EXECUTIVE BRIEFING · PRODUCTION AUDIT

Not All Bots Are the Same

55 Days of Server Logs, 2.84 Million Requests, and the Realities of Machine Traffic

Target Site: tygartmedia.com

Audit Window: Sep 18 – Nov 12, 2026 (55 Days)

Total Ingestion: 2,838,206 Requests (56 Log Files)

Prepared For: Executive Leadership & Operations

02 The Client-Side Analytics Blind Spot

TELEMETRY GAP

Client-side analytics tools like Google Analytics or Henkan Vision are blind to over 99.7% of automated machine traffic.

WHAT THE DASHBOARD SAW (7-DAY WINDOW)

68

TOTAL IGNORED VISITS

- **63 Classified Bots:** Only clients executing JavaScript and loading pixels.
- **5 No-Consent Sessions:** Opted out or blocked telemetry.
- **False Security:** Gives management the illusion that bot traffic is negligible.

WHAT THE NGINX ORIGIN LOGS RECORDED

23,030

SINGLE-HIT MACHINE IPs (8d 2h 15m SLICE) · 142,615 (55 DAYS)

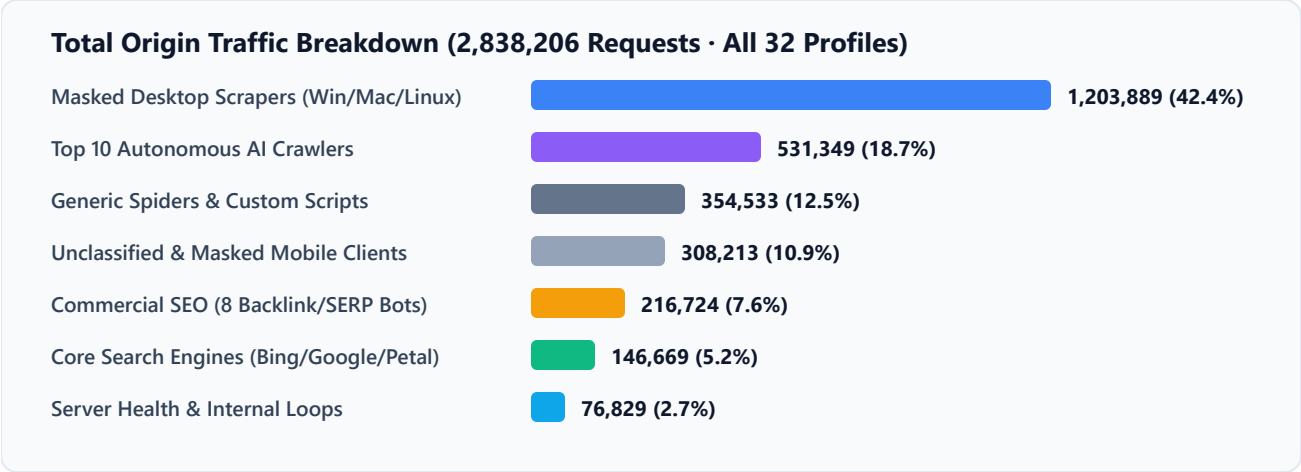
- **359,112 Total HTML Page Hits:** All HTML requests handled by Nginx in 8d 2h 15m (Nov 4–12).
- **Immediate HTML-Only Exit:** Bots fetch HTML and close connections without executing JS tags.
- **~365x Discrepancy (Approximate Window):** Origin logs 23,030 single-hit IPs in 8.1 days vs. 63 bot visits in Henkan's 7-day window.

03 The Traffic Census: 2.84 Million Requests

MACRO BREAKDOWN

HTML document requests represent 75.3% of origin traffic—an inverted ratio proving machine-dominated consumption.

2,838,206	2,137,787	700,419
TOTAL ORIGIN REQUESTS	HTML PAGE REQUESTS (75.3%)	STATIC ASSETS (24.7%)



04 The Rise of the AI Bots: 531,349 Requests

AI CRAWLERS

The Top 10 autonomous AI bots drive 4.5x more traffic than Googlebot & Bingbot combined (117,454) and 3.6x more than all Core Search (146,669).

Bot Name	Operator	Classification	Total Hits	Unique Paths	Peak RPM	Bounce Rate
ExaSearchBot	Exa Labs	Neural Search & Embeddings	**91,607**	8,909	74	93.4%
LinkupBot	Linkup SAS	Enterprise AI Search	**73,435**	4,949	22	7.0%
GPTBot	OpenAI	Model Training & Search	**72,246**	18,853	143	82.0%
MetaBot	Meta Platforms	Social Preview & Llama AI	**70,988**	2,467	61	87.0%
ShapBot	Shap	Content Aggregator	**61,461**	13,840	1,044	22.7%
ClaudeBot	Anthropic	Knowledge Retrieval	**55,550**	17,844	226	10.2%
PerplexityBot	Perplexity AI	Live Answer Engine	**46,100**	8,653	118	34.2%
Applebot	Apple Inc.	Siri Knowledge Graph	**30,572**	11,212	16	82.6%
Amazonbot	Amazon	Alexa & Search Indexing	**22,553**	4,095	41	96.8%
Bytespider	ByteDance	AI & Search Indexing	**6,837**	1,188	45	58.0%

05 Not All Bots Are the Same: Crawlers vs. Agents

BEHAVIORAL TAXONOMY

Automated traffic has bifurcated into two completely different architectural models: Tree-Walkers vs Surgical Fetchers.

THE CRAWLER: SITEMAP TREE-WALKER

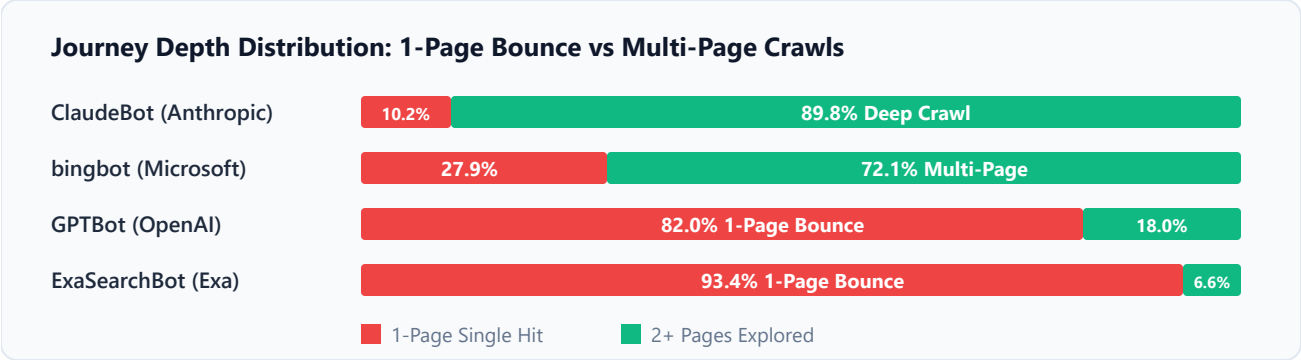
ClaudeBot · bingbot · LinkupBot

- **Low Bounce (7.0% – 27.9%):** Explores deep topic hierarchies.
- **Multi-Page Depth:** 72.1% to 93.0% of sessions explore 2+ pages.
- **Objective:** Comprehensive index construction & graph mapping.
- **Robots Behavior:** Steady multi-page traversal with `/robots.txt` checks.

THE AGENT: SURGICAL DEEP-LINK FETCHER

ExaSearchBot · MetaBot · GPTBot

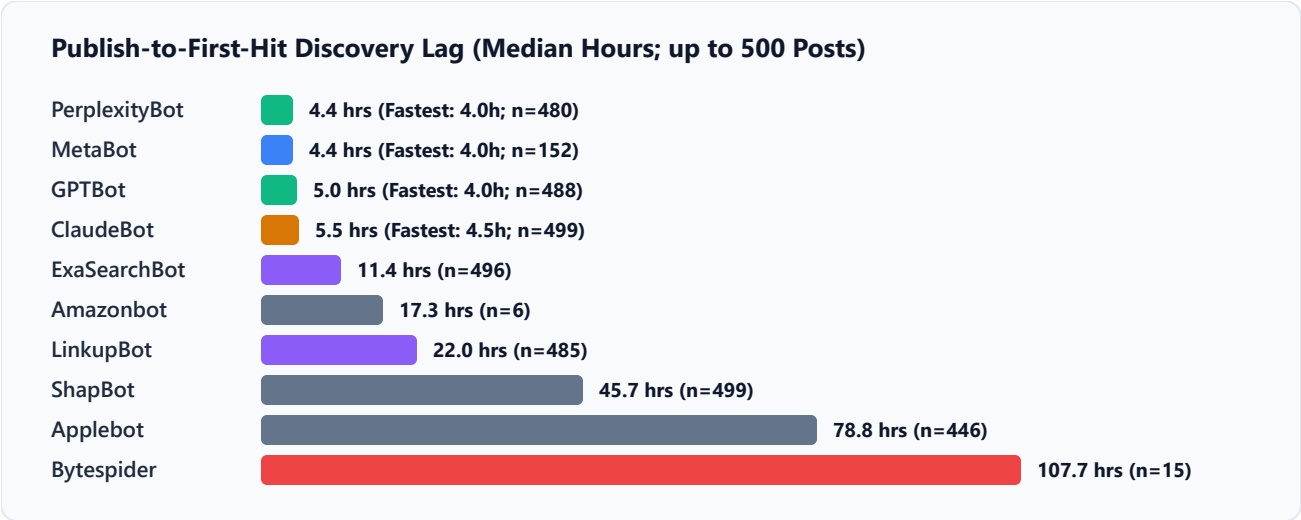
- **Extreme Bounce (82.0% – 93.4%):** Single page requested and gone.
- **Minimal Depth:** Only 6.6% to 18.0% advance past a single page.
- **Objective:** Answering immediate user prompts & live RAG grounding.
- **Robots Adherence:** High concurrency bursts from distributed IPs.



06 The Discovery Race: Publish to First Hit

INGESTION SPEED

Evaluation of up to 500 published articles (matched sample sizes n=6 to n=499) shows leading AI answer engines arrive within 4.4 to 5.5 hours.



Key Operational Implication: Content changes and breaking news propagate to AI answer engines (Perplexity, Meta, OpenAI) on the exact same business day. Origin HTML caching (such as nginx FastCGI caching) must accommodate instant revalidation.

07 What Bots Read vs. What Humans Read

CONTENT CONSUMPTION

While human readers consume regional Pacific Northwest news, automated agents are obsessed with Claude AI pricing.

HUMAN VISITOR PROFILE

Local Journalism & Business Trends

- Port of Tacoma logistics and industrial real estate.
- Regional Pacific Northwest infrastructure guides.
- Everett and Snohomish county business expansions.
- Evenly dispersed reading across editorial categories.

BOT AGENT PROFILE (SINGLE-HIT PRICING ANOMALY)

High-Frequency Single-Hit Scraping of LLM Pricing

- **/claude-ai-pricing/:** 1,268 single-hit scraper requests.
- **/claude-student-discount/:** 1,228 single-hit scraper requests.
- **/claude-pricing-tiers-compared.../:** 973 single-hit scraper requests.
- **/claude-rate-limits/:** 950 single-hit scraper requests.

Why? AI answer engines must continuously update token costs and subscription limits to avoid hallucinating stale rates in user citations.

08 The Phantom Fleet: 1,203,889 Masked Desktop Requests

STEALTH TRAFFIC

42.4% of origin traffic disguises itself under desktop Chrome/Edge User-Agents on Windows, Mac, or Linux.

613,606	317,686	272,597
MASKED WINDOWS DESKTOP (80.0% BOUNCE)	MASKED MAC DESKTOP (78.2% BOUNCE)	MASKED LINUX DESKTOP (69.2% BOUNCE)

FORENSIC SIGNATURES OF MASKED SCRAPERS

1. HTML-Skewed Traffic: 747,173 page hits (62.1% HTML) alongside selective `/favicon.ico` and plugin asset pulls.

2. Dense /24 IPv4 Clustering: Up to 255 single-hit IPs pack individual `/24` prefixes across the 142,615 single-hit pool.

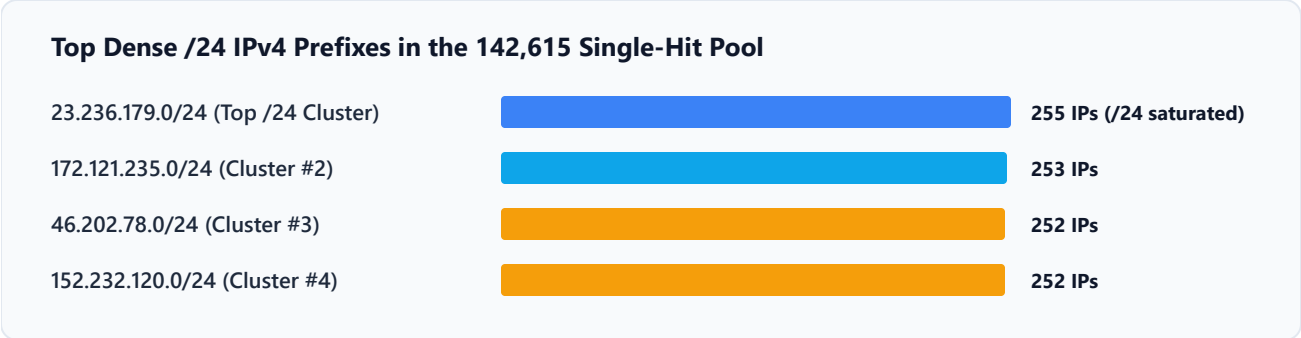
3. Large IP Rotation Pools: 172,428 Windows, 123,874 Mac, and 22,655 Linux IPs forge desktop browser UAs (77.0% bounce).

4. Zero JS Telemetry Execution: Clients disconnect after HTTP retrieval without firing client-side analytics beacons.

09 Dense /24 Prefix Clustering: 142,615 Single-Hit IPs

INFRASTRUCTURE AUDIT

Scrapers cycle through entire /24 IP blocks to bypass traditional per-IP rate limiting controls.



IPv4 /24 Prefix	Cluster Signal	Unique Single-Hit IPs	Observed Behavioral Profile
23.236.179.0/24	Saturated /24 Rotation	**255 IPs** (Entire /24 saturated)	Single-hit page extraction under forged desktop UAs.
172.121.235.0/24	High-Density /24 Pool	**253 IPs**	Distributed single-hit scraper cluster.
46.202.78.0/24	High-Density /24 Pool	**252 IPs**	High-concurrency single-page content harvester.
152.232.120.0/24	High-Density /24 Pool	**252 IPs**	Rotating single-hit article collector.

10 The Probe Layer: Background Radiation of the Web

SECURITY TELEMETRY

Automated scanners continuously probe for exposed credentials, CMS vulnerabilities, and staff directories.

48,164
/wp-login.php (31,082 EXT · 17,082 ORIGIN)

798
/.env SECRET PROBES (100% BLOCKED)

8,202
CORPORATE/CONTACT PROBES (10 SLUGS; 7,315 EXCL. /contact-us)

RECONCILING IP 8.229.250.67 VS. EXTERNAL PROBES

17,082 hits from `8.229.250.67` are internal origin `HEAD` healthchecks, not a foreign attacker. Actual external exposure is **31,082 distributed hits** across **10,746 IPs** (**16,805 POSTs**). Origin nginx hardening and rate-limits stopped compromise, yet external probes still tax PHP workers and DB connections.

11 Security Deep Dive: The /wp-config.php Audit

INCIDENT FORENSICS

Forensic verification of 22 HTTP 200 responses confirms zero credential or source code leakage.

THE ALARM

22 Hits Returned HTTP 200 in September

- Security scanners flag any 200 response on `wp-config.php` as critical severity (database credential leakage).
- Occurred during early server configuration prior to explicit drop rules.

THE FORENSIC REALITY

PHP Executed Empty: 0 Payload Bytes Sent

- Nginx logs show `body_bytes_sent` of 5 to 31 bytes (standard protocol headers only).
- Because `wp-config.php` defines PHP constants without `echo` or HTML markup, the interpreter returned an empty string.
- **Verdict: Zero credentials or code were transmitted.** Hardened with explicit 404/deny rules.

```
location ~* (wp-config\.php|xmlrpc\.php|\.env) { deny all; return 404; }
```

12 The AI Overlap Map: Universal Core vs. Exclusives

INDEX CONSENSUS

Out of 32,020 unique URLs requested by AI crawlers, only 60 are indexed by all 10 bots, while 42.1% are requested by only one.

60	5,316	13,164	13,480
UNIVERSAL CORE (ALL 10 BOTS)	CONSENSUS CORE (5-9 BOTS · 16.6%)	MODERATE OVERLAP (2-4 BOTS · 41.1%)	SINGLE-BOT EXCLUSIVES (1 BOT)

WHY DO AI BOTS SPECIALIZE ON EXCLUSIVE URLS?

- **ShapBot (7,844 Exclusives):** Requests non-trailing-slash `301` article variants (48.4% of its hits) alongside plugin CSS/JS assets (`contact-form-7`, `fluentform`).
- **GPTBot (2,367 Exclusives):** Follows user citation links into historical technical documentation.
- **ClaudeBot (1,827 Exclusives):** Methodically indexes author bios, category indexes, and relational taxonomies.

13 Traffic Attribution: Who Are They?

FORENSIC ATTRIBUTION

Classifying origin traffic strictly from log-observable User-Agent, URL, and /24 prefix signatures (no external rDNS/ASN lookups).

Traffic Cluster	Confidence	Forensic Evidence & Behavioral Signature
Microsoft Bing & Copilot	[DECLARED UA]	Declared `bingbot/2.0` UA (98,890 hits, 2,130 IPs; 27.9% bounce). External rDNS/ASN validation not executed in this log-only pass.
Googlebot Search	[DECLARED UA]	Declared `Googlebot/2.1` UA (18,564 hits, 4,884 IPs; 70.4% bounce). External rDNS/ASN validation not executed in this log-only pass.
Corporate Directory Probers	[INFERRED]	**Automated directory enumeration.** Systematic probing of `/team`, `/people`, `/staff`, and 7 sibling slugs. Single IP `187.77.23.88` hit 771 times on each of 10 slugs (7,710 total probes).
Dense-/24 Masked Scrapers	[INFERRED]	**Saturated single-hit /24 prefixes** (`23.236.179.0/24`: 255 IPs; `172.121.235.0/24`: 253 IPs; `46.202.78.0/24` & `152.232.120.0/24`: 252 IPs) forging desktop Chrome/Edge UAs (77.0% 1-page bounce).

14 Six Concrete Operational Recommendations

ACTION PLAN

Immediate architectural enhancements to reduce origin load, harden security, and optimize for AI search.

01

Origin-Harden Login & Probes

Use nginx deny/444 rules, `POST limit_req` (429), and `block-phtml.conf` (410) to cut 31,082 external login hits (of 48,164 raw).

02

Harden Nginx Against Sensitive PHP

Explicitly deny `wp-config.php` and `.env` in Nginx. Guarantees zero credential transmission.

03

Origin AI Micro-Caching

Use nginx `fastcgi_cache` (up to 2h TTL) for anonymous HTML. Absorbs 531,349 Top-10 AI hits at origin before PHP-FPM.

04

Adaptive Rate Limiting

Cap aggressive scrapers (e.g. ShapBot at 1,044 RPM) to 15 req/sec while letting polite crawlers pass.

05

Drop Scanner Slugs at Nginx

Return static 404 or 444 for `/team` and `/people`. Eliminates expensive database 404 lookups.

06

Unify Server & Tag Telemetry

Ingest Nginx log summaries into BI pipelines to eliminate the >99.7% blind spot in capacity planning.

STRATEGIC SUMMARY

The Modern Web is Machine-First

Managing web infrastructure in 2026 requires looking past client-side marketing dashboards. By understanding the distinct travel patterns of autonomous agents, hardening origin endpoints, and optimizing origin caching for rapid AI ingestion, organizations can secure their infrastructure while maximizing their visibility in the era of artificial intelligence.

Tygart Media Operations Intelligence · White Paper & Slide Deck Available in Repository