**IEUK Report – Log Analysis**

Following the recent growth in popularity of the company’s podcast and newsletter, the website has seen a significant increase in incoming traffic. While this has brought a rise in subscriptions, it has also led to recurring server downtime. With only a small engineering team in place, these outages are affecting both productivity and user experience.

To investigate the root cause, I analysed a set of server logs to determine whether the traffic is legitimate customers or artificially inflated by bots; and to propose realistic solutions to manage the traffic and a potential increase in traffic as the company gains popularity.

The strategy that I developed to tackle this problem was to write a python script that highlights any suspicious traffic from the server logs, such as suspicious user agents; high frequency of requests in a short time frame; empty referrer headers (which I will be ignoring for this task as all logs provided have empty referrer headers); and extremely low response times. My script would then return the total number of suspected bot requests, the total number of requests, and the percentage of requests that are likely to be from a bot.

After writing my script, I ran it with the provided sample-log.log file to test it. Using that file, my program returned the following results:

* Total number of bot-like requests: 5,667
* Total number of requests: 432,096
* Percentage of bot-like requests: 1.31%

At first glance, these numbers appear to be very low and thus indicate that a very low amount of traffic for the website is generated by bots. However, given the short time frame I had to complete this task, and the lack of skills and knowledge I have to create a script to distinguish bot requests from human requests, it is more likely that my script doesn’t account for all bot requests, just the obvious ones.

While my findings strongly suggest that non-human activity is not an issue for this company, if they later find evidence to reject my findings and that there *is* in fact a large amount of non-human traffic, then I would suggest the company to do the following:

* Utilising Captcha to verify users’ humanity
* Utilising tools like Cloudflare to block bots
* Caching static resources (such as images) to reduce server load
* Scaling the server vertically (i.e. upgrading CPU, RAM, storage, etc) and horizontally (i.e. sharing the traffic across multiple servers), though I appreciate this suggestion may not be feasible due to the company’s small budget.
* Block specific IP addresses with suspicious user-agents.