BLUEHEXAGON

The CISO Manifesto

Redefining Metrics for Threat Detection

CISO Contributors:

Rich Mason
Richard Seiersen
Tom Baltis
Anne Marie Zettle Moyer
Greg Shannon

Introduction from Nayeem Islam, CEO Blue Hexagon

This whitepaper and the CISO Manifesto event began as a series of discussions with CISOs Rich Mason and Richard Seiersen as we were getting ready to launch our deep learning platform. As a long-time advocate and practitioner of deep learning in my previous roles at Qualcomm, Amazon and IBM, I believed that we could use deep learning to perform inline threat inspection of network traffic with high efficacy verdicts. When we actually tested our product in real-world environments, the performance astounded even me -- inline threat inference in **less than a second at greater than 99.5% accuracy**.

As we began to share our product performance with CISOs everywhere, I realized that there was a definite sense within the CISO community that traditional approaches to security haven't kept pace with the threat environment and that the metrics that have been used for years are no longer valid. In fact, for far too long, vendors have promoted numbers that put their product in the best possible light based on parameters that may or may not have anything to do with keeping *organizational* data safe. As an industry, we can do better. CISOs need security vendors to do better. As Rich Mason put it, "we need to migrate from vendor-defined, practitioner-delivered metrics to practitioner-defined, vendor-delivered metrics".

Blue Hexagon together with CISOs Rich Mason and Richard Seiersen initiated a discussion to rethink the way the industry measures threat detection on March 3 during the eve of RSA with our CISO Manifesto event. It was a discussion that featured a panel of respected security leaders. Former Honeywell CISO Rich Mason served as moderator, with panelists: author and CISO Richard Seiersen; VP Security Engineering, Mastercard, Anne Marie Zettlemoyer; chief scientist for the CERT division at Carnegie Mellon University Greg Shannon; and Delta Dental CISO Tom Baltis.

Attendees at the Blue Hexagon event included CISOs from across the spectrum of industry-- financial services, healthcare, entertainment, consumer services, manufacturing-- including Bank of America, Netflix, Mastercard, Humana, Mitel, and Palo Alto Networks. The attendees packed the room and engaged in a vigorous discussion about their frustrations, successes, and values.

This CISO Manifesto is a result of discussions with this audience-- during and outside of this event .

One theme dominated the discussion: measurement is vital to the success of a security program. That measurement shouldn't be of things that a vendor thinks is important, but what is important to an organization, its security leaders and its customers. I'm excited to see that the key areas that CISO leaders believe are important are where Blue Hexagon is focused on delivering value -- time to detect, doing it as close to the source of attacks as possible and with very high efficacy.

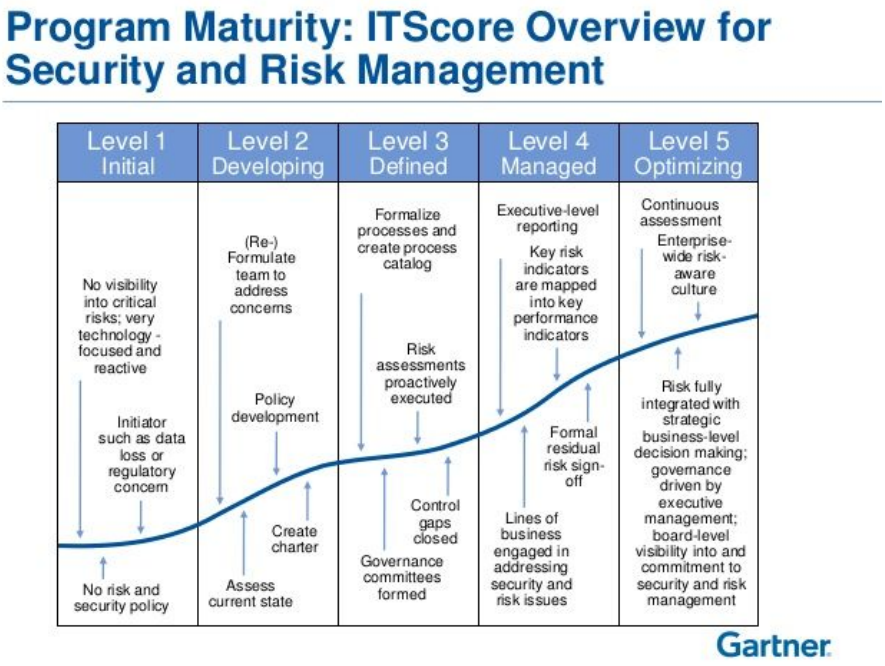
Thank you to Rich Mason, Richard Seiersen, and many CISO attendees for their close collaboration on security metrics. We're looking forward to additional metrics discussions, and demonstrating how we deliver on them alongside many other security vendors.

Metrics Matter

In almost every discipline, measurement is a mandatory and key component of the actual development process. For example, when engineers build a bridge, the measurement of whether it can withstand gale force winds is incorporated into the architecture and engineering process. When a drug manufacturer creates a new drug, measuring how well the drug performs and its impact on users is part of the drug trials. There is typically no barrier between the concept of measurement and science; to decouple this would be odd. Yet, in cybersecurity, measurement is not always a fundamental element of a security strategy.

"You're managing by anecdote until you measure. If you're not measuring, you're not making a data influenced decision" - Anne Marie ZettleMoyer, VP Security Engineering, Mastercard

In fact, in the Gartner Maturity Model for Security and Risk Management (figure 1), key risk indicators mapped into key performance indicators aren't incorporated until maturity level 4. This seems a little late considering that if security leaders cannot measure whether security controls and processes are working from the very start, they won't know whether they are actually working.



Metrics and measures are terms that in many ways serve as the building blocks for how security performance is both assessed and achieved. Therefore, it's paramount to have a fundamental understanding of them. Measures and metrics are performance indicator tools that can be quantitative or qualitative. Measures measure only one thing (e.g., I have five security incidents this month). In contrast, metrics describe a quality and require a measurement baseline (e.g., I have five more security incidents this month than I did last month). Measures are useful for demonstrating workloads and activity, and metrics

are useful for evaluating process effectiveness and measuring success against established objectives at a system (high) and indicator (low) level.

When considering metrics, it is important to consider the context of the audience the measurements are for. We often gravitate towards metrics for measuring things that practitioners as are familiar with but may not be important to other constituents such as the CEO and board of directors. Security teams need to consider the different audiences where metrics need to be produced. In fact, in most cases, metrics at the operations level need to cascade up to business metrics in order to ensure security strategy success is ultimately tied to business objectives.

Key takeaways:

- Measurements need to be considered from the very start of kicking off a security program rather than waiting for an advanced maturity level. It's not intuitive, you need to staff with appropriate analysts and data scientists who can look at data and generate insights on what is and is not working.
- A "measure" is a number that is derived from taking a measurement. In contrast, a "metric" is a calculation between two measures
- When considering metrics, it is important to consider the context of the audience.

Distance Matters

The perspective of whether distance matters -- whether threat prevention and mitigation should occur as far away from the defended assets as possible -- generated the most varied responses from panelists and the CISO audience.

If you take an example of the human immune system and how diseases spread, it is better to stop pathogens as quickly as possible in the external barriers such as the skin and all mucous membranes before it enters and spreads throughout the body.

We are in fact already using epidemiology terms such as viruses. The epidemiology examples of false/true positive and negative results also extend to cybersecurity. Putting security controls closest to the entry point of attacks means the ability to stop threats before it is executed, thereby preventing lateral spread throughout the entire network (similar to virus spreading). In fact, if you can detect a threat at the network quickly, you can stop the initial incursion and further damage from lateral movement.

	Epidemiology	Cybersecurity
True positive	The patient has the disease and the test is positive.	The cybersecurity product identified a threat accurately.
False positive	The patient does not have the disease but the test is positive.	The cybersecurity product mistakenly identified benign traffic as a threat.

True negative	The patient does not have the disease and the test is negative	The cybersecurity product identified benign traffic accurately as not a threat.
False negative	The patient has the disease but the test is negative.	The cybersecurity product did not identify a threat correctly as one.

However the decision to focus on network security or endpoint security may vary based on the architecture of the enterprise environment, and what is needed to support the business. For example, an enterprise supporting cloud infrastructures may consider the cloud and digital workloads as the perimeter. Enterprises that have more legacy infrastructures may have a crunchy perimeter where focusing on security controls around the network perimeter instead of endpoints matter more. Most environments are likely to be hybrid environments, therefore it is very important to be strategic about where to place security controls. Regardless of classical or modern workload architectures, it is important to have the security controls close to potential incursion points where the “defined perimeter” is at.

Many vendors look to skirt any real measurement of value with the “defense in depth” platitude – adding one more layer must be good! Malcolm Harkins Chief Security and Trust Officer of Cylance is calling this “Expense-in-Depth”. We need to be measuring each layer, and throw out the layers that add limited value and introduce unnecessary cost and complexity. We can do this with every layer -- identifying where a threat is stopped and to determine the value of security controls.

Key takeaways:

- Distance does matter in cybersecurity. The quicker you stop a threat closer to the source or entry-point, the better.
- Consider the enterprise architecture and the placement of the right types of controls that are needed to support the business.

Time Matters

Time to detect should keep pace with the speed that attackers are unleashing attacks. How fast should threat detection be? If we agree faster is always better, who defines this? Our group of CISOs believe that it is the attacker that defines how fast threat detection should be.

A great deal of security metrics today is presently focused on this concept of speed --- dwell time, mean time to detect, mean time to respond, and mean-time to contain. But there is great disparity between our goals and actual. Dmitri Alperovitch CTO at Crowdstrike has advocated a 1-10-60 target where threats are detected in 1 minute, followed by investigation within 10 minutes and complete removal or containment of the threats within 60 minutes. Johna Till Johnson, CEO and Senior Founding Partner of Nemertes Research recently set the bar at a combined 8 minutes for all three in order to achieve a 98% percentile rating, based on her research with 600 correspondents.

These are aspirational goals today; the reality is that the average dwell-time is cited as anywhere from 50-200 days depending on which source you are referencing. Mandiant's 2018 report cites average dwell time as 101 days. Additionally, In 2016, the median duration between the start of an intrusion and it being identified by an internal team was 80 days, but in 2017 this number decreased to 57.5 days.

Here are the areas that are important for consideration:

- **Mean time to detect** is valuable if the fidelity is high. If a security vendor brings up alerts every half a second, and are flooding the Security Operations teams with alerts they cannot process, it is an issue. False positives matter. Triggering a number of alerts, where a high percentage is false positives, means that the investigation and removal/containment times will be impacted.

  Additionally, the mean time to detect needs to be appropriate to the organization's risks and protection of its crown jewels. A security team might require extraordinary speed with threat detection and response when there are physical ramifications, such as with SCADA environments. For certain organizations, an attack that wipes out the business infrastructure/system may be worse than an attack that exfiltrates data out.

  Are "mean" times good enough or should organizations be taking more of a Six Sigma approach to shrink the tails of the bell curve. According to Richard Seiersen, if you use averages, you only get average results. The recommendation by Seiersen is to use "survival analysis" or "time to event" analysis. In short, it allows us to say things like, "Over the past year 50% of advanced threat were detected in 10 days or more, and 10% were detected in 100 days or more."

- **Principle of first observation** is important. Most security vendors deliver very poor efficacy the first time they see a threat sample; in fact unknown threats will bypass signatures and sandboxes the very first time. Subsequently, when dynamic analysis results catch up and a signature for the unknown threat is created and tested, the verdict is updated. Some security vendors such as Cisco even touts this as a feature, calling it "continuous analysis and retrospective security capabilities". But, if you take the analogy of a bullet proof vest, it is similar to the first bullet getting past and blocking the next 99 bullets. You would be 100% dead, yet in cybersecurity terminology, vendors will report this as 99% efficacy. The metric needs to be more accurately represented as "time-to-verdict" rather than "time-to-detect", and the performance based on first observation is what should be measured.

- One key consideration is not just the detection of the threat but the **actual prevention**. Prevention can mean different things -- stopping malware from being downloaded, stopping actions on objectives, or stopping lateral movement. While real-time prevention on the same product that performs the detection will be the fastest, most organizations will rely on orchestration of an action downstream, either directly to other products or via security and orchestration platforms (SOAR).

  Time to prevent will therefore depend on the polling intervals and architecture of various related security products. For example, taking the same example above on network threat detection, prevention can be applied at the network device layer (via firewalls, secure web gateways and proxies) and endpoints (typically via endpoint cloud infrastructures). The combination enables the actual execution of a malware to be stopped on the endpoint, but also ensures that further communication by the malware (C2 communications) is stopped on the firewall. Orchestration of all these components can be simplified via security automation and orchestration, but may also extend the time to prevent.

- **Speed along with adaptability to new techniques** is the next challenge. Adversaries will undoubtedly be improving their own metrics, including speed. AI will also be an enabler to them. As we head towards a near real-time arms race, security products will need to adapt in response to an agile threat landscape and the next-generation of attack techniques.

Key takeaways:

- Mean time to detect upon first observation, with high fidelity is the ideal metric for threat detection
- Mean time to prevention is important but may depend on various factors such as the security products participating in the actual prevention.

Putting CISO Manifesto Metrics to Action

In response to the key metrics defined by our CISOs, here is an example of a test plan for network threat protection solutions.

Capability	Why this is important	Metric
Detection time	Time matters. The verdict time should be calculated based on first observation of the threat sample.	Verdict Time
	Can the security solution keep up when thousands of threats are being unleashed by attackers at the same time.	Verdict time at scale
Detection Breadth	Hackers use padding and other evasive techniques to make files larger to exploit sandbox limitations.	Detection of threats in Small Files
		Detection of threats in Large Files
	Value matters. The ability for a threat detection solution to address breadth of file-types is critical. While threats can be included in a variety of documents, EXEs, DLLs, Office documents and PDF are the dominant file types.	Detection of threats in EXEs
		Detection of threats in DLLs

		Detection of threats in Office Documents
		Detection of threats in PDF Documents
Threat categorization	Threat detection solutions should identify threat categories for SOC and threat analyst teams to understand the behavior of the threats.	Threat Category (example: trojans, ransomware, downloaders etc)
	Threat detection solutions should also identify families so that SOC and threat analyst teams have information for further analysis and remediation efforts	Threat Family
Detecting old, fresh and fresh mutated threats	Value matters. The ability for a threat detection solution to address both known and unknown threats is important	Old malware (2-3 years old)
		Fresh malware (less than 24 hours old)
		Fresh mutated malware
Preventing threats	Time and distance matter. If threats are detected at the network, can prevention be quickly enabled at other security products -- endpoints, proxies, or firewalls to stop the execution of the malware, C2 communications and lateral movement	ICAP Integration
		Endpoint protection
		Firewall Integration

Conclusion

Whether you are a practitioner who has ideas about how the vendor community should change their approach to delivering security products, or you are a vendor who has ideas about how we must change how we test and showcase our features to better serve customers, it's clear metrics are important. This whitepaper provides the perspective of what security leaders believe are important, and examples of metrics to consider when testing network threat detection products.