|  |  |
| --- | --- |
| Why Non-Human Traffic Is Bad | Arkose Labs  **Engineering Skills Project – Report**  For IEUK 2025 | Abstract  About a small music startup suffering from abnormal traffic issues and website downtimes.  By Alesha Sangar |

Engineering Skills Project – Report

To make my program work, download the sample log file on Github first: [**https://github.com/brightnetwork/ieuk-task-2025**](https://github.com/brightnetwork/ieuk-task-2025)

Introduction:

I was hired by a small music media startup facing website crashes from a traffic surge due to successful podcasts and newsletters. With only three engineers and no monitoring tools, the site was vulnerable. My role was to analyse traffic, confirm bot activity, and recommend low-cost solutions. I created a Python program to examine server logs, which revealed that most traffic was non-human and caused repeated downtime.

Assumptions regarding the business:

The startup faces resource limitations and lacks:

* **Bot protection systems**
* **Monitoring software**
* **Sufficient technical staffing**

Prompt action is vital to prevent further disruption.

Findings

Log analysis confirmed most traffic was from bots. These bots displayed distinct traits:

* **High request rates** far above human browsing patterns
* **Identifiable user-agents** like curl, python-requests, Googlebot
* **Frequent errors** (404/500) caused by malformed or excessive requests

While search engine bots are generally helpful, others degrade performance or scrape data. I developed a program to flag suspicious IPs based on these behaviours.

Recommendations

To manage traffic without extra cost, I propose four simple, effective actions:

1. **Rate Limiting** – Restrict excessive requests per IP.  
   *Tools: NGINX, Apache, Cloudflare.*  
   *Cost: Free.*
2. **CAPTCHA** – Block bots on forms and login pages.  
   *Tools: reCAPTCHA, hCaptcha.*  
   *Cost: Free.*
3. **Basic Bot Detection** – Identify non-human traffic via user-agents and session patterns.  
   *Tools: FingerprintJS, custom scripts.*  
   *Cost: Free.*
4. **Traffic Monitoring** – Log and analyse patterns regularly.  
   *Tools: GoAccess, UptimeRobot.*  
   *Cost: Free.*

Conclusion

With basic detection, manual oversight, and simple technical solutions, the company can reduce bot-related issues affordably. Ongoing monitoring is essential. Short-term, I recommend hiring additional engineers to respond quickly if problems return.

Bibliography

* <https://www.cloudflare.com/en-gb/learning/bots/what-is-bot-traffic/>
* <https://datadome.co/guides/bot-protection/tools/>

[**Code to my project**](https://d.docs.live.net/3055c667723f6e88/Important%20things/Work%20(Engineering)/IEUK%202025/Skills%20Project/ieuk-task-2025-main/Solution%20code%20(Notepad).txt)