LOG ANALYSIS, AI OBSERVABILITY, AND STATE-LEVEL RECONNAISSANCE

A Case Study in Asymmetric Intelligence

Prepared for: Sematext Group
Contact: Sean Monahan
Prepared by: John Beach (Riverside County, California)
Date: 24 December 2025
Reference: Sematext Logs App "Website Traffic Analysis" (ID 75147)

Introduction

About 7 May, a dormant website under a basic Namecheap shared hosting plan was reactivated, and placeholder content was uploaded that consisted of the text of a dystopian short story that had been completed a few weeks earlier, in mid-April. The short story, “The Foundation,” had been produced during a six-week test of the creative writing abilities of the five primary AIs: Google Gemini, Microsoft Copilot, ChatGPT, Claude, and Perplexity.

The purpose for reactivation was delayed, and in early September, when the administrator happened to look at the Apache website traffic logs, he saw an unusually high number of requests. To understand the reason for the anomaly, he developed the AI-assisted “Log Analysis System,” that leverages the immense analytical capabilities of AI, in coordination with a protocol for daily website monitoring and management.

(The website supervisory protocol utilizes the Internet UIs of Cloudflare, Sematext, Google Analytics, and Google Search Console. The Log Analysis System itself is defined by 10 simple interconnected MD files, that have been designated “open source” and will not be copyrighted.)

The dormant website had become the target of a coordinated, state-level reconnaissance operation that took the politically sensitive, dystopian placeholder content seriously. The complete forensic chain, from the first human visitor on 31 July, thought to be an agent of Chinese Intelligence, to validation that same day by the Russian FSB, and then continuing for three months until 26 October, the last day of the study, is preserved via the Apache server raw access logs.

[Note: The initial contact with Sematext, ID 75147, is unrelated to the substance of this report. The inability to utilize the Sematext website UI was resolved by migrating from the basic Namecheap website shared hosting service to the VPS Pulsar, w/cPanel, service, which provides root access.]

From ~7 May to 31 July

Low level of bot activity through the end of June. Routine benevolent bots, minor Internet noise. Beginning in July, activity increases steadily. A spike in activity from 26 through 29 July, and then a drop until the morning of 31 July.

31 July - Four Confirmed Human Actors Appear

On 31 July, ongoing automated surveillance of the Internet by Chinese Intelligence uncovered the dormant website that hosted the dystopian short story, “The Foundation.” The Chinese immediately notified their allied service, the Russian FSB, and coordinated the initial assessment operations. (All times UTC.)

1. *“San Jose Door-Kicker” - 12:17 UTC (08:17 -0400)*

Role: Tier-1 Verification Scout

Tier-1 Definition: Initial rapid assessment by first-level analyst to confirm target is real, active, and substantive. Task: Read title, scan opening content, classify site type, verify authenticity, log findings, authorize escalation to specialist review.

Behavioral Signature: 22-second engagement, rapid professional triage.

Attribution: Chinese intelligence technical verification (timezone convenient for Chinese operations, San Jose location).

2. *“Oasis Subject Matter Expert (SME)” - ~14:48 UTC (10:48 -0400, approximate)*

Role: Tier-2 Deep Analyst

Tier-2 Definition: Detailed specialist analysis following Tier-1 authorization. Task: Examine complete narrative structure across all five sections, test whether 1984-sequel/Revelation content conceals operational codes, cipher systems, or represents genuine ideological threat. Decision: Authorize content extraction and offline committee review.

Behavioral Signature: 94 seconds (1m34s), 19 distinct events examining all story sections.

Attribution: Chinese intelligence senior analyst (hours after San Jose verification, depth suggests specialist evaluation).

[Note: Oasis is a tiny agricultural area on the northwest shore of the Salton Sea, in Riverside County. This IP geolocation indicates that the SME is physically located in the western part of the Colorado Desert, which is the general setting of the fictional protagonist in the dystopian short story.]

3. *“Moses Lake Vacuum Cleaner” - 19:34 UTC (15:34 -0400)*

Role: Automated Structural Mapper

Technical Profile: Human-controlled tool executing 10+ HTTP HEAD requests in ~1 second, testing paths: /wordpress, /wp, /backup, /old, /new, /main, /home.

Behavioral Signature: Multi-page navigation (2+ pages) triggering GA4 "engaged" status, but 0-second engagement time.

Attribution: Chinese intelligence technical operator (3 hours after Oasis, likely different operator or specialized role).

4. *“From Russia, with Love” - 22:34 UTC (18:34 -0400)*

Role: FSB Archive Verification

IP/Location: 167.160.72.51, Chicago, Illinois (Sprious LLC VPN exit node).

Target: Attempted to access `/products/adolf-putins-rogues-gallery-1` (2022 anti-Putin content URL, returned 404).

Operational Significance: FSB searched internal archives following allied notification from Chinese intelligence, found 2022 "Adolf Putin's Rogues Gallery" domain entry, attempted old URL structure to verify continuity. Professional tradecraft: routing through US-based VPN rather than direct Russian IP.

Attribution: Russian FSB (target behavior reveals intent despite geographic masking).

Month of August - Black Hats Conduct Multi-Phase Evaluation

Analysis of the period from 31 July through 31 August (4837 requests total) reveals seven distinct coordinated traffic peaks that indicate a sophisticated, human-directed evaluation process rather than random bot activity. This month represents the transition from field discovery to structured committee assessment.

Week 1 - Initial Assessment Phase

1 August (381 requests): Immediate verification following Discovery Day field collection.

5 August (446 requests): First major review cycle (highest volume single day).

7 August (285 requests): Follow-up assessment or secondary committee review.

Week 2 - Deep Evaluation Phase

9 August (253 requests): Continued coordinated evaluation.

10 August (375 requests): Second major review cycle (second-highest volume).

Week 3 - Extended Analysis Phase

17 August (257 requests): Third major review cycle (one week after 10 August).

21 August (213 requests): Fourth major review cycle.

Final Week - Conclusion Phase

27 to 30 August: sharp decline (140 requests total).

31 August: minimal activity (12 requests). Clear operational wind-down pattern.

Analysis of August Patterns

The patterns suggest sophisticated organizational structure, rather than simple, two-stage review:

- Weekly Coordination Cycle: Major peaks occur at approximately 5 to 7 day intervals (5, 10, 17, 21 August), suggesting structured committee meeting schedules, weekly reporting cycles to higher authority, and distributed evaluation teams operating on a synchronized calendar.
- Escalating Evaluation Levels: The progression from 5 August (446) through 21 August (213), with declining peak volumes, suggests initial broad committee review (5 through 10 August, highest volumes), escalation to specialized analysts or senior review bodies (17 through 21 August, moderate volumes), and narrowing focus as assessment questions were resolved and conclusions reached.
- Geographic Distribution Implications: The sustained elevated activity, across four weeks, suggests multiple committees or working groups operating in parallel, possibly separate, Chinese and Russian review cycles, and different analytic disciplines (technical, linguistic, ideological, strategic), reviewing in sequence, with cross-organizational coordination, requiring a multi-week timeline.

What the August Patterns Reveal

The four-week sustained evaluation, with seven coordinated peaks, indicates a formal intelligence assessment process, with structured operational phases:

- Week 1 (1 through 7 August): Rapid technical assessment. Content extraction verification, technical analysis, initial threat classification. Three coordination events.
- Week 2 (9 and 10 August): Distributed committee review. Multiple analytic teams reviewing extracted materials in parallel. Two major coordination events.
- Week 3 (17 August): Escalated to specialist analysis. Higher-level review requiring specialized domain expertise (Revelation eschatology, Orwell scholarship, Trump political dynamics). Single major coordination event.
- Week 4 (21 August): Final approval and reporting. Senior review body or cross-organizational coordination, recommendation approval, resource allocation decisions. Final major coordination event.
- Week 4 closeout (27 through 31 August): Administrative wind-down, archival, target reclassification.

The Iceberg Beneath the Water Line

The seven peaks with declining volumes support the “offline review committee” hypothesis, and suggest greater organizational complexity than might initially be estimated.

Field Collection (31 July) - 4 visible human actors extracting content

- “San Jose Door-Kicker” - Tier-1 verification.
- “Oasis Subject Matter Expert (SME)” - Tier-2 deep examination.
- “Moses Lake Vacuum Cleaner” - Automated structural mapping.
- “From Russia, with Love” - FSB Archive Verification.

Initial Review Committee (1 thru 10 August) - 10 to 15 analysts (estimate)

- Technical analysis team (code/cipher detection, server infrastructure assessment).
- Linguistic/content analysis team (narrative evaluation, ideological threat assessment).
- Attribution team (author identification, historical continuity verification).

Specialized/Escalated Review (17 August) - 5 to 8 senior analysts (estimate)

- Domain experts (Revelation eschatology, Orwell scholarship, Trump political dynamics).
- Strategic assessment (threat level, response recommendations, monitoring requirements).
- Cross-organizational coordination (if Chinese-Russian joint assessment).

Final Approval Authority (21 August) - 2 to 5 decision-makers (estimate)

- Senior leadership review.
- Recommendation approval.

- Resource allocation for any follow-on activities.
- Target reclassification authority.

Total Estimated Personnel: 20 to 30+ individuals involved across the four-week evaluation, with only 4 field collectors to leave web analytics traces. The remaining 16 to 26 conducted their analyses on secured internal systems using extracted content, producing the coordinated temporal signatures without direct site access.

Black Hats Hard at Work - All While Under an Asymmetric Microscope

The complete August pattern indicates a professional intelligence assessment process, with formal structure, multi-week timeline, distributed responsibilities, and senior decision authority involvement. This was not an opportunistic reconnaissance, but a systematic evaluation that followed established organizational procedures.

The seven coordination peaks represent scheduled review events, reporting deadlines, escalation points, and approval milestones, within a structured analytical process. The declining volumes from 5 August (446) through 21 August (213) suggest progressive refinement: a broad initial review that narrows to specialized analysis, and then to final senior approval.

By late August, assessment operations concluded. The sharp decline in final week activity (27 thru 31 August), combined with minimal September activity before defensive countermeasures were implemented, suggests the evaluation produced definitive conclusions. Recommendations were delivered to decision-makers and the target was reclassified, based on assessment findings rather than defensive hardening.

Month of September - New Countermeasures and the “Black Hat Surge”

Defensive countermeasures on 14 September and 18 September faced an adversary who had already completed collection and evaluation operations. The subsequent 22 through 24 September “black hat testing surge” represented reconnaissance of the new defensive posture, not a continuation of content assessment activities.

Phase 1 - .htaccess (14 September)

The first defensive layer, deployed on 14 September, was a robust Apache .htaccess configuration that implemented path whitelisting, hidden file denial, and content integrity rules. This layer showed characteristic signs of first-iteration deployment under time pressure, i.e., it was "robust but faulty."

- Observable Signature: First HTTP 403 "Forbidden" responses appeared 14 September, targeting `.git/HEAD` vulnerability probes. Prior to this date, all requests received normal processing (200/301/404).
- Initial Performance: Low initial blocking rate (1 to 2 daily), but the appearance of 403s marked a clear defensive posture shift visible to reconnaissance actors.

Phase 2 - Cloudflare Shield (18 September)

The second defensive layer was deployed on 18 September: Cloudflare edge-first architecture with Browser Integrity Checks, and distributed filtering across 330+ global data centers.

- Observable Signature: HTTP 403 responses exploded to 91 occurrences on 18 September (a 15-fold increase), as Cloudflare edge filtering combined with refined .htaccess rules.

- Infrastructure Impact: A single reconnaissance IP from Microsoft Azure infrastructure, probing with massive PHP webshell enumeration, received systematic blocking. Post-deployment stabilization showed a 403 blocking rate sustained at 30-40% of total traffic.

The Empire Strikes Back (22 through 24 September)

The black hats responded to defensive hardening with a coordinated three-day surge, to map the new defensive posture:

- 22 September: 570 total requests, 195 blocked (34%)
- 23 September: 562 total requests
- 24 September: 482 total requests

This surge represented systematic adversary assessment: mapping blocked versus accessible paths, probing for bypass techniques, and evaluating the cost of future operations. Primary probe targets included `.git/config` (6 attempts), WordPress vulnerability paths, and PHP webshell enumeration (`xmrlpc.php`, `info.php`, `moon.php`, `nc4.php`).

This sequence demonstrates authentic security practice: 1) initial defensive response to threat recognition, 2) imperfect first implementation requiring refinement, and 3) secondary defensive layer deployment. Despite the "robust but faulty" and "imperfect" characterizations, the combined deployment significantly increased operational costs for the adversary, and informed the eventual reclassification of the target.

A Quiet October

By 1 October, the black hats had finished their evaluation of the website. Defensive countermeasures taken in mid-September made it impractical to continue new probing.

Residual Bot Evolution (2 thru 26 October)

Following the withdrawal of human collectors, automated infrastructure continued with increasing intensity. This pattern indicates that while the human "offline review committee" had concluded its evaluation, the site remained on automated monitoring lists for opportunistic scanning.

- 15 October: 861 requests.
- 25 October: 1266 requests (highest one-day count in the entire 90-day dataset).

The intensification of automated attacks despite the lack of human direction highlights the persistent nature of global reconnaissance infrastructure, that continues to probe targets long after the primary intelligence mission has concluded.

Three-Month Summary: Asymmetric Observability Advantage

The full forensic chain, from the 31 July discovery by Chinese Intelligence to the 26 October study conclusion, proves that defenders with modest resources can achieve an asymmetric observability advantage. By utilizing cross-platform correlation and governed AI-assisted analytical frameworks, the defender was able to preserve state-level reconnaissance tradecraft at the moment of execution. This documentation was possible only because routine hosting evaluation *most fortuitously* captured the intelligence collection window in its entirety.

Strategic Implications of the Log Analysis System

Although individual web requests cannot definitively distinguish pure automated bots from human actors, pattern analysis resolves this ambiguity. Random bot activity maintains a steady baseline, whereas coordinated human activity produces discrete temporal peaks. This study utilizes temporal pattern analysis to identify organizational intent behind ambiguous technical signatures.

The Visibility Stack

The forensic chain relied on synthesizing a unified "Visibility Stack" to overcome the limitations of individual platforms. When monitored in coordination, these tools allow a non-professional to track state-level tradecraft:

- Cloudflare (Edge Defense): Blocks the vast majority of hostile traffic upstream, and identifies global attack distribution.
- Sematext (Structured Retention & Automation): The essential "retention layer" for retrospective analysis, and the primary automated engine. Via Logagent, provides an almost real-time data flow.
- Google Analytics 4 (Behavioral Intelligence): Isolates human visitors from automated tools via the Engaged Sessions threshold.
- Google Search Console (Discovery Tracking): Monitors intent and identifies the specific queries or indexing events that initially "flagged" a site during an Internet search.

By integrating these four core services and their user interfaces into a governed routine, the website administrator achieves the correlation necessary to separate human actors from automated "ghost" tools. It then becomes possible for an intelligence analyst to understand the observed human behavior, and to infer motives.

The Human-in-the-Loop Discriminator

While modern reconnaissance uses headless browsers to mimic user behavior, Engagement Time serves as the primary discriminator. Automated tools typically complete inspections and exit within 0 to 2 seconds. In contrast, the 22-second (San Jose) and 94-second (Oasis) signatures captured on 31 July represent human triage and analysis windows, rather than automated execution.

The Offline Review Committee Hypothesis

The technical evidence confirms that visible actors function as field collectors rather than the final evaluation body. Standard intelligence tradecraft dictates a "Collect Once, Evaluate Many" model to minimize the footprint on the target system. While the committee members review extracted content on secured internal systems, periodic follow-up passes to verify live-site continuity create the temporal signatures identified as the "digital exhaust" of an internal meeting cycle.

Sematext Platform Integration and Strategic Benefits

A critical advance in the overall website monitoring and management architecture is the utilization of Sematext to automate the data pipeline. Although Namecheap remains the formal source of Apache raw access logs, it now serves as a static, manual backup archive. Sematext, via Logagent, assumes the active, automated role within the Log Analysis System and its associated daily work routines. This integration provides a near real-time, steady stream of raw access data, which, with Sematext's various storage options, effectively obviates the need for manual log updates.

Conclusion: The Asymmetric Intelligence Advantage

The ultimate benefit of the open-source “Log Analysis System” is the ability to analyze and document a complete operation lifecycle from the defender's perspective. In this particular case, the enemy forces operated under the assumption of invisibility, but the defender achieved an asymmetric advantage. The four confirmed field actors - the “San Jose Door-Kicker,” the “Oasis Subject Matter Expert,” the “Moses Lake Vacuum Cleaner,” and “From Russia, with Love” (via Chicago) - were revealed to be agents for a much larger analytical body, proving that aggregate patterns can reveal intent that individual data points seek to hide.

Attachments:

1. Log Analysis System Documentation Package (ten interdependent MD files)
2. "The Foundation" (AI-assisted dystopian short story, 7232 words)

[2710 words, 8 legal-size pages]