EXECUTIVE BRIEFING: ORIGIN SERVER BOT INTELLIGENCE

Target: tygartmedia.com · **Scope:** 55-Day Comprehensive Nginx Origin Log Audit (2,838,206 Requests) · **Date:** November 2026

Origin Request Census: 2,838,206 Total Requests · **2,137,787 Page Hits (75.3%)** · 700,419 Assets (24.7%) · **1,203,889 Masked Scrapers (42.4%)** · **531,349 AI Bots (18.7%)** · **146,669 Search (5.2%)** · 142,615 Single-Hit IPs

THE FIVE CRITICAL FINDINGS

1. **The Client-Side Analytics Blind Spot (>99.7% Invisible):** Client-side tools (Henkan Vision) saw 68 ignored visits in 7 days (63 bots); origin logs recorded 23,030 single-hit IPs across a near-equal 8d 2h 15m slice (142,615 in 55 days). Automated scrapers fetch raw HTML and disconnect immediately without firing JavaScript beacons.
2. **The Masked Browser Scraper Majority (42.4%):** 1,203,889 hits forge desktop Chrome/Edge User-Agents on Windows/Mac/Linux across dense `/24` IPv4 prefixes, showing a 77.0% combined 1-page bounce rate (80.0% Win, 78.2% Mac, 69.2% Linux) while evading basic UA blocks.
3. **Crawlers vs. Agents (Architectural Split):** Crawlers (LinkupBot, ClaudeBot, bingbot) traverse site hierarchies with low bounce rates (7.0–27.9%; 72.1–93.0% multi-page). AI retrieval agents (GPTBot, MetaBot, ExaSearchBot) execute single-page deep links with 82.0–93.4% bounce rates.
4. **The 4-Hour AI Discovery Race:** Measuring up to 500 published articles confirms AI answer engines index new articles within hours: **PerplexityBot (4.4h, n=480)**, **MetaBot (4.4h, n=152)**, **GPTBot (5.0h, n=488)**, and **ClaudeBot (5.5h, n=499)**. AI ingestion is a continuous real-time loop.
5. **Probe Radiation & Security Verification:** Logged 48,164 hits on `/wp-login.php` (17,082 internal origin `HEAD` checks from `8.229.250.67`; 31,082 external probes across 10,746 IPs, 16,805 POSTs) and 798 `/.env` probes. Audit of 22 HTTP 200 `/wp-config.php` hits confirmed **0 payload bytes sent and 0 secrets leaked**.

KEY AUTONOMOUS ACTORS COMPARISON

Actor / Bot	Operator	Classification	Total Hits	Bounce Rate	Discovery Lag	Primary Behavioral Signature
ExaSearchBot	Exa Labs	Neural Search	91,607	93.4%	11.4 hrs (n=496)	Prompt-driven single-page neural fetcher
LinkupBot	Linkup SAS	Live Web API	73,435	7.0%	22.0 hrs (n=485)	Disciplined, low-frequency sitemap sweeper
GPTBot	OpenAI	Training & Search	72,246	82.0%	5.0 hrs (n=488)	Rapid discovery; heavily re-crawls Claude pricing
MetaBot	Meta Platforms	Social / Llama AI	70,988	87.0%	4.4 hrs (n=152)	Distributed pool; instantaneous viral headline fetcher
ClaudeBot	Anthropic	Knowledge Ingestion	55,550	10.2%	5.5 hrs (n=499)	Methodical, polite, multi-page hierarchical walker
PerplexityBot	Perplexity AI	Live Answer Engine	46,100	34.2%	4.4 hrs (n=480)	Tied fastest AI lag (4.4h); prompt-triggered live citations
bingbot	Microsoft	Bing & Copilot	98,890	27.9%	N/A	High-volume index maintenance across entire site

THREE IMMEDIATE OPERATIONAL PRIORITIES

1. Origin-Harden Login & Probes
Use nginx deny/444 rules, `POST limit_req` (429), and `block-phtml.conf` (410) to stop 31,082 external login hits (of 48,164 raw) before PHP.

2. Harden Root PHP Directives
Add `location ~* (wp-config|xmlrpc|\.env) { deny all; return 404; }` in Nginx to guarantee zero PHP booting on sensitive targets.

3. Deploy Origin FastCGI Caching
Use nginx `fastcgi_cache` (up to 2h TTL) at origin to absorb 531,349 AI hits without PHP-FPM spikes during breaking news coverage.