WHITEPAPER

THE PATH TO CONTINUOUS COMPROMISE ASSESSMENT

By: Claudio Deiro

Table of Contents

EXECUTIVE SUMMARY

Today, the cyber world is affected by an ever-increasing number of successful attacks. In 2018 we've seen 1,244 documented data breaches with a global average cost of almost four million dollars.

While defense mechanisms abound, there is a lack of tools capable of providing visibility on the state of compromise of our networks. Without visibility tools that provide timely and precise feedback, it is impossible to react appropriately to threats.

Given the complexity of contemporary computer systems and networks, providing an effective diagnosis may be perceived as a big undertaking. But in today's cloud-enabled world, this is possible. An organization's metadata is the enterprise's gold mine. When an organization's metadata is appropriately collected, processed and stored it can provide clear, accurate and actionable indicators of compromise. This highly prescriptive data analysis model is called Continuous Compromise Assessment®.

In this paper, we review the need to implement a Continuous Compromise Assessment model as a key component of any cyber defense strategy, the steps required for appropriate data analysis, and why implementing this methodology is now possible, including the following:

1. The cost of hardware is decreasing exponentially.
2. Cloud computing enables resource and knowledge sharing, therefore all processing and data storage happens in the cloud.
3. Artificial intelligence algorithms have now reached a level of maturity that makes them suitable for practical applications.

DO WE EVEN HAVE A PROBLEM?

With more than 1,200 vendors attempting to address the wide variety of cybersecurity problems, the answer to this question is clear: There is a problem. This paper is not meant to dissect the problem itself as Lumu recently published a paper on the subject: **The Need for a Breakthrough in Cybersecurity** [1]. Though worth reading, one of the most relevant points of the paper is the incongruent relationship between the investments made in cybersecurity and its indirect correlation with data breaches.

Despite the limited evidence presented here, it should suffice to point out that the problem the industry is facing is real, and with real life consequences.

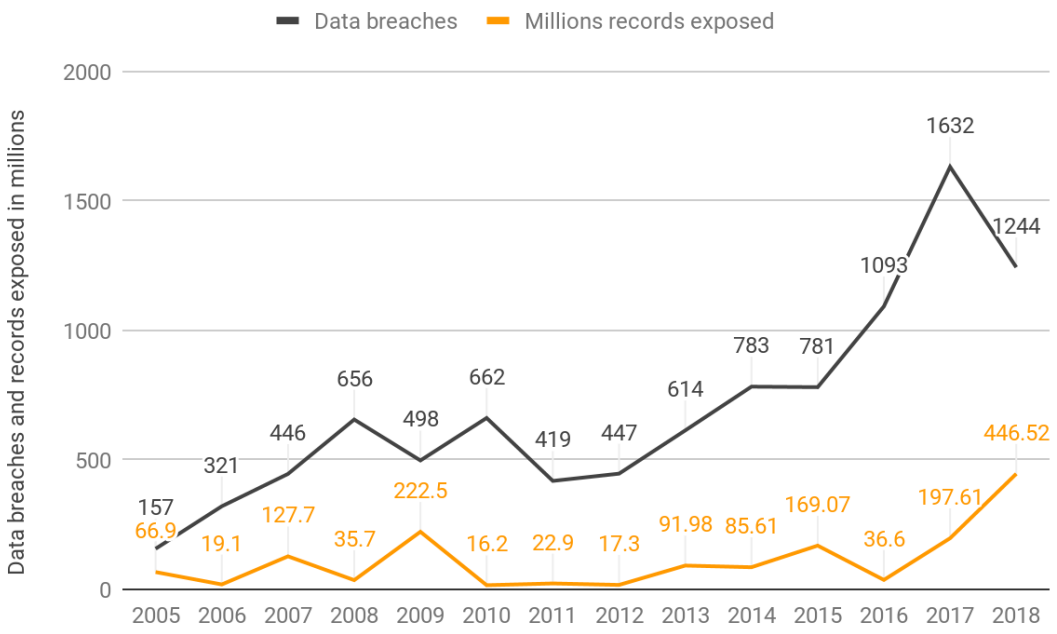


Fig 1. The number of data breaches per year is clearly increasing.[2]

GLOBAL AVERAGE TOTAL COST OF A DATA BREACH

Measured in US$ millions

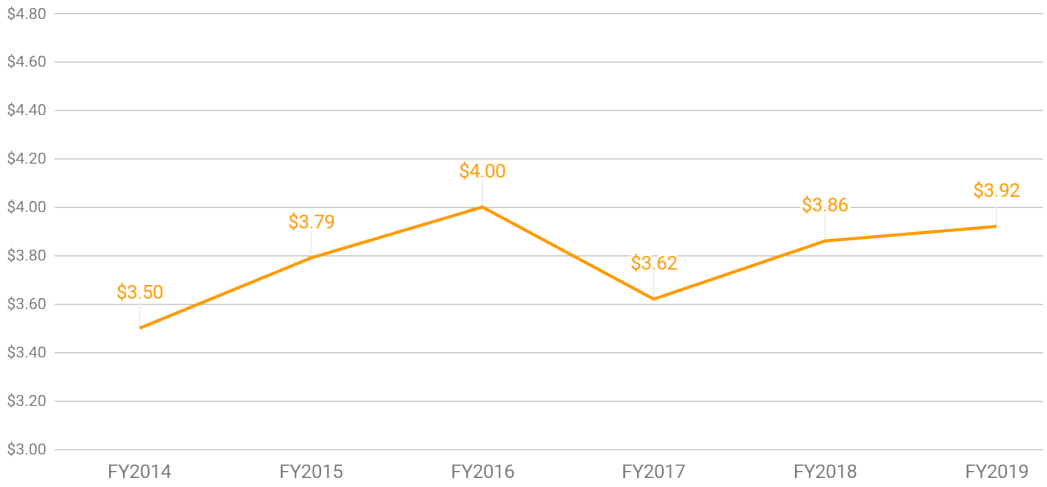


Fig 2. The global average cost of a data breach. [3]

TOTAL COST OF A DATA BREACH BY ORGANIZATIONAL SIZE

Measured in US$ millions

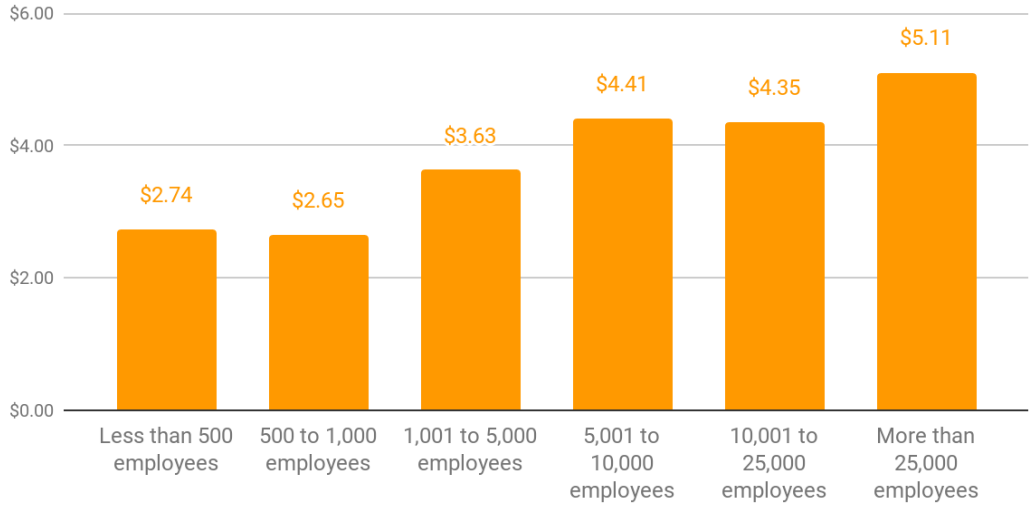


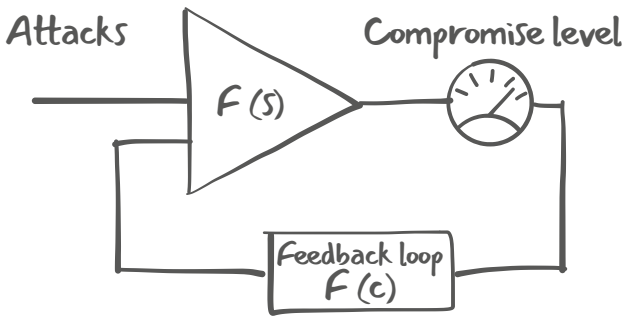
Fig 3. Total cost of a data breach by organizational size. The cost relative to the company size is much higher for smaller organizations. [3]

THE COUNTRY OF THE BLIND

An important contributor to the cybersecurity problem outlined above is the limited visibility on the state of compromise of our networks. As such, organizations need to diligently find ways to develop enough visibility to understand what goes on within their networks. There is an emphasis on building ever-better walls to signal a boundary not to be crossed, and to give time to guns to intervene. There is a missing element, one that it is perceived as obvious: eyes. Eyes to document the infractions and guide the guns.

In contemporary cybersecurity we have an abundance of walls, but we are almost blind and we have almost no guns. While the scarceness of means to efficaciously retaliate is an issue beyond the reach of technical solutions alone, involving international treaties, nation states goodwill, and the like, we—as cybersecurity professionals—have the ability, and the duty, to provide our customers the tools to timely identify compromises and breaches.

Cybersecurity as a Control System



An alternative point of view is to look at cybersecurity's problem as a control system [4]. For the system to work—that is to say to keep the compromise level within acceptable limits and prevent serious damage—a precondition is the existence of sufficiently precise and timely feedback. Nowadays the feedback is often the message of a security researcher that notifies the victim that its data was found on the dark web. It might be a uranium centrifuge unexpectedly failing: too late. Alternatively, it may be a stream of thousands of alerts each day triggered by heuristic rules: too much.

> "ORGANIZATIONS NEED TO DILIGENTLY FIND WAYS TO DEVELOP ENOUGH VISIBILITY TO UNDERSTAND WHAT GOES ON WITHIN THEIR NETWORKS".

Compromise as a Disease

To use a healthcare metaphor, compromise is like an infection. Firewalls and EDRs—the current staples of cybersecurity—are preventative measures. Now, "no amount of prevention will help you when prevention fails" [5], and therefore you become sick. To fight back the infection we need diagnostic tools and antibiotics.

In cyber the cure can be easy—format a machine, change access credentials, strengthen firewall rules, run cleanup tools—to diagnose the infection is not. Our machines and our networks are a mess, much like our bodies. Much like their real life counterparts, computer viruses have learned to mutate, making the traditional signature-based approach to detection reactive and very much dependent on the agility of the provider. Thousands of processes, on thousands of machines, interchange thousands of packets every second. Finding the malicious threat is like finding the proverbial needle in the haystack. Behind a single IP address in the cloud there can be hundreds of applications. Some of them may be malicious, or infected themselves. But how do we find out?

There is simply no way for the unaided human eye to make sense of all this noise.

This is also the reason why a whitelist approach—only permitting what is known to be safe—won't work outside of very special cases. The whitelist would eventually outgrow any management capacity, and what at one moment is known to be safe may become infected, and infectious, a moment later.

A POSSIBLE PATH TO CONTINUOUS COMPROMISE ASSESSMENT

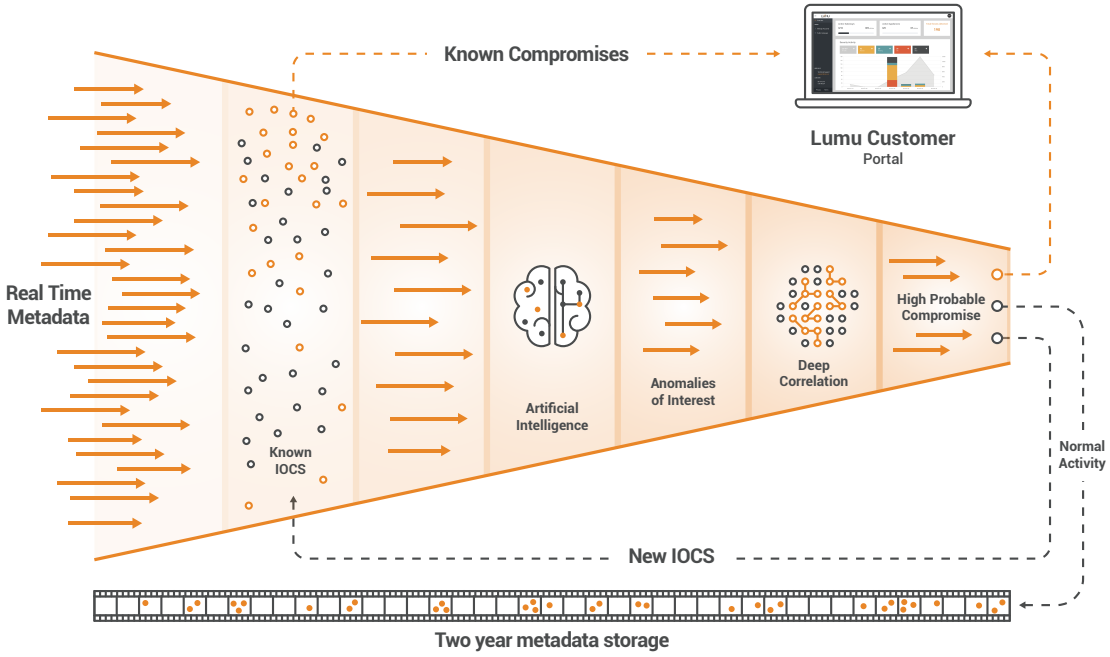


Fig 4. Lumu's patent-pending "Illumination Process"

Organizations often overlook the power of their own network metadata. Today, this is the most promising path for transformational improvement in the world of cybersecurity. This goldmine contains incredible potential, as long as it is used correctly. The process outlined below exemplifies how to best leverage network traffic:

1. For starters, metadata is collected in real time to later be contrasted against a large pool of known, certified IoCs, coming from the organization itself as well as private and public sources of curated detailed cyber threat intelligence. Alerts are generated upon the identification of matching data.
2. Once alerts are generated, organizations have a clear indication that prompt action is required.
3. All incoming metadata should then be put through artificial intelligence and heuristic inference engines that would allow us to understand anomalous behavior, in order to reduce false negative rates. For example, unusual traffic patterns generated by an asset coming in contact with points within the network that are out of the ordinary and/or with a certain frequency. What results from this filtering process is a list of anomalies of interest, which may represent compromise.
4. Anomalies of interest should be later put through a deep correlation process which consists of taking the traffic deemed as likely to be related with malicious actors and confirming its compromise nature. For example, cybercriminals often use the same group of IP addresses or a specific segment in the network, as well as the same domains in rotation. The deep correlation step generates only alerts of high-probability.
5. The last step in this process is to store the residual network metadata for a period of time and leverage emerging IoCs for further correlation and analysis. This step is critical because it will enable organizations to constantly improve their Continued Compromise Assessment process.

The implementation of Continuous Compromise Assessment can have a transformative effect on the cybersecurity industry. The most natural question at this stage is 'Why has this not been done before?' First, we did not know the impact of the interconnected world and the extent of malicious attacks. Knowing what we know now, Continuous Compromise Assessment will be an absolutely critical step in any enterprise's cybersecurity strategy. Secondly, this process is only possible as of very recently. About 10 years ago, carrying out this process would not have been effective or practical. There are a few reasons why, which are enumerated below:

1. Data Storage Cost & Computing Power

The cost of storing data has decreased 3,000 times in the last 20 years while computing power has increased 10,000 times since the year 2000 [6]. These conditions have helped build the perfect scenario for the collection and administration of large volumes of metadata as well as the execution of deep learning capabilities.

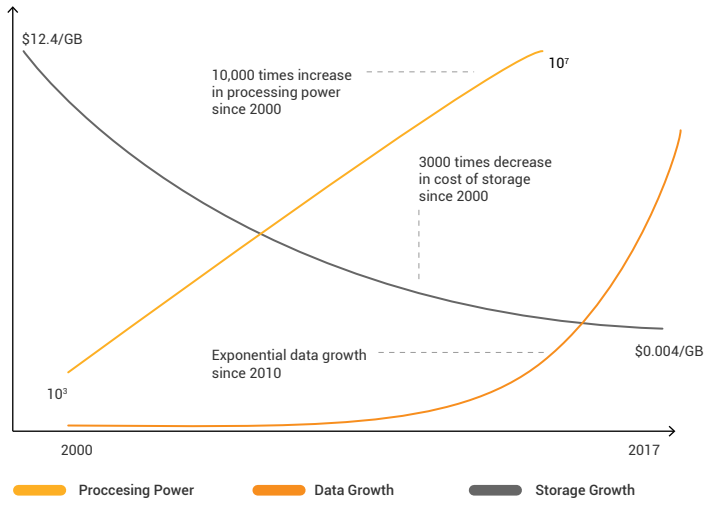


Fig. 5 Increase in processing power and reduction in data storage cost [6]

2. The Cloud

There are good reasons to place processes and storage in a cloud computing environment. A critical factor for the success of a system of this kind is the time between when new intelligence is available and when such intelligence is incorporated into the system. With on-premise systems there will always be a delay in the distribution of such information.

On the other hand, with a cloud-based system, new intelligence will be available to all users as soon as it is available to the system.

"THE IMPLEMENTATION OF CONTINUOUS COMPROMISE ASSESSMENT CAN HAVE A TRANSFORMATIVE EFFECT ON THE CYBERSECURITY INDUSTRY".

The most important factor that makes a cloud deployment ideal for a system like this is the removal of all maintenance and management burdens on its users. The valuable time of skilled security professionals should no longer be spent on system maintenance, including monitoring disk space usage, or writing rules to catch the latest infection. All these menial tasks are transferred to the cloud environment. Security professionals can concentrate on investigating and remediating incidents.

3. Machine Learning & AI Renaissance

Artificial intelligence was all the rage in the late '60s and early '70s. It then went out of fashion, for a decade (or four). But now, with immensely more powerful hardware and somehow curbed expectations, it is riding high again. Besides the fact that talking about AI is trending, the use of machine learning and anomaly detection for Continuous Compromise Assessment may bring real advantages.

Today's machine learning algorithms are sufficiently well understood and can be supported by enough computing power to be successful in practical contexts.

Storage cost, computing power, combined with cloud, machine learning and AI, make this approach absolutely practical, effective and possible. As they say, the devil is in the details and a detailed look at how data is collected merits a discussion.

How to Leverage Your Metadata & Overcome Challenges Along the Way

DATA COLLECTION: DO YOU REALLY WANT TO SEE IT ALL?

The ground truth is in the network data. Unlike logs, that can be tampered with or simply deleted, or EDRs (Endpoint Detection and Response), that have to play at the same level as the

attacking software, there is no way for an attacker to interfere with packet capture and analysis. So a conscientious network administrator should capture and analyze everything. Right?

As the ones familiar with Betteridge's law of headlines [7] have already guessed, the answer is no: it would cost too much. You would need roughly double the amount of bandwidth and computing power of the original network, simply to analyze all its data.

Fortunately, "traffic patterns reveal a lot about any organization and are much easier to collect than actual communication data" [8]. This means that a much more effective approach is possible. In this approach, what is collected and analyzed for the whole network is the metadata.

AN APPROACH TO COMPROMISE DETECTION BASED SOLELY ON INDICATORS OF COMPROMISE (IOCS) IS BOUND TO FAIL

Please note that a partial approach, where only the critical systems are under control, would not suffice. Attackers inside your network would be able to move laterally and take control of less critical systems until they are in the position of reaching the resources they are after without raising suspicion. In case you are interested, reading the account, in her own words, of how Phineas Fisher hacked Hacking Team [9] can provide a good idea of the steps an attacker may take, from a peripheral firewall to an unsecured test database to a backup storage and finally—with some additional steps—to everything. It is therefore necessary to include in the analysis all network devices, including the ones considered less critical.

Metadata Collection and Consolidation

Collecting the metadata seems to be a rather trivial process. After all, most organizations most likely already generate and collect such metadata. However, there exist at least two problems with collecting metadata:

1. Metadata comes in different formats, making it difficult to collect, organize and consume.
2. Some organizations do not have processes in place for metadata collection.

The first problem can be overcome using de facto standards, like Cisco's NetFlow [10] and Elastic's Packetbeat [11]. We can address the second problem using a stack of existing software components that can be easily customized to fulfill the user's needs.

Metadata Analysis: The Trouble with Indicators of Compromise

When looking at an organization's network metadata a particular IP or domain is identified, it is easy to conclude that this network has been compromised. Well, not so fast. An approach to compromise detection based solely on Indicators of Compromise (IoCs) is bound to fail for at least three good reasons:

1. **Reactive approach:** IoCs should be identified, confirmed, and divulged. Looking for IoCs does not help the first victims of an attack, or the targets of customized attacks. An IoC-only approach will present a high false negative rate.
2. **Noise:** IoC lists suffer from a high noise level. Part of the reason is that they are often compiled in automated form. But the high reuse and sharing rates for network resources imply that just seeing an IP is most often not sufficient proof of a compromise. An IoC-only approach will present a high false positive rate.
3. **Lack of context:** Context is often necessary for the interpretation of the data. And without interpretation it is impossible to really understand what is going on and take the appropriate corrective actions.

Metadata Analysis - Anomaly Detection

The occurrence of a compromise will cause a change in the behavior of the network, possibly a very subtle change to escape detection. For example, a botnet agent will phone home to let the botmaster know a new bot is available. A worm will try to contact neighbors to infect as many machines as possible. A coin miner will contact the C&C to get new jobs and report results.

The emergence of anomalous behaviour gives us a chance to discover compromise, if we can appropriately learn how the network behaves and detect changes.

Unfortunately, the network is ever-evolving. The behavior of a network can change for reasons as benign as the installation of new software, or a new version of already existing software.

Or the deployment of a new web application. And possibly almost every user will need a unique combination of applications and will present a unique behavioral pattern.

Relying solely on anomaly detection would therefore generate an unreasonably high number of false alerts.

Metadata Analysis - Suspicious Behavior Detection

When analysis begins, a compromise could already be there, possibly in a fairly high number of machines. On the other hand, the changes introduced in the behaviour of the network can be so subtle, ever so gradual, that the anomaly detection system does not trigger.

Therefore, we need a set of models, heuristics and rules that can detect suspicious behavioral patterns.

Metadata Analysis - Deep Correlation

At this stage, we have a series of anomalies, such as a machine contacting a web server never seen before; and suspicious patterns, such as a series of machines constantly posting small amounts of data to an unknown external server. There is little we can say with this information. Maybe the website is little known, but showed up in the results of a particular search, and hosts relevant content. Maybe custom software is in use that is sending home some telemetry or diagnostic data.

Now, let's imagine that the little known web server is hosted by a known bulletproof hosting provider [12, 13], and in other instances a visit to this host has been followed by malicious activity. Most would agree that it would be wise to take a look at that machine. If, instead, it turns out that the website was recently created by a reputable owner, the anomaly can probably be safely ignored. On the other hand, imagine that the same posting pattern is also observed in other unrelated users, and the receiving system is not known to be managed by a legitimate organization. Would you look into it?

The examples attempt to show how correlating metadata with available intelligence enables filtering out a significant part of the detected events, leaving for human investigation only events that with high probability are compromises. It also allows us to enrich the provided information with context, so that human analysis is easier.

Metadata Storage - The Importance of Memory

After analysis, the metadata is stored for a substantial amount of time. This, while generating non-negligible costs, allows for forensic analysis and reanalysis. The importance of forensic analysis is pretty obvious.

The idea with reanalysis is that the metadata will be scrutinized again as new intelligence or algorithms become available. Threats that unfortunately escaped the first round will therefore be discovered, hopefully before significant damage occurs.

It is of course vital that the data is stored correctly, so that the needed information could be efficiently retrieved. The storing technology should also allow for a high degree of flexibility, as it may not be obvious what exactly will be needed in the future. Fortunately, the technological advances in big data treatment make it possible to meet these requirements.

IF NOT NOW, THEN WHEN?

Almost every report written in cybersecurity ends with a high sense of urgency. It is the nature of the industry we are in. This one is no different. There is no doubt that it is time for a breakthrough in cybersecurity. The current state is simply unsustainable. This is not meant to sell the famous fear. On the contrary, the solution is largely in the hands of practitioners. In addition, because a lot of missing pieces have recently fallen into place: cost of storage, computing power, cloud infrastructures and functional machine learning.

As an industry, it is our responsibility to respond with the democratization of cybersecurity, where the tools for executing sophisticated attacks are readily available. This also means that advanced and efficient technology and intelligence are available to all who want them, and those who dare to put them to work.

REFERENCES

[1] https://lumu.io/resources/the-need-for-a-breakthrough-in-cybersecurity/

[2] https://www.idtheftcenter.org/data-breaches/

[3] https://www.ibm.com/security/data-breach

[4] https://www.bankinfosecurity.com/interviews.php?interviewID=4475

[5] https://securityboulevard.com/2019/04/new-tech-network-traffic-analysis-gets-to-ground-truth-about-data-moving-inside-the-perimeter/

[6] https://medium.com/@rpradeepmenon/an-executive-primer-to-deep-learning-80c1ece69b34

[7] https://whatis.techtarget.com/definition/Betteridges-law-of-headlines

[8] Bruce Schneier (2000) - Secrets and Lies: Digital Security in a Networked World

[9] https://www.exploit-db.com/papers/41914

[10] https://www.cisco.com/c/en/us/products/ios-nx-os-software/ios-netflow/index.html

[11] https://www.elastic.co/products/beats/packetbeat

[12] http://www.blueangelhost.com/blog/bulletproof-hosting/

[13] https://krebsonsecurity.com/tag/bulletproof-hosting-providers/

Illuminating threats and adversaries

www.lumu.io

Lumu Technologies Inc. | 8350 NW 52nd Terrace Suite 301, Miami, FL 33166 | info@lumu.io | +1 (877) 909-5868