SANS Security Analytics Survey

Dave Shackleford
Today’s security event monitoring and correlation tools are being taxed more than ever before. The constant deluge of data being output by the many devices in our security infrastructures and input into log managers, SIEM and other security management interfaces creates quite a load on our analytics systems.

In the noise of all this security and event-related data, it is difficult to sort through and obtain real, actionable events that need attention. According to the last survey on Log and Event Management conducted by SANS in 2012, respondents indicated an inability to find events in their enterprises. This hadn’t been a problem in prior years, and they blamed this problem on too much operational, security, log and other data to sort through.

As a result of that survey, SANS decided to focus a new survey on the next generation of security tools—ones that process much larger data sets, are capable of deep-dive analytics, and rely more on intelligence than attack signatures.

Results of the current survey show that the market is in need of analytics and intelligence wrapped around the data that is being (and can be) collected in respondent organizations. In it, only 10 percent of respondents felt truly confident in their “Big Data” intelligence and analytics capabilities.

Their biggest impediment is in the process of collecting the correct data in order to make the necessary associations, followed by lack of vulnerability awareness and context. Yet these capabilities are important for a comprehensive detection and response system. The system should also be affordable and able to reduce manpower for strapped IT security departments.

---

1 www.sans.org/reading-room/analytics-program/SortingThruNoise
The 647 participants who took the survey represent a broad range of industries, organization sizes and IT security budgets.

The top three segments represented were government and military (19%), financial services and banking (17%), and a mix of consulting, cloud services, legal, and insurance included among others (16%), as shown in Figure 1.

The majority of respondents were affiliated with very large businesses (50,000 or more employees), and most of these organizations were international in scope. Mid-sized businesses came in a close second (2,000 to 4,999 employees, and 500 to 1,999 employees), and these were primarily domestic organizations in a single country (see Figure 2).

While the largest group (40%) are unsure of their IT security expenditures, those who did know indicated a low percentage of their total IT budget being spent on security. The next largest group (21%) spend less than 5% of their IT budgets on security. This is pretty standard in IT operations. Other studies indicate 2–3% of the IT budget being devoted to security is common.
On a positive note, another 16% of respondents spend between 6–10% on security. However promising, this represents the exception, because security cannot consume a large part of IT budgets or businesses will not be able to innovate. The market would be wise to respond to this trend by helping IT security teams with products and services that can provide a lot of “bang for the buck” in terms of providing better visibility, reporting and information aggregation. The full breakdown of budget allocation is shown in Figure 3.

The respondents themselves also represented a variety of job titles and management levels that are mostly staff positions (rather than being consultants). As shown in Figure 4, the largest group (56%) holds security administrator and security analyst titles, while 27% are in management positions ranging from manager to CSO/CISO, and another 17% work in senior IT operations roles (CIO/director/manager).

Other roles—network engineer, forensic professionals and privacy officers—were also represented. This indicates broad interest and involvement across IT operations, security and governance organizations seeking a better way to manage their security information.
Based on the results of the SANS 2012 Log Management Survey, 58% of respondents were using dedicated log management and 37% were using a security information and event management (SIEM) system.\(^2\)

Unfortunately, many “last-generation” log management and SIEM platforms can’t handle the deluge of data we are feeding them.

Therein lies the crux of the problem—we are literally drowning in data. Plus, we’re dealing with more unique and different types of data that require customized analysis, making the problem even worse. And, as if the situation wasn’t bad enough, we’re under enormous pressure to detect attacks and respond to them faster than ever before.

### Collecting and Processing “Big Data”

In a nutshell, any definition of “Big Data” analytics is really describing a set of tools and techniques that can store and process large data sets effectively and efficiently for a variety of purposes—particularly for analytics.

Adrian Lane of Securosis defines Big Data analytics as “a framework of utilities and characteristics common to all NoSQL platforms“ that includes the following (not a complete list):\(^3\)

- Handles very large data sets (Volume)
- Provides extremely fast insertion (Velocity)
- Manipulates multiple data types (Variety)
- Provides complex data analysis capabilities
- Uses distributed parallel processing

In this definition, many Big Data analysis architectures consist of two fundamental technology areas: distributed programming frameworks, such as Hadoop, and new types of nonrelational data stores, such as NoSQL. However, this definition does not necessarily reflect the growing spectrum of Big Data solutions, particularly for information security event monitoring and management platforms. In particular, there are a variety of different data storage and analysis toolkits emerging that do not rely on Hadoop and/or NoSQL that still allow for much more rapid and precise collection and processing of large data sets. In particular, the collection needs to be performed extremely rapidly and scale across multiple collection platforms that can handle large numbers of events.

As for processing, the large event count needs to be processed rapidly, and specific types of attributes need to be applied during processing (such as universal time stamping). In addition, security data analytics platforms should allow for rapid contextualization (sorting data into consistent fields and formats for analysis), as well as metadata extraction during the processing phase.

---

\(^2\) [www.sans.org/reading-room/analysts-program/SortingThruNoise](www.sans.org/reading-room/analysts-program/SortingThruNoise)

Analyzing

Large data sets—with technologies that enable rapid and accurate analysis, correlation and reporting—are the next stages for managing the data being collected for better proactive and reactive capabilities. Key elements should include the following:

- **Broader Security Event Data Types.** Security event monitoring uses a broad variety of data and data types, including logs and events from network security controls such as firewalls, intrusion detection systems (IDSs), intrusion prevention systems (IPSs) and traditional network monitoring tools that provide behavior and performance statistics. Host-based controls such as local firewall/IDS and other access controls are commonly included in security event data sets, as are a myriad of logs (including email and database servers, network devices and web content filtering, to name just a few). Antimalware controls and vulnerability assessment data are commonly included as well. This data typically resides in traditional relational databases, central log stores or some sort of SAN or NAS platform. While this is certainly a lot of data, we’re still missing a vast array of other IT data, especially data related to sensitive information, so organizations need even more data types for advanced correlation.

- **Correlation for Incident Response.** The benefits of advanced correlation with multiple data types are many: They can obtain system details; perform vulnerability scans; acquire time, date, and user information; and determine network attributes such as source and destination ports and addresses. But how can we look for events we don’t know about? This dilemma is one of the primary cases for additional analytic capabilities—one that would help us generate new correlation rules and discover new trends and behavioral patterns in the environment.

- **Analytics for Predictive Research and Root Cause Analysis.** Most security event management tools and activities have focused on detecting events happening right now and might develop some behavioral trends. But we’ve been sorely lacking in the ability to predict future trends based on current and past behavior in order to stop threats, rather than searching for them after the fact. This is one of the goals of security analytics. Analytics also ensures more data on which to perform root cause analysis and forensics after the fact, and can constantly update itself based on this analysis, thus leading to a more proactive monitoring approach.

- **Wider Scope of Analysis (Security Intelligence).** According to Gartner’s Joseph Feiman, Enterprise Security Intelligence is “the collection of data from all IT systems in the enterprise that could be security relevant, and the application of the security team’s knowledge and skill, resulting in risk reduction.” This intelligence often comes from the many platforms and endpoints from which user and event data is being collected and analyzed. However, there is also additional intelligence being built into security management systems by third parties that can provide external analysis of events that are, for example, unknown.

---

Output

Finally, any analytics solution needs to allow security analysts to make use of the processed and correlated data. Most analytics platforms provide a dashboard that allows for rapid (as close to real-time as possible) analytics queries, as well as longer-term forensic analysis of the data. Automated and rapid response capabilities that include triggered alarms on pattern matches, scheduled and manually-created reports, and flexible dashboard views are also typical.

So, where do we start to evaluate the current state of security event monitoring and determine where security analytics and Big Data can take us? We ask about organizations’ threat landscapes and how they’re collecting and sorting their security-related data to become more proactive in their security and response.
Advanced threats (often called advanced persistent threats, or APTs), are a primary driver for broader collection and analysis capabilities being required by organizations. APTs can involve multiple events happening across the enterprise that would otherwise not be connected without the use of advanced intelligence.

These advanced threats affected two-thirds of respondents in the past two years. Survey comments indicate that experiences differed widely, with one respondent seeing almost 500 attacks in a nine-month period, and many seeing somewhere between one and 20, as illustrated in Figure 5.

One respondent’s organization was notified of malicious behavior by an outside party. This number is much higher in the Verizon Data Breach Investigations Report (which found that 69% of intrusions were first detected by external parties).5

There are several interesting questions that arise from this result set. First, what separates the organizations who are detecting only a few attacks from those who were detecting hundreds? This ability to detect attacks could be related to the tools and personnel in place, the overall event monitoring maturity, the size of the organization or simply the volume of attacks. Also, for the 33% of organizations that did not detect any attacks, the question becomes: Are they really not experiencing any attacks, or are they simply missing the indicators?

---

5 www.verizonenterprise.com/DBIR/2013
**Duration of Attacks**

When asked about the effectiveness of security tools and the time it took to detect the presence of malicious code, including the shortest, longest, and average length of time before detection, we received some interesting results:

- **Shortest time to detect**—Over half of respondents (198 of 374 who had experienced attacks) indicated they were able to detect the threat in the same day; 63 indicated they found the threat in one week or less; and several others took between one and 10 months to detect.

- **Longest time to detect**—Of the 353 respondents to answer this question, 24 indicated more than 10 months were needed to detect and respond to advanced threats; 37 said the longest time was a day; 80 indicated a week or less; and 55 said one week to one month was the longest time on record for attack detection.

- **Average time to detect**—Of the 345 respondents to answer this question, 72 averaged one day for attack detection; 105 took one week or less on average to detect; and 113 respondents did not know the average detection time.

What do these responses indicate? Currently, it appears that many organizations feel they are detecting threats fairly rapidly. However, there’s nothing to indicate that they are detecting all the threats, and there are likely a number of different tools coming into play here. For example, some organizations may be relying on antivirus tools to provide alerts and blocking/quarantine actions, which would indicate a “real-time detection,” but anything beyond the same day may seem to indicate that the threats were not detected by signature-based tools of this nature.

**Time to Respond**

The majority of respondents who were able to identify the time involved in response to an attack said the shortest time to respond to a detected attack was within the same day. One week or less was the most common response for the longest response time, and this was also consistent with the average stated time to detection. Based on these results, with the average detection being one week, and the average response time after detection being another week, the window of opportunity for attackers is two weeks long. Yet, according to several reports, the window between initial attack and exploit has narrowed to just hours.
Impediments to Response

When asked about the primary impediments to discovering and following up on attacks, the number one reason cited was the lack of ability to gather appropriate security and operational data with which to make proper associations, as shown in Figure 6.

The next most frequently cited impediments were a close tie between lack of system and vulnerability awareness, and lack of relevant “normal” behavioral baselining.

A number of other reasons were cited, ranging from the aforementioned problems to lack of skills in incident detection and response, difficulties with normalization, and lack of available intelligence data.
When asked to tell us which data types they were collecting now, planning to collect in the next 12 months, and not planning to collect at all, some respondents didn’t know the answers. However, the majority had clear answers.

The majority of teams collect server, application, and endpoint logs, as well as network device and vulnerability scanner events and logs, according to responses. Quite a few collect some types of identity data as well, and many also plan to collect such data over the next 12 months, as shown in Table 1.

<table>
<thead>
<tr>
<th>Answer Options</th>
<th>Currently collect</th>
<th>Plan to collect within 12 months</th>
<th>Don’t plan to collect</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log data from network (routers/switches) and servers, applications and/or endpoints</td>
<td>77%</td>
<td>15%</td>
<td>2%</td>
<td>4%</td>
</tr>
<tr>
<td>Monitoring data provided through firewalls, network-based vulnerability scanners, IDS/IPS, UTM, etc.</td>
<td>77%</td>
<td>16%</td>
<td>1%</td>
<td>4%</td>
</tr>
<tr>
<td>Access data from applications and access control systems</td>
<td>45%</td>
<td>27%</td>
<td>12%</td>
<td>12%</td>
</tr>
<tr>
<td>Unstructured data-at-rest and RAM data from endpoints (servers and end-user devices)</td>
<td>12%</td>
<td>19%</td>
<td>39%</td>
<td>24%</td>
</tr>
<tr>
<td>Security assessment data from endpoint (aka from NAC/MDM scans), application and server monitoring tools</td>
<td>37%</td>
<td>29%</td>
<td>15%</td>
<td>16%</td>
</tr>
<tr>
<td>Assessment and exception data (not on the whitelist of approved behaviors) taken from mobile/BYOD endpoints (aka from NAC/MDM scans)</td>
<td>14%</td>
<td>25%</td>
<td>30%</td>
<td>24%</td>
</tr>
<tr>
<td>Monitoring and exception data pertaining to internal virtual and cloud environments</td>
<td>22%</td>
<td>27%</td>
<td>22%</td>
<td>24%</td>
</tr>
<tr>
<td>Monitoring and exception data pertaining to public cloud usage</td>
<td>14%</td>
<td>17%</td>
<td>37%</td>
<td>28%</td>
</tr>
<tr>
<td>Other</td>
<td>3%</td>
<td>2%</td>
<td>4%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Table 1. Current and Planned Data Collection for Monitoring

Nearly 30% of respondents plan to collect security assessment data from endpoints, mobile device security data, and monitoring and security data related to virtualization and cloud infrastructures over the next 12 months. Most do not plan to collect (or aren’t sure about their plans to collect) unstructured data from servers and endpoints, including RAM.
How They’re Collecting and Analyzing Data

Nearly half of the respondents have dedicated log management platforms, SIEMs, and scripted searches as part of their data collection and analysis processes. Fewer than 10% use unstructured data repositories and specific Big Data frameworks for analysis and search, as shown in Figure 7.

This indicates an early stage in the market, where most organizations haven’t yet put advanced analytics systems to real use in any comprehensive, integrated fashion. With the combination of overwhelming data volumes and challenges in gathering and correlating operational and security data, respondents clearly need an integrated way to organize their reporting data. From our most recent log management survey, a dedicated log management platform is not meeting the needs created by Big Data collection. With the second highest number of responses, it appears that SIEMs are most poised to step into that void, however dedicated security analytics vendors may also fill the space where “manual” and “not automated” systems are in use.

Figure 7. Types of Security Data Analysis Tools in Use
Intelligence

With more data available for analysis and better data context, we might discover that we have more indicators of compromise in the environment than we thought. Threat intelligence data can be of use, and there has been much discussion in the information security field about the need for threat intelligence in most organizations today. SANS survey respondents indicated a broad range of attitudes toward threat intelligence collection and use when evaluating security event data.

Nearly 40% of respondents said they don’t use external threat intelligence tools; 32% indicated that internally-gathered threat intelligence data plays a role in event analysis; and 27% have threat intelligence data provided by external third parties. Almost the same number (26%) rely on analytics tools to do the heavy lifting of threat intelligence for them; and others leverage SIEM platforms and manual techniques, as shown in Figure 8.

This was a choose-all-that-apply answer set, so these results aren’t surprising. The threat intelligence capabilities of many organizations are still immature, and those that are developing their capabilities in this area are using a variety of techniques and platforms. One of the key areas that the market will need to develop is standards for threat assertions based on data collected and a common syntax for alerting and reporting this data.

Intelligence and Visibility in Action

With more intelligence, we can answer the important questions: Who is attacking us? What trends are others seeing in the wild? Are we specifically being targeted, and if so, how?

With more intelligence data to work with, security teams can analyze the environment based on the following:

- Timing of events
- Sequences of occurrence (and time)
- Differences in data from various sources
- Real statistical analyses (such as scatter plots)

For example, security analytics could enable more advanced searches for indicators of compromise, such as the following:

- Phishing in mail logs, by using trends and correlation capabilities to assess the affinity of senders to the organization
- Slow data exfiltration in proxy/firewall logs, by examining the number of bytes and sessions over time
- HTTP-based malware command and control channels (C&C) in web proxy logs, by looking for long URLs without Referrer fields and Base64 POST variables in the headers
- Insider threats, by looking for trusted insiders abusing their access for malicious purposes

---

7 http://csrc.nist.gov/cyberframework/rfi_comments/040813_si_organization.pdf
**Need for Speed**

Users of any tools developed also need the ability to sort the data and generate alerts based on policies in real-time or as close to real-time as possible. This means a lot of data must be processed, sorted and analyzed.

Of the 27% of respondents that actually measure aggregate volumes data that needs to be processed during average and peak events, volumes ranged from 1 to 99 gigabytes (GB). However, the next largest volume during peak event times was 100-499 GB; followed by more than 1 terabyte.

**Tools and Services**

When asked what specific solutions they used for advanced analytics and intelligence, many answered with a variety of vendors in a number of product and service categories, ranging from SIEM, to log management, to third-party threat intelligence services. Security teams are well on their way to leveraging more data for security event monitoring and management using a combination of home-grown and vendor solutions.

The use of threat intelligence data in correlation efforts will likely increase over time as organizations search for additional correlation capabilities to find suspicious activity and connect the dots. As an additional data point, 51% of respondents said they are currently using third-party intelligence services; 39% are not; and 10% weren’t sure (see Figure 9).

How are security teams automating their pattern recognition within event management and data analytics platforms? Currently, many simply are not. Almost 36% of respondents are acting on “hunches.” Still, nearly 50% are “fairly automated” using a variety of tools both internally and from external providers; and 9% are “highly automated” in their ability to recognize malicious patterns of activity and behavior. See Figure 10 for more detail.
This, again, indicates a lack of maturity in the market. Some report that they are working with intelligence services, SIEMs and log managers, but only 9% say are satisfied with their pattern recognition process.

This is supported by their answers to another question, in which the majority of respondents are using less than 25% of their solution’s capabilities as measured by the number of inquiries for both active monitoring (i.e., detecting events in progress) and reporting (i.e., forensics and follow-up). As the use of the analytics platform increases, so does the usage for reporting and forensics (see Figure 11).
**Overall Satisfaction**

When we asked survey respondents how satisfied they were with their current analytics capabilities, we got a wide range of responses. In general, teams are satisfied with storage capabilities and data access, performance and response time, and reduction of false positives. Some interesting miscellaneous write-in responses were received, including one that said “This appears like the up-and-coming spending bandwagon,” and another that fell behind in IDS platform updates, putting the organization behind the curve in its analytics capabilities.

Areas that seem to need improvement include reporting, context, training/intelligence expertise, integration and cost, as shown in Figure 12.

![Figure 12. Satisfaction with Current Analytics Platforms](image-url)
Respondents also told us where they benefited most from using security analytics and advanced threat management systems. Combating external malware-based threats was the top selection, followed by detection of advanced threats and compliance monitoring. Other selections included insider threat identification and risk and compliance monitoring (see Figure 13).

Focusing on external malware and advanced threats, therefore, makes sense. Many organizations use their monitoring platforms for compliance and risk monitoring as well, so this result is not surprising either. The response also indicates some level of security data analytics and intelligence. Some organizations are getting relief from advanced threats that are externally based, although fewer than 50% are reporting experiencing benefits.
Given the advanced nature of attacks discussed at the beginning of this paper, security teams are feeling the pressure to detect and react more quickly to events that are harder to find within their enterprises.

Building better correlation rules, reporting on historical data, and developing more effective baselines will likely become more important in the future. But, adoption of these more advanced analytics functions will take time.

When asked about plans for future investments in security analytics and Big Data platforms, most organizations are still focused on the fundamentals—better SIEM, more training to detect patterns of malicious activity, vulnerability management and network protection tools, and endpoint visibility. Many are considering dedicated analytics engines (23%) and intelligence products and services (26%) at the moment, as shown in Figure 14.

What are the next steps toward fully realizing the benefits of Big Data and security analytics? We need to collect more data, process it more quickly, and query it for more intelligent and contextual visibility into associated actions that create events.

New tools are now emerging that are up to the task. Most organizations, however, are just beginning to evolve from traditional log management and security event monitoring to more advanced security analytics and Big Data processing.
Dave Shackleford is the owner and principal consultant of Voodoo Security and a SANS analyst, senior instructor, and course author. He has consulted with hundreds of organizations in the areas of security, regulatory compliance, and network architecture and engineering. Dave is a VMware vExpert with extensive experience designing and configuring secure virtualized infrastructures. He has previously worked as CSO for Configuresoft, CTO for the Center for Internet Security, and as a security architect, analyst and manager for several Fortune 500 companies. Dave is the author of the Sybex book *Virtualization Security: Protecting Virtualized Environments*, as well as the coauthor of Hands-On Information Security from Course Technology. Recently Dave coauthored the first published course on virtualization security for the SANS Institute. Dave currently serves on the board of directors at the SANS Technology Institute and helps lead the Atlanta chapter of the Cloud Security Alliance.
<table>
<thead>
<tr>
<th>Course Name</th>
<th>Location</th>
<th>Dates</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>SANS Amsterdam August 2020 Part 1</td>
<td>Amsterdam, NL</td>
<td>Aug 03, 2020 - Aug 08, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS FOR508 Canberra August 2020</td>
<td>Canberra, AU</td>
<td>Aug 17, 2020 - Aug 22, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Amsterdam August 2020 Part 2</td>
<td>Amsterdam, NL</td>
<td>Aug 17, 2020 - Aug 22, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Virginia Beach 2020</td>
<td>Virginia Beach, VAUS</td>
<td>Aug 30, 2020 - Sep 04, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Philippines 2020</td>
<td>Manila, PH</td>
<td>Sep 07, 2020 - Sep 19, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS London September 2020</td>
<td>London, GB</td>
<td>Sep 07, 2020 - Sep 12, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Baltimore Fall 2020</td>
<td>Baltimore, MDUS</td>
<td>Sep 08, 2020 - Sep 13, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Munich September 2020</td>
<td>Munich, DE</td>
<td>Sep 14, 2020 - Sep 19, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Network Security 2020</td>
<td>Las Vegas, NVUS</td>
<td>Sep 20, 2020 - Sep 25, 2020</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS OnDemand</td>
<td>OnlineUS</td>
<td>Anytime</td>
<td>Self Paced</td>
</tr>
<tr>
<td>SANS SelfStudy</td>
<td>Books &amp; MP3s OnlyUS</td>
<td>Anytime</td>
<td>Self Paced</td>
</tr>
</tbody>
</table>