Sector Skills Project Report – Engineering

There is evidence to suggest non-human activity in the site’s user traffic. Assuming that there is a custom (non-standard) format for the log data, I ran my code in a Docker container to flag data that occurs too many times: within the 4-day period, some IP addresses were used as many as 5,400 times, and some pages (HTTP paths) were visited around 15,000 times. This would be unlikely to occur naturally for a business of this size. Even if some pages were especially popular, the same IP address should not make this many requests and could be a sign of a denial of service (DoS) attack.

One potential solution would be to place a restriction on requests from the same IP address, such as limiting an IP address to 50 requests per day. This could be a quick, easy way to circumvent the problem but may impede user experience for users with VPNs. Another solution could add a CAPTCHA, requiring users to prove they are human before loading their first page. The free and standard tiers would allow 100,000 monthly assessments for £5.97 ($8), but the log file suggests 100,000+ daily assessments, at a higher-tier cost approaching £0.75 ($1) per 1,000 assessments (these prices were calculated by converting from USD). One final solution could rate-limit pages which are visited too frequently; this would help against a DDoS attack if many different IP addresses make requests but may inconvenience genuine users.

The ideal solution would likely involve a blacklist to restrict frequent IP addresses from sending requests and frequent pages from being visited, since these address the most obvious symptoms. Although it would sometimes be inconvenient for users, slower service or minimal outages are preferable to full site downtime, and is cost-effective for a small company compared to using external providers for services like CAPTCHAs and 2FA. This will not account all anomalies like high bounce rate or high session duration, but it’s adequate to fix most problems in the short-term.

References:

Writer, S. (2023). *How to Distinguish Bot vs. Human Traffic | Arkose Labs*. [online] Arkose Labs. Available at: [https://www.arkoselabs.com/blog/blog-how-to-distinguish-bot-vs-human-traffic/](https://www.arkoselabs.com/blog/blog-how-to-distinguish-bot-vs-human-traffic/%20) [Accessed 17 Jul. 2025].

‌

Cloudflare.com. (2024). *What is bot traffic? | How to stop bot traffic*. [online] Available at: <https://www.cloudflare.com/en-gb/learning/bots/what-is-bot-traffic/>.

NGINX Documentation. (n.d.). *NGINX Docs | Configuring Logging*. [online] Available at: <https://docs.nginx.com/nginx/admin-guide/monitoring/logging/>

‌Apache.org. (2011). *Log Files - Apache HTTP Server Version 2.4*. [online] Available at: <https://httpd.apache.org/docs/2.4/logs.html>.

Docker (2019). *Get Started, Part 1: Orientation and setup*. [online] Docker Documentation. Available at: <https://docs.docker.com/get-started/>.

Google Cloud. (2025). *Compare features between reCAPTCHA tiers*. [online] Available at: <https://cloud.google.com/recaptcha/docs/compare-tiers>.

‌

Twilio. (n.d.). *Verify Pricing*. [online] Available at: <https://www.twilio.com/en-us/verify/pricing>.

Radware (2024). *Bot Protection: Attack Examples & 8 Ways to Defend Your Network | Radware*. [online] Radware.com. Available at: <https://www.radware.com/cyberpedia/bot-management/bot-protection/>.

‌

‌