IEUK Engineering Task 2025

Harry Foster

July 16, 2025

1 Findings

The analysis of the provided log data revealed that approximately **8.3%** of all requests were deemed suspicious. The IP address with the highest volume of requests, `45.133.1.1`, was found to frequently change its User-Agent string—behavior commonly associated with bots.

By identifying requests using well-known automation tools such as `Scrapy` and `Selenium`, and then flagging all associated IPs, it was found that approximately **5%** of total requests originated from IPs linked to bot-like behavior.

Another detection method involved filtering for IP addresses with exceptionally high request frequencies. Overall, this approach revealed that most detected bots originated from IP addresses based in Russia and the United Kingdom, with smaller proportions from the United States, Canada, and North Korea.

2 Assumptions

The analysis assumes that requests with non-standard User-Agent strings—such as `curl` or `wget`—are likely generated by automated scripts rather than by legitimate users. These tools are not typically used by browsers or official applications. In contrast, User-Agent strings resembling `Mozilla/5.0` are assumed to be more likely legitimate, though spoofing is still possible.

It is also assumed that normal users rarely exceed **25 API requests per minute**. IPs generating sustained high request volumes are considered suspicious, based on the assumption that no legitimate use case requires such frequency. This forms the basis for flagging such IPs as potentially abusive.

3 Proposed Solutions

A practical and low-cost mitigation strategy is the implementation of **API rate limiting**. Setting a limit of **25 requests per minute** could block approximately **6%** of suspicious traffic. This can be configured using `nginx` for self-hosted environments, or via rate-limiting tools provided by third-party hosting services.

Another effective measure is **User-Agent filtering**. About **2%** of all requests contained clearly suspicious User-Agent strings. Blocking such requests can significantly reduce automated or scripted abuse with minimal risk of affecting legitimate users.

Lastly, unnecessary or vulnerable API endpoints should be disabled to reduce the attack surface. Examples of exposed endpoints include `/api/version/test` and `/api/internal/logs/1`.