Indicator of Benignity:
An Industry View of False Positive in Malicious Domain Detection and its Mitigation

Daiping Liu, Palo Alto Networks
Danyu Sun, University of California, Irvine
Zhenhua Chen, Palo Alto Networks
Shu Wang, Palo Alto Networks
Zhou Li, University of California, Irvine

February 2026

False Positives: The Overlooked Drag on Security Operations

Analyzing 6-year FPs (2019 ~ 2024) reported to Palo Alto Networks.

Operational Scale

- Protects over 65,000 organizations globally
- Analyzes ~7 billion DNS queries per day
- Detects 1.6 million new malicious domains daily

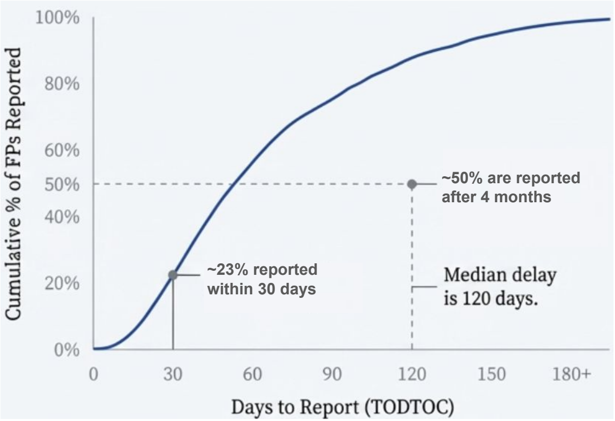
FP Report Scale

- Received 123,491 FP reports over 6 years
- Confirmed 121,073 to be true FPs (~98%)

The Hidden Cost: These FP incidents can lead to significant financial and operational impact for users.

FPs Are a Slow-Burning, Long-Tail Problem

FPs have a long reporting delay



Implication: Evaluating a detector's real-world FP rate requires long-term deployment of ~4 months.

The problem is highly fragmented

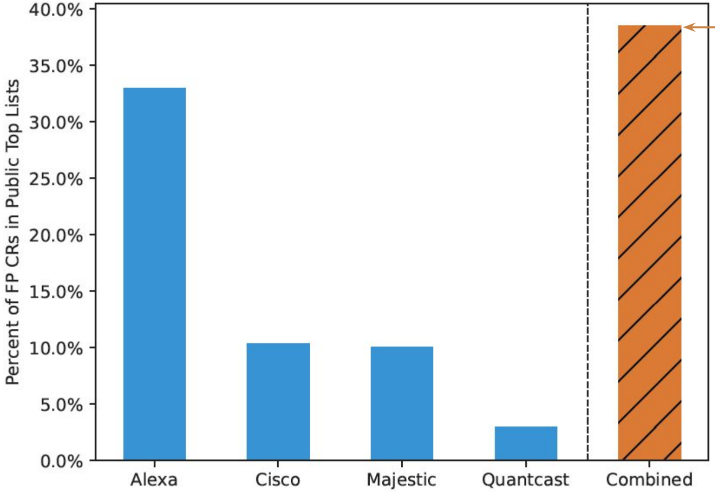
"97.7% of false positive FQDNs were reported only once, by a single user."

"Most FPs occur under unique root domains. 123,491 FP reports came from 113,738 unique root domains."

Implication: FP mitigation needs to be generic and scalable; grouping by domain or user is not an effective strategy.

The Industry's and Academia's Go-To Solution is Failing

Relying on popularity-based top domain lists (*e.g.*, Tranco) to prevent FPs is fundamentally inadequate.



~38% Maximum Coverage

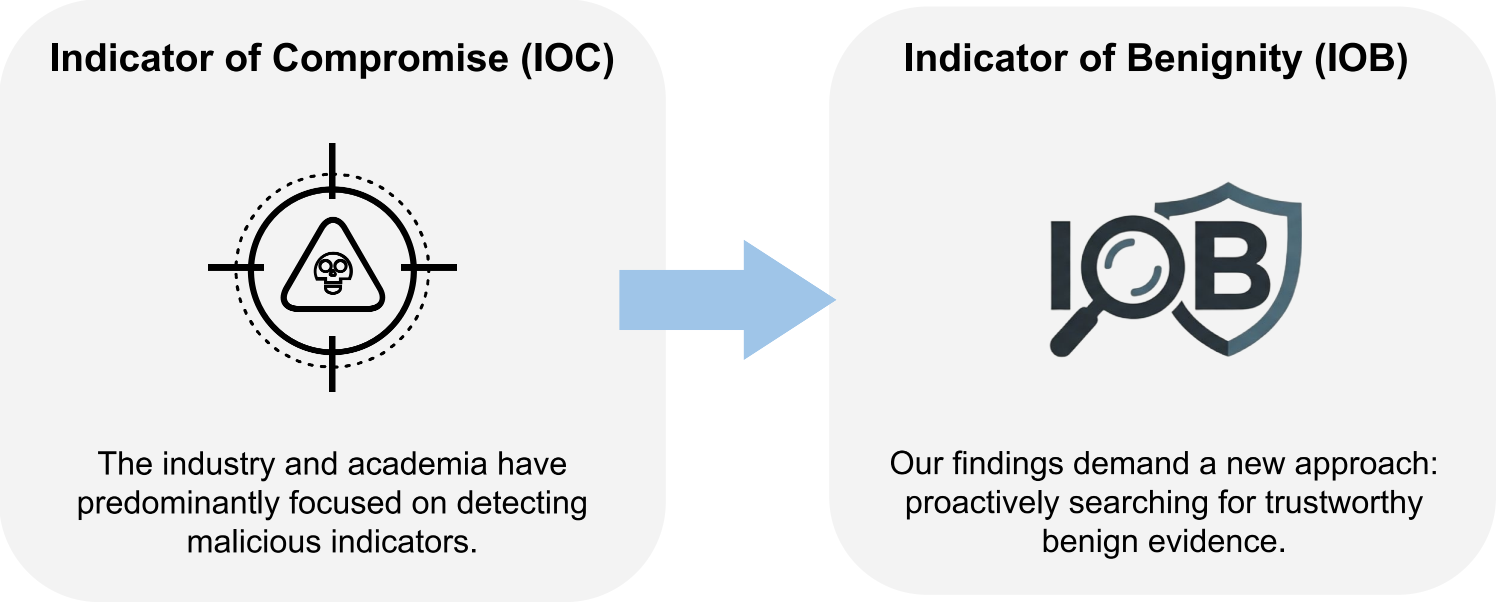
Combining all major public top lists can only cover ~38% of real-world FPs.
Simply expanding these lists introduces an unacceptable risk of false negatives.

Our Core Insight: Benign Domains Leave a Digital Footprint

~55%

For approximately 55% of confirmed FPs, clear indicators of their benign nature could be found on the public Internet.

Shifting the Paradigm from IOC to IOB



Indicator of Compromise (IOC)

The industry and academia have predominantly focused on detecting malicious indicators.

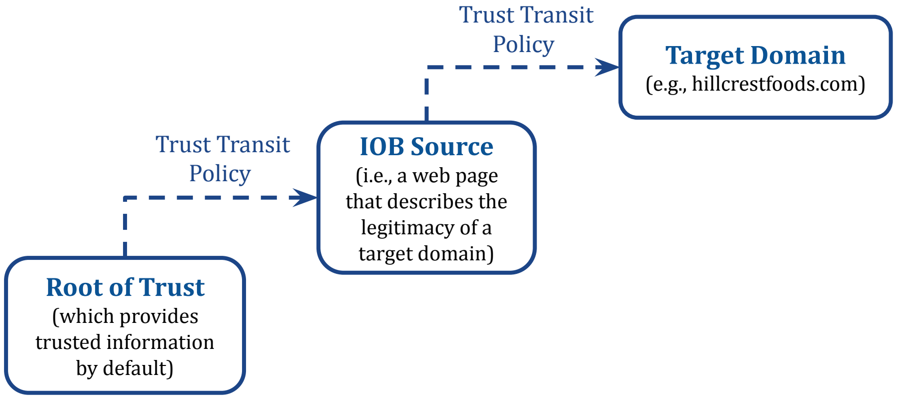
Indicator of Benignity (IOB)

Our findings demand a new approach: proactively searching for trustworthy benign evidence.

An IOB is a piece of web content from a trusted source that attributes a domain to a legitimate application or organization.

The Transitive Trust Model

The main challenge is trusting information on the open Internet. We developed a **Transitive Trust Model** built on a simple principle: a domain is benign if it is owned by or certified by a trusted source.



Root of Trust (which provides trusted information by default)

Trust Transit Policy

IOB Source (i.e., a web page that describes the legitimacy of a target domain)

Trust Transit Policy

Target Domain (e.g., hillcrestfoods.com)

Analogy: This works like SSL certificates establish a chain of trust for encryption. We start with a **Root of Trust** and use policies to transit that trust.

Key Definitions

Root of Trust: Currently defined as official government organizations (.gov, etc.).

Trust Transit Policy: Defines how trust can flow between entities (e.g., from a root domain to its subdomain), while accounting for exceptions like public web hosting.

Introducing IOBHunter

An automated system designed to proactively discover and verify Indicators of Benignity from the open Internet.

How it Works:

1. Search

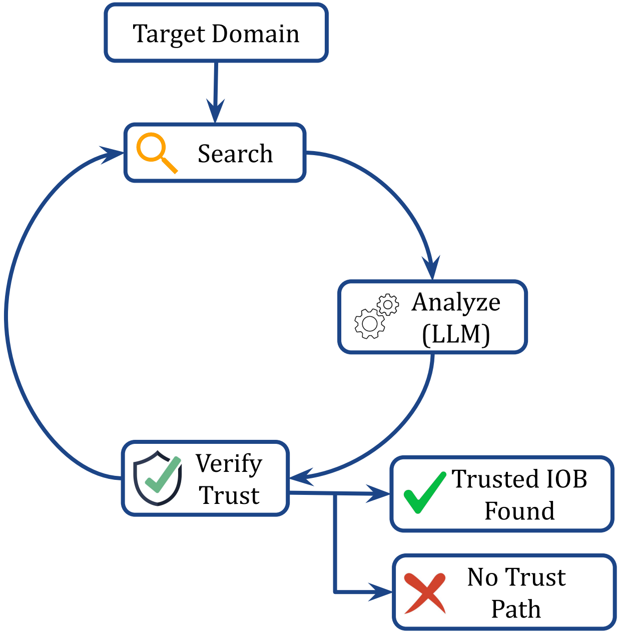
For a given domain, IOBHunter uses search engines to find websites that might contain IOBs.

2. Analyze

IOBHunter leverages **Large Language Models (LLMs)** with **Chain of Thought (CoT)** prompting to parse unstructured web content.

3. Verify

IOBHunter iteratively checks the trustworthiness of IOB sources.



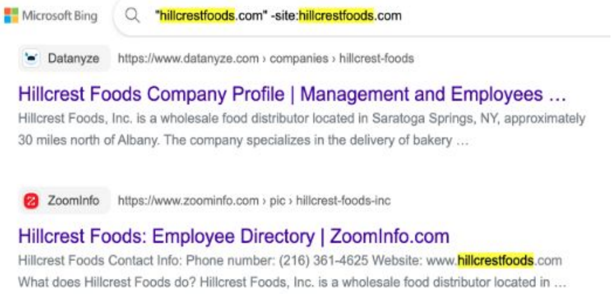
A Real-World Example: Tracing Trust for hillcrestfoods.com

1. The Target

An FP is detected: hillcrestfoods.com

2. Search Round 1

IOBHunter searches for **hillcrestfoods.com**. Finds IOBs on **datanyze.com** and **zoominfo.com**, both attributing the domain to a legitimate company.



"hillcrestfoods.com" -site:hillcrestfoods.com

Datanyze https://www.datanyze.com › companies › hillcrest-foods

Hillcrest Foods Company Profile | Management and Employees ...

Hillcrest Foods, Inc. is a wholesale food distributor located in Saratoga Springs, NY, approximately 30 miles north of Albany. The company specializes in the delivery of bakery ...

ZoomInfo https://www.zoominfo.com › pic › hillcrest-foods-inc

Hillcrest Foods: Employee Directory | ZoomInfo.com

Hillcrest Foods Contact Info: Phone number: (216) 361-4625 Website: www.hillcrestfoods.com What does Hillcrest Foods do? Hillcrest Foods, Inc. is a wholesale food distributor located in ...

Are datanyze.com or zoominfo.com trustworthy? They are not a Root of Trust.

3. Search Round 2

Path A

IOBHunter searches for **datanyze.com**. Finds that oag.ca.gov mentions datanyze.com.

Path B

IOBHunter searches for **zoominfo.com**. Finds that sec.gov mentions zoominfo.com.

4. Conclusion

sec.gov → zoominfo.com → hillcrestfoods.com

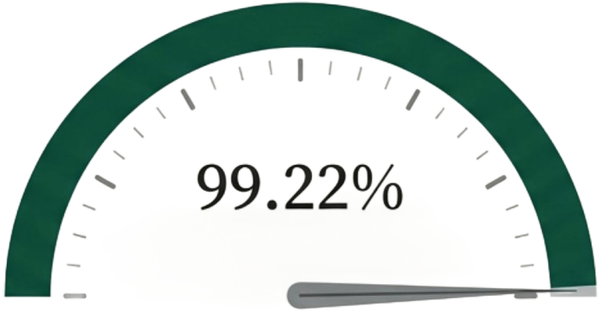
A chain of trust is established: sec.gov → zoominfo.com → hillcrestfoods.com

Result: IOBHunter confirms hillcrestfoods.com **has a trusted IOB and is a false positive**.

Validating Accuracy on Ground Truth Data

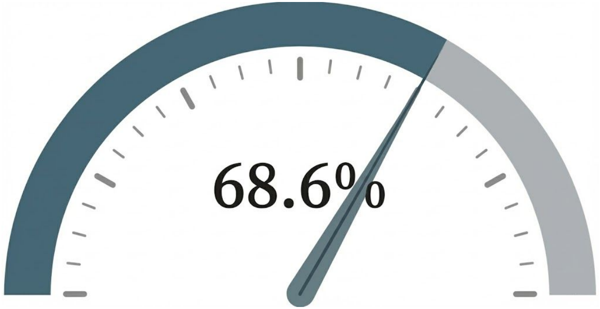
Tested against 67,308 accepted and rejected CRs.

Precision



When IOBHunter finds an IOB,
it’s almost always correct.

Recall



For one-third of FPs, the found
IOB cannot be vouched.

Gemini-2.5-Flash Qwen-3-32B GPT-4o

Consistent high performance achieved across multiple LLMs.

Deconstructing the Nuances

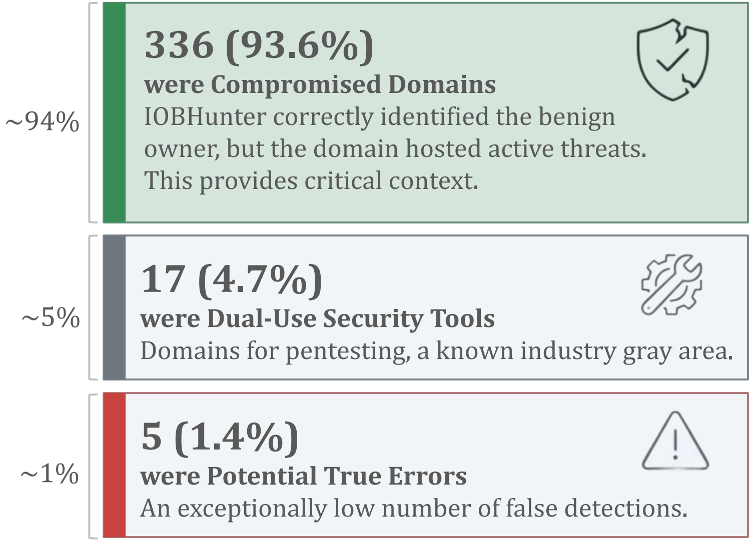
The ‘Missing’ ~32% (Relatively Low Recall)

By Design: IOBHunter intentionally ignores evidence from untrusted sources like social media to prevent manipulation.

The Takeaway: The lower recall is a direct result of our current strict transitive trust model, which prioritizes precision and reliability.

What a ‘False IOB’ Really Means

Total ‘False IOBs’ Found: 359



~94%

336 (93.6%)
were Compromised Domains
IOBHunter correctly identified the benign owner, but the domain hosted active threats. This provides critical context.

~5%

17 (4.7%)
were Dual-Use Security Tools
Domains for pentesting, a known industry gray area.

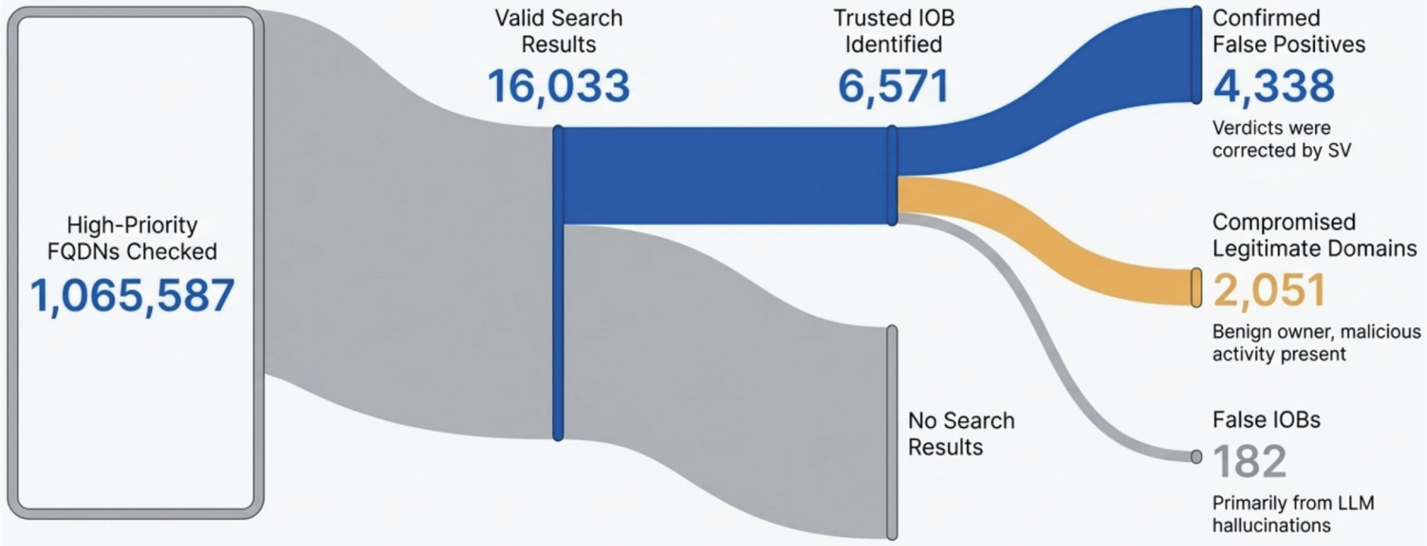
~1%

5 (1.4%)
were Potential True Errors
An exceptionally low number of false detections.

IOBHunter’s ‘errors’ are nuanced findings that separate legitimate owners from malicious activity, a critical distinction for security teams.

The Real-World Impact: Two Months in Production

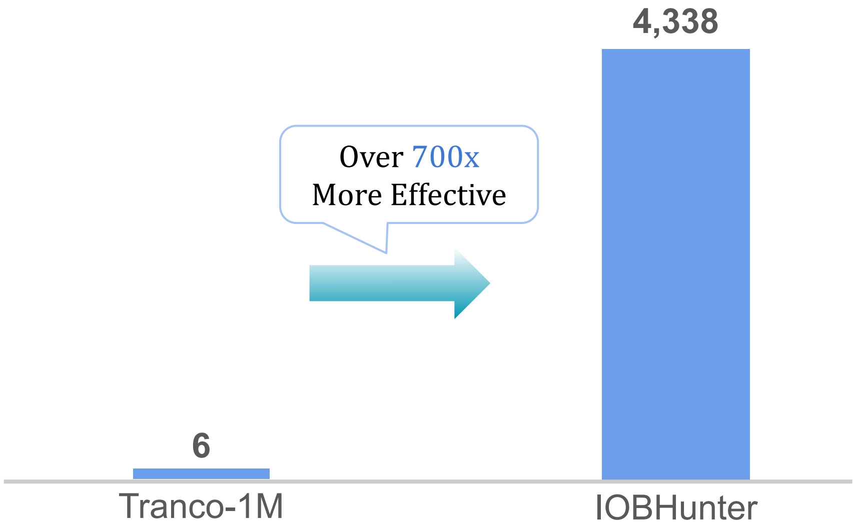
IOBHunter was deployed in shadow mode for two months (Mar-May 2025), analyzing malicious domains detected by Palo Alto Networks.



IOBHunter automatically identified and corrected thousands of FPs before they impacted users.

IOBHunter Catches What Allostlists Miss

Of the **4,338** FPs IOBHunter found, how many would a 1-million Tranco list have caught?



IOBHunter is not a replacement for allowlists;
it is a fundamentally new and complementary capability.

It’s Not Just FPs. It’s About Context

IOBHunter’s value extends beyond simple FP mitigation. By identifying the complete profile of a domain, it provides critical context for security teams.

False Positive

This is a false detection.

Provides a crucial and complementary verification layer to avoid FPs.

Compromised

Found **2,051** domains that were legitimate but **compromised**.

Enables IT to configure flexible policies over simple “block” or “allowlisting”, balancing security and collateral damage to legitimate services.

Alert Triage

Evidence chain can be fed directly into SOC workflows.

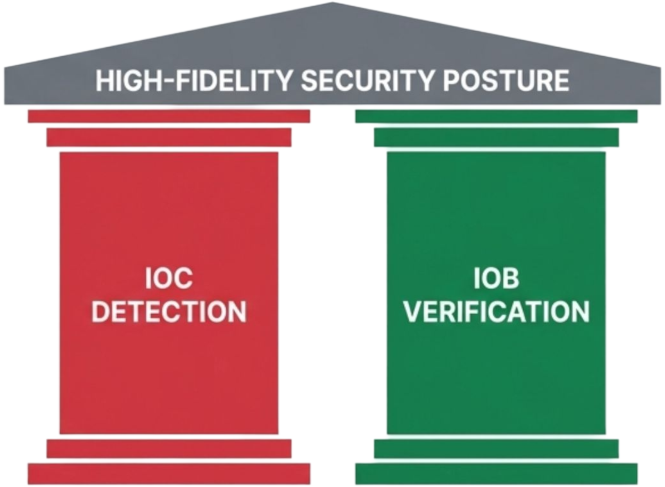
Automates manual research process for suspicious domains, saving analysts time and accelerating incident response.

Conclusion

A New Strategy for Accuracy in Threat Detection

Summary of Key Lessons

1. **The FP Problem is Real and Underestimated:** The scale, delay, and long-tail nature of FPs make them a persistent challenge that current methods fail to solve.
2. **Popularity is Not a Proxy for Benignity:** Relying only on top lists is an insufficient and outdated strategy, missing over 99% of FPs found by IOBHunter.
3. **Proactive IOB Detection is the Path Forward:** Automatically finding and verifying IOB is a powerful, precise, and scalable new approach, and should become a new crucial verification layer to our security posture.



The Future of IOB

1. The transitive trust model can be improved by expanding the trustworthiness to other sources, such as social media, and assessing the robustness of root of trust.
2. IOBs can be integrated as features into next-generation machine learning detectors.

****Call to Action****: To motivate and support future research on this topic, a partial dataset of FPs has been made available at:

https://github.com/dpliu/iobhunter-dataset