots.

- Request volume follows a daily rhythm, useful for performance tuning.

- Error spikes suggest investigating server health in afternoon hours.

6. Recommendations

- **Optimize performance** via CDN, caching, and load balancing.

- **Enhance security** by monitoring POST requests and high-frequency IPs.

- **Fix broken links** and investigate server errors to reduce failure rates.

- **Future work:** IP geolocation and user-agent analysis.