INTERNATIONAL INSTITUTE FOR ANALYTICS™

DISCUSSION SUMMARY
RESEARCH & ADVISORY NETWORK

Stronger Cybersecurity Starts with Data Management

CHRISTOPHER SMITH
Director of Cybersecurity Strategy, SAS

EVAN LEVY
Vice President of Business Consulting, Data Management, SAS

May 2016

Interviewed by Robert Morison, IIA Lead Faculty

Discussion Overview

A paradigm shift is underway in the cybersecurity industry. Cybersecurity professionals are moving from a focus on attacker prevention to attacker detection. Preventing the "bad guys" from getting in is still important, but cyber adversaries are increasingly able to bypass even the most sophisticated network defenses. Once inside, it is more important than ever to find these attackers fast, before their activities get buried in the daily volume and pulse of network communications. This is where security analytics holds promise. Security analytics provides the necessary and timely visibility into normal and abnormal network behavior. This visibility enables devices and entities acting suspiciously to be quickly identified and investigated.

What security leaders and security teams sometimes fail to appreciate, however, is the degree to which security analytics is fundamentally a challenge in data management. To explore the role of data management in security analytics, IIA talked with experts from SAS - Christopher Smith, Director of Cybersecurity Strategy, and Evan Levy, Vice President of Business Consulting, Data Management.

Why is cybersecurity a data management challenge?

Chris: During the course of business, organizational networks may generate petabytes of data per second from the normal communications of connected devices. And, with the growth of mobile devices, sensors, and cloud-based services connecting to these networks, the volume of this data will only increase – and so will the attack surface.

To protect themselves, organizations have deployed a variety of security tools – by one count, an average of 67. Yet each tool looks at only part of the picture, and each has its own data and reporting formats. Each may also generate an alert when suspicious interactions occur. Even with traditional tools designed to consolidate and perform basic correlations on this data, security teams are still buried. They simply can't keep up with all the data, all the detail, all the alerts they receive. The data management challenge isn't just to capture and store all the content, but to make it easy for security analysts to navigate data and investigate activity. It's impractical to expect security analysts to understand all of the log files and data formats generated by all of their tools. The analyst should be able to focus on what's happening and why, without having to double as a data scientist.

This is where security analytics holds promise. Information security professionals are under the gun to react quickly to the constant barrage of cyberattacks. The technologies for "big data" security analytics seem to have come along just in time. But this isn't just about data volumes and velocity. It is also about handling the data variety, and using advanced security analytics to focus and drive rapid incident response.

Evan: Effective security analytics requires a steady supply of integrated data. This data is going to be structured, as in network traffic flow data, and semi-structured, as in security event descriptions. It will need to be drawn from multiple internal systems and external sources. That doesn't necessarily mean moving all of the data, storing it in one place, or getting all of it perfect ahead of time. It does mean

having the interfaces and processes for pulling the data together for analysis.

Getting the right data can be a challenge from the start. It's often assumed that security operations know all of the systems containing the data they need, has access to those systems, and is familiar with all of the data elements and naming conventions. That's not always the case. These systems may have log and data files that are uniquely formatted. The formats may be particular to a category of systems or to a particular vendor. Moreover, field names, values, and meanings can vary across systems.

Let's take network traffic flow data as an example. Network traffic flow is a key data component for security analytics because it offers a rich information baseline for identifying anomalous network behavior. Most often, however, this data is under the purview of the networking team, not the security team. It requires coordination to understand if the routers and switches are currently generating these records, and how those records are being generated. There could be a one-to-one or one-to-many relationship between traffic flow record and network packet generated. This sampling ratio can ultimately impact the accuracy of the analytic results.

Given the magnitude of this data management challenge, how should organizations break it down?

Evan: There are quite a few pieces to the puzzle, but it's useful to put them in two groups—getting a handle on the data sources, and then making the data accessible for use. The data sourcing challenges are:

- **Naming** – The data coming from all those different devices, systems, and monitoring tools have different names and formats that need to be matched and reconciled. This is Data Management 101, and it's got to be done.
- **Number** – Complicating the reconciliation problem is the fact that the number of sources is large and likely to grow and change. If your infrastructure contains 30 security tools today, in six months it may be 35, and some may then be decommissioned in favor of more advanced replacements. It almost goes without saying that the data management architecture has to be flexible.
- **Structure** – Because cybersecurity data may be structured or semi-structured, it's essential for a security analytics solution to handle both. Security analysts need to be able to interpret what's happening without having to remember that Event Code 42 means one thing in System A and another in System B.
- **Storage** – The methods used for storage and access also vary across tools and devices, and all the source data has to be stored together in a manageable way. It's not just a matter of storing a large volume of content, but also of anticipating how you will use the data and what detail to store for how long.
- **Volume** – Security data is already large and growing with the proliferation of tools and network-connected devices. Not just the technology, but the whole cyber-analytics process, has to be able to scale to keep pace.

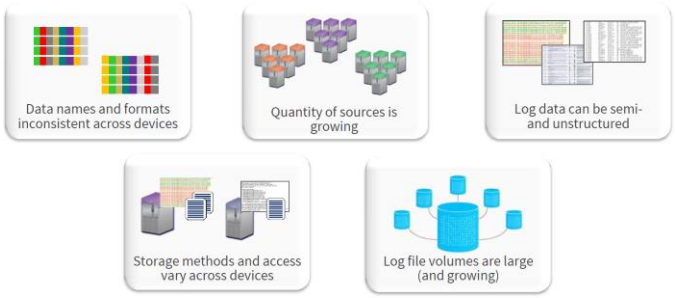

Figure 1. Cybersecurity Data Sources

Chris: Building on what Evan said, there may be data *quality* issues along with problems with data names and formats. It's really important to have device and user context enrich the network traffic data for more accurate analytic results. What does this mean from a data perspective? It may mean bringing in user authentication information or the functional role of the device in the organization. In many organizations, neither of these data sources are maintained with the accuracy and currency needed to generate quality results. To vary an old saying, "there's no such thing as a free lunch with data quality when it comes to cybersecurity."

Those are the data sourcing challenges. Let's look at the challenges with data access and use.

Evan: In its native state, a lot of cybersecurity data isn't ready for efficient use by security analysts. The problem manifests itself in several ways:

- **Format** – The data may not be formatted and structured to support ease of retrieval and analysis. The organizing principle is often time – when was the data created? Finding specific data can be like searching in a library where the books are shelved in the order they happen to have been acquired.

- **Query** – The analysts must be able to query the data, to drill down and look across data sources. They want to query large volumes of data without having to do endless data preparation. So the platform must be flexible. But most cybersecurity tools are more like transactional systems, organized to collect data and process it in predetermined ways.

- **Integration** – The underlying challenge is integrating the data from multiple sources so relevant data can be analyzed together. This needs to be as automated as possible for the sake of efficiency and consistency, because no two analysts would integrate the data the same way. Fortunately, most large organizations have tools, skills, and experience with data integration, but not necessarily – or not yet – within security operations.

- **Inventory** – We shouldn't assume that analysts, especially if they specialize in specific types of problems or investigations, know of all the data that's available. We're talking about perhaps thousands of available data elements, and dozens or hundreds relevant to a specific investigation. Analysts need active catalogs of data sources and data elements, including format, meaning, and known uses.

- **History** – Finally, it's essential to keep enough history accessible to analysts for purposes of establishing baselines and doing retrospective investigations. You don't want to store all the detail indefinitely, but if a breach is discovered 100 days after the fact, you want to be able to do some level of historical reconstruction.

Figure 2. Data Access and Usage

Chris: Let me emphasize two points here. The first is making things easy for the security analysts. It's great if some of them have deep data science skills, but we can't expect that. And we shouldn't ask them to spend a lot of time doubling as data integration and remediation and management specialists. Get the data management professionals involved in preparing an easily understood and productive data

environment. Let the security analysts focus on what they do best and where they add the greatest value – investigation and analytics.

That environment should bring together all the data, and it must be very flexible in terms of enabling a variety of analyses and accommodating new data sources. But it can't just be open-ended. As Evan mentioned, we need to anticipate how the data will be used in order to structure it for analysis. How big is the security analyst team, and how are they organized? What kinds of work are they going to be doing? Also keep in mind that some of the query capabilities have to be real-time. When spotting something that may look bad, the analyst may want to explore the what and why as quickly as possible.

The second is about what and how much data to store. It's important to consider the broader picture here. There may be certain sub-teams in the security organization that only review some pieces of information and don't retain others. But these other pieces – such as successful multiple logins on a specific device – may be of interest elsewhere. Without all of the necessary data, it will be difficult to determine how to shore up your systems and network for the future. Also, what and how much you keep may be dictated by industry-specific regulatory requirements or even corporate boundaries. For example, your legal department may want data to be kept for a shorter period than the security team. And, of course, your data history is going to be limited by existing hardware constraints.

We've transitioned into talking about solutions. What are the most important actions to take?

Evan: Keep in mind the goal that Chris just articulated: enable security analysts to be accurate and productive without being burdened by data management activities.

The first step is to adopt and adapt data management standards for cyber. Establish the "card catalog" of data sources and their content, with standards for formatting and naming and combining. Of course, the sources are going to change; it's a moving target. But this kind of "production support" is something large companies know how to do with their business applications and email systems and data center platforms. This isn't new – it just requires discipline.

You also need robust query tools and an analytics platform. The security analysts will make regular use of alerts and reports automatically generated by the source cyber tools. They will also want to fold in business context outside of cybersecurity log files and systems, for example, application-based security detail and usage content. That way, they can build the complete picture – not just that a breach occurred, but what it touched and what it did.

Chris: Among all the data management challenges we've discussed, the only things really different in cybersecurity are the subject matter and the specific business objectives of the analysis. The challenge is that data management hasn't historically been in the wheelhouse of information security professionals. These individuals, by nature, love borders. As a result, they've siloed themselves off from other areas in the organization. Data management practices and knowledge exist within other domains. Cybersecurity analysts should align with counterparts elsewhere in the organization. This way, they can accelerate cyber data management practices by importing successful techniques and leveraging lessons learned.

For obvious reasons, there's only so much transparency that can be maintained between security operations and other functional areas. But the cybersecurity arena has evolved beyond where hard and fast organizational borders should exist. Security

teams need to learn to become better data gatherers. In doing so, they will improve their data hunting capabilities.

Where do data management programs for cybersecurity fall short, and why?

Chris: They fall short when people treat data management as a destination, not a process. As we've said, cybersecurity is a moving target, and so is its data. Attackers are creative. Tomorrow's issues will be different from – or in addition to – today's. Data sources evolve and multiply, and analytical methods improve. A successful data management program for cybersecurity has to be in motion, comfortable with change, and able to cope with the surprises and challenges that will inevitably arise.

Evan: I see initiatives fall short of their potential because of skills issues. Collecting and storing data takes a different skill set from structuring the content to be queried, analyzed, augmented, and combined. Sourcing, accumulating, and storing raw content is very technology centric. Structuring the data for understanding, query, and analysis is very analyst-centric. Many organizations have a gap because that second skill set doesn't get the attention and staff that are necessary.

By way of wrap-up, what are the top three things to know and do about data management for cybersecurity?

Evan: First and foremost is to understand what analysts need to accomplish, not just what data they need. Meet the analysts where they are. Rich data and cutting-edge analytic tools go to waste if analysts lack the skills to use them. By the same token, sophisticated analysts shouldn't be constrained with rudimentary tools and everyday data management tasks.

Second, be ready for the number of data sources and the overall data volume to keep growing. Don't let the technological environment, staff capacity, or skills limit capability. You've got to handle whatever useful data comes your way. Cybersecurity is that important to the business.

Third, as we've emphasized, cyber is a moving target. Perfection isn't in the cards. The goal is to be as effective as possible without wasting resources – and to be prepared and flexible to meet the next new need, the next new challenge.

Chris: I'll add three more. One, even though you're working with an enormous amount and variety of fast-moving data, pay attention to its quality and usefulness. Two, don't just throw technology at the problem. Today's technology is powerful and versatile, but no data management technology is going to cover all your needs – or meet tomorrow's needs. Data management for cybersecurity remains a people-process-technology challenge. And three, you need a data management strategy to provide visibility into the network and overall computing environment, and to enable advanced analytics. Visibility and analytics are the best defenses we have.

Additional Information

To learn more about this topic, please visit sas.com/cybersecurity.

About the Interviewees

CHRISTOPHER SMITH

Christopher Smith is Director of Cybersecurity Strategy at SAS Institute. Chris has 22 years of information technology and security experience in the public and private sectors. Prior to joining SAS in 2010, he served as Chief Technology Officer for the US National Park Service and as the Lead Enterprise Architect for various US federal agencies, the Presidential Transition Team, and the White House. Chris holds a bachelor's degree in Information Systems from University of Maryland, and is finishing a master's degree in Cybersecurity Engineering there. He holds the CISSP, INFOSEC, C|CISO, CCSP, and CEH security and technology certifications.

EVAN LEVY

Evan Levy is an acknowledged speaker, writer, and consultant in the areas of Enterprise Data Strategy and Data Management. In his current role as Vice President of Business Consulting, Data Management, Evan advises clients on strategies to address business challenges using data, technology, and creative approaches that align IT with the business capability. Businesses are experiencing exponential growth in data volumes, sources, and systems – Evan offers practical real world experience in addressing these challenges in a manner that utilizes the company's existing skills, coupled with new methods to ensure IT and business success.

IIANALYTICS.COM