
Black Sheep Wall: Clearing the Fog-of-War for Cyber Intelligence Collection

By: [Bernhards Blumbergs](#) Date : April 9, 2024 Categories : [Public exposure](#)



Navigating the vast information space can be a tedious task and establishing the appropriate situational awareness may become even more frustrating as we try to find our path amidst the thickening fog-of-war. Alas, the cyberspace is not StarCraft where we can cheat the game and use **black sheep wall** to reveal the whole battlefield. Nevertheless, there is still a way to step closer towards establishing a contextually sensible situational awareness.

Spotting Black Sheep in the Night

As a society we are an integral part of the information age as we create and consume vast amounts of information daily. The ever-growing expansion of data has led us to the absolute necessity for big data analytics and training the machine learning and natural language processing algorithms to help us make sense of it all. We are already moving away from writing search queries to performing prompt engineering to derive and generate information.

When we browse and search for information online, we see it from a limited perspective of our cognitive perception and what information is being fed to us by the content providers and targeted content delivery algorithms. Despite being fully immersed in the vast ocean of data, we are only exposed to a fraction of it. Does this give you the feeling of being a black sheep yourself once you, in a moment of enlightenment, realize you have lost yourself in the darkness of information space?

Establishing situational awareness and collecting cyber intelligence is becoming ever more important as our lives and daily activities happen online. Imagine the following situation – you and your colleague are doing remote work-from-anywhere and are working together on an open-source intelligence collection against a specified target organization and its affiliated information space. Your colleague is enjoying a work day from Japan with cherry trees in their full bloom, while you are stuck in coldness in Northern Finland. You both enter the same domain name in your browsers and the web page content is delivered to you. Wait a minute – the content does not match – you realize after a moment of banging your head against the keyboard. You both attempt using a VPN and Tor proxy to debug the information space and every time there is a discrepancy in the content displayed by the treacherous website.

Why is our visibility in the same information space distorted, you may ask? The immediate answer is: it is all about your perspective. In reality, it is not that simple and making sense of it all is even harder, especially if we need to establish as complete as possible situational awareness and attempt to expose the true nature of the information source.

Information space visibility is typically limited to a single or narrow set of vantage points, giving only a partial view of the true nature of the information space. Within the context of situational awareness and cyber intelligence collection, the available tools and solutions possess the same design and visibility drawbacks.

It is All About Your Perspective

Nowadays, websites mostly serve dynamic content, which will change based on the connection parameters and source origin. The dynamic nature may commonly be observed for a broad range of information sources, since:

- social networks serve content tailored to the user
- news portals deliver regionally relevant content or its translations
- cloud services will rely on global load balancing
- content delivery networks may restrict access from certain regions
- or access may be blocked based on your connection parameters.

This all creates a highly dynamic information space where the visibility of the content will depend on where and how you access it.

In essence, it all depends on your perspective.

Common situational awareness and cyber intelligence collection solutions reach out to the specified domain name URL from their own limited set of vantage points. As an operator, you get only one search bar to query the results and get a single-faceted perspective on the information space. For intelligence collection, this is not sufficient as it gives a very narrow and biased view of the information space you are trying to analyze.

Ascending the Peaks of Cyberspace

As a part of the ongoing research and development cycle, I have patented a principle and developed a data collector prototype solution – **b-swarm**, which will be released publicly under GPLv3 license in the middle of 2024. In essence, a vantage point is understood as a combination of access origin and access technique, which is designed to provoke the information source to deliver different content. Such vantage techniques include approaches, such as:

- Docker-based container cloud platform deployments in various global IP address ranges
- the use of private or public VPN or proxy connection brokers
- routing traffic through Tor network exit nodes
- or changing HTTP request header field parameters.

Once deployed and collection is launched against the specified information space resources, the collector instances will reach out to the resources from all deployed vantage points simultaneously, attempt to trigger content changes and collect received content. Data and metadata collected from all vantage points together form a single snapshot of the target resource and will represent how it changes based on the vantage point. This scalable approach provides broader visibility and may reveal the dynamic nature of the target information resource.

From the awareness and intelligence perspective, this permits activities, such as:

1. identification of changes within the same snapshot to perform activities, such as, the evaluation of content changes due to geographical distribution and access restrictions, assessment of resources used for cybercrime and targeted attacks;
2. identification of dynamic changes between a sequence of snapshots over time to perform activities, such as, observation of the content availability, tracking and identification of changes in the website content, which may lead to the disclosure and tracking of misinformation and disinformation campaigns, as well as (D)DoS and defacement attacks.

A Few Steps Towards Awareness

There is no way to evade the applicability of machine learning to parse the collected snapshot data. However, there is a necessity to avoid the overhyped trend to apply machine learning to every single aspect of life no matter if it makes sense or not. From this perspective, the collected snapshot data and metadata are analyzed to reduce dimensionality and perform clustering to identify the clusters and outliers.

The machine learning-assisted analysis of collected data significantly eases its assessment by a human analyst. Ultimately, there is only a human analyst who can make sense out of the identified changes and the dynamic nature of a resource, as it is heavily context-driven and depends on the reasons for such data collection in the first place.


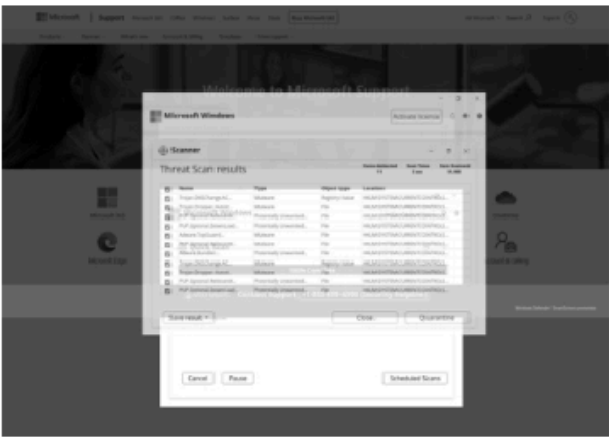

Such a novel approach to contextual data collection may apply to any organizations, agencies, or special services for whom information space awareness, data analysis, and intelligence collection are the key operational activities.

Tip

Collecting data and metadata from the same resource from a large pool of deployed vantage points will grant a better situational awareness and reveal the dynamic nature of the information space.

See the Unseen

As a part of continuous prototype development and data collection, a regular automated cycle of analyzing selected top benign domains and publicly released malicious phishing URL feeds is being performed. Collector Docker instances are automatically deployed across all 37 global IP networks of the Google Cloud Platform to establish broad visibility from geographically distributed vantage points.



AgentID: 056fc5ec-b68c-11ee-8b15-4200a9fe0102:europa-west1:IPNET:007-phish-testing:007
URL: http://zcxzcxzcx.d2jk5f4fer48s8.amplifyapp.com
Visited URL: https://zcxzcxzcx.d2jk5f4fer48s8.amplifyapp.com/

AgentID: 0cb8a6bc-b68b-11ee-90aa-4200a9fe0102:asia-east1:IPNET:007-phish-testing:007
URL: http://zcxzcxzcx.d2jk5f4fer48s8.amplifyapp.com
Visited URL: https://zcxzcxzcx.d2jk5f4fer48s8.amplifyapp.com/store.htm

Content SHA256 match: False
Content Ppdeep similarity: 0
Image Mean squared error: 63.1407
URL match: False | False
SSL match: True

A depiction of a malicious example, where the malicious content delivery depends on the victim's geolocation.

As a first use case, the malicious resource `hxxp://zcxzcxzcx.d2jk5f4fer48s8.amplifyapp[.]com` was identified, which serves a scamming website impersonating Microsoft Defender threat scanner with detected threats and requiring immediate user action. The collected snapshot showed that the website displayed phishing content for connections originating from Japan, South Korea, Taiwan, and Australia, but excluded – Indonesia, Singapore, India, and Hong Kong. This intelligence may reveal the targeted regions, and scope of the malicious campaign, and support the human analyst towards further in-depth investigation and identification of threat actor *modus operandi*.

url https://yandex.net [0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0]



AgentID: 1ac5c629-b5ee-11ee-9d8c-4200a0fc0102:europa-north1:IPNET:005-topdomains-testing:005
URL: https://yandex.net
Visited URL: https://ya.ru/showcaptcha?cc=1&mt=15CA76329820D9599A8F6E8D17EE38C560B26834C6010D913

AgentID: 5873f224-b5cf-11ee-907c-4200a0fc0102:us-central1:IPNET:005-topdomains-testing:005
URL: https://yandex.net
Visited URL: https://ya.ru/?nr=1&redirect_ts=1705597136.00000

Content SHA256 match: False
Content Ppdeep similarity: 0
Image Mean squared error: 107.7433
URL match: False | False
SSL match: True

A depiction of a benign example, where the content delivery depends on the victim's geolocation.

As a second use case, the benign resource `hxxps://yandex[.]net` was identified, which is an Internet services and search engine platform originating from the Russian Federation. The snapshot showed that this resource displays the crawling bot detection prompt only if the request originates from European IP networks. While such a broader perspective gives an additional assessment of the website's behavior, it is not unusual to observe content providers placing restrictions or content filtering based on their policies or collected metrics. Within the current geopolitical situation, such restrictions might be anticipated and may give additional perspective to the human analyst towards intelligence collection and maintaining awareness over the cyber domain especially when it comes to how the Russian Federation is also controlling external access to its online resources.

Conclusions

While traditional single-point-of-view information space awareness solutions yield applicable results, they cannot fundamentally deliver a broader contextual perspective and reveal the dynamic nature of the information resource. Expanding your field of vision from a reasonably larger set of globally distributed vantage points will enrich and bolster your capabilities.

 Tip

Do not trust what you see as your view is limited and distorted. Ask for a second opinion. Better yet – ask for many opinions, compare them, and make a systematic analysis to derive the most appropriate situational awareness.

Credits

Hero image by [Jose Francisco Morales](#) on Unsplash.

Give Us Feedback or Subscribe to Our Newsletter

 Info

If this post pushed your buttons one way or another, then please give us some feedback below. The easiest way to make sure that you will not miss a post is to **subscribe to our monthly newsletter**. We will not spam you with frivolous marketing messages either, nor share your contact details with nefarious marketing people. 😊

Public Exposure Comment Policy

We welcome healthy discussion and debate, but please be constructive with your feedback.

Got it

What do you think?

2 Responses



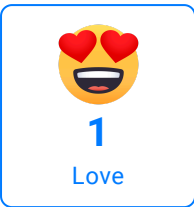
1

Upvote



0

Funny



1

Love



0

Surprised



0

Angry



0

Sad

0 Comments

 Login ▼



Start the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS 

Name

3 Show

Best Newest Oldest