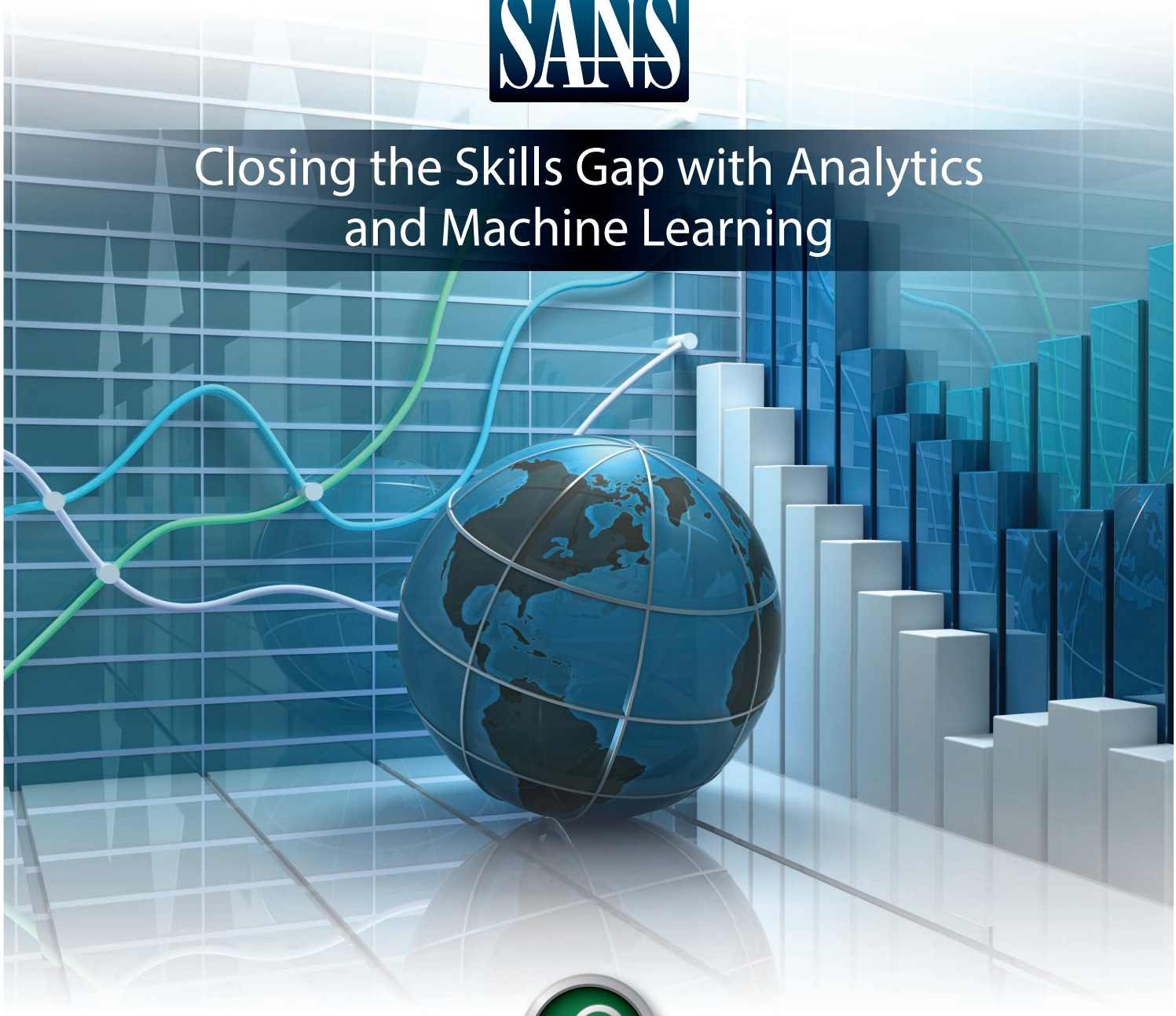

Closing the Skills Gap with Analytics and Machine Learning

A SANS Product Review

Written by Ahmed Tantawy

October 2017

Sponsored by
RSA

Executive Summary

Over the past few years, there has been a lot of discussion about a shortage of skilled professionals in the security field. A report from Frost & Sullivan and (ISC)[2] estimates there will be more than 1.5 million unfilled cyber security positions across the globe by 2020.[1] Shortage of skills is the top impediment to successful detection and remediation, followed by lack of management support, according to SANS surveys on cyber threat intelligence (CTI) and security analytics.[2, 3]

While the skill gap widens, attackers continue to find new ways of evading current security controls. For example, the latest trend is the use of file-less malware, a malicious software that exists only in memory, without any software on the file system, which is more difficult to detect. Another example of new attack techniques surfaced in April, when the Shadow Brokers hacker group leaked tools developed and used by the U.S. National Security Agency, including several zero-day exploits that targeted enterprise firewalls, antivirus products and Microsoft products. The NSA tools were also used in new exploits, including the WannaCry ransomware attack in May 2017, which again targeted computers running Microsoft Windows operating systems, encrypting their entire contents and then demanding ransom payments.

The specialized analytics skills needed to catch these sophisticated attacks in progress are hard to come by, leaving organizations vulnerable to threats that can linger and persist without IT's knowledge. The only way to close this gap is through automation. With this unfair advantage, it is important that IT departments leverage automated analytics and machine learning solutions that connect the dots between seemingly random events and provide much-needed context, visibility and actionable advice. However, while most organizations using security analytics report better visibility into events, they also feel they need much more integration and automation across systems to fully realize the capabilities of intelligence and analytics. In the 2016 SANS survey on security analytics, only 4 percent considered these capabilities to be fully automated.[6]

In this paper, we explain how to utilize and integrate analytics and machine learning to reduce the load on security professionals, while increasing visibility and accurately predicting attackers' next steps.

CYBER THREAT INTELLIGENCE

The SANS Forensics Cyber Threat Intelligence course defines CTI as the "collection, classification and exploitation of knowledge about adversaries."[4] This includes, in particular, information about adversaries' tactics, which can help targets detect and block the attackers. As one of the course's primary authors describes it, "CTI is analyzed information about the intent, opportunity and capability of cyber threats."[5]

[1] "(ISC)[2] Study: Workforce Shortfall Due to Hiring Difficulties Despite Rising Salaries, Increased Budgets and High Job Satisfaction Rate," http://blog.isc2.org/isc2_blog/2015/04/isc-study-workforce-shortfall-due-to-hiring-difficulties-despite-rising-salaries-increased-budgets-a.html

[2] "Cyber Threat Intelligence Uses, Successes and Failures: The SANS 2017 CTI Survey," March 2017, www.sans.org/reading-room/whitepapers/analyst/cyber-threat-intelligence-uses-successes-failures-2017-cti-survey-37677

[3] "SANS 2016 Security Analytics Survey," December 2016, www.sans.org/reading-room/whitepapers/analyst/2016-security-analytics-survey-37467

[4] "FOR578: Cyber Threat Intelligence," www.sans.org/course/cyber-threat-intelligence

[5] "Cyber Threat Intelligence Uses, Successes and Failures: The SANS 2017 CTI Survey," March 2017.

[6] "SANS 2016 Security Analytics Survey," December 2016.

Lighten the Load

Security operations centers (SOCs) have it tough these days. They are performing the multiple functions involved in prevention, detection, response, remediation, vulnerability management and compliance—all with two to five full-time SOC employees, according to the 2017 SANS SOC survey.[7] The survey found that most of them still rely heavily on IDS and logs, as well as manual processes for analysis and security metrics. Here are some challenges with that approach:

Focusing only on logs and IDS to break down and analyze information is not the best game plan to distinguish and detect sophisticated attacks.

- **Inability to connect the dots and see attacks in progress.** By leveraging machine learning-based techniques or knowledge-based techniques, such as rule-based pattern matching and statistical anomaly detection, SOC staff get better at discovering previously unknown threats that are often overlooked by just analyzing the logs.
- **Need for visibility into more than just logs.** Network packet data, for example, is also needed to detect the movement of sensitive data or command and control channels. Aggregated intelligence from third-party and internal sources of intelligence—such as reporting and alert data from security devices—is also important.
- **Having to search across multiple platforms and tools that don't scale.** Another challenge is IT staff don't have time to shift between platforms and attempt to connect the dots themselves. They need a single platform with one view they can pivot from without having to log in and out of different systems.
- **Inability to scale with manual correlation.** Statistical analysis is another critical function that should be automated, particularly when security analysts need to react quickly, determine whether the incoming alerts are real threats or not, take appropriate actions against those threats and repair exploited vulnerabilities.

[7] "Future SOC: SANS 2017 Security Operations Center Survey," May 2017, www.sans.org/reading-room/whitepapers/incident/future-soc-2017-security-operations-center-survey-37785

CTI Beyond the SOC

Many SOC functions are enabled and improved through the use of CTI. According to the 2017 SANS CTI survey, organizations are using intelligence data to not only support their security operations, but also for awareness, threat and vulnerability management, IT operations, and budgeting. See Figure 1.

How is CTI data and information being utilized in your organization?
Select all that apply.

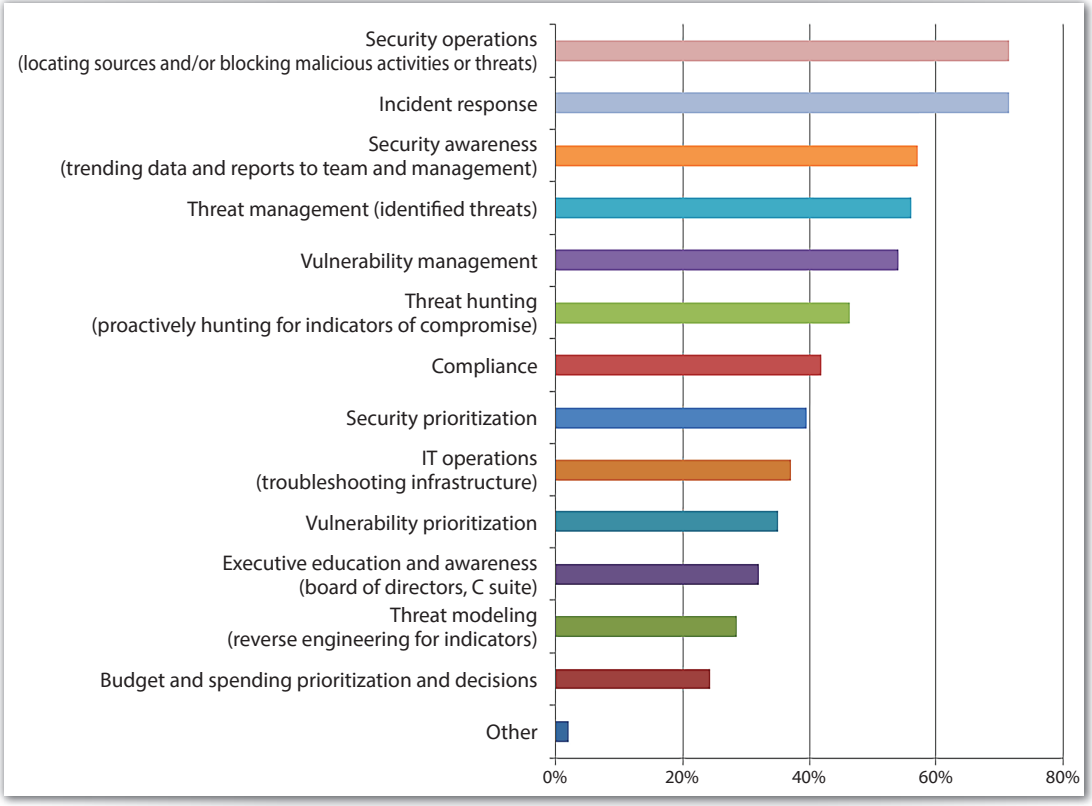

Figure 1. Multiple Uses for CTI[8]

[8] "Cyber Threat Intelligence Uses, Successes and Failures: The SANS 2017 CTI Survey," March 2017, www.sans.org/reading-room/whitepapers/analyst/cyber-threat-intelligence-uses-successes-failures-2017-cti-survey-37677, Figure 7.

Analytics and Machine Learning Defined

At its core, analytics and machine learning are mathematical techniques implemented on computer systems to enable information mining, pattern discovery and detection of connections between patterns, as well as the ability to draw inferences from the aggregated, analyzed data.

ANALYTICS
The process of examining large, varied datasets to uncover information to help organizations make well-informed business decisions.

Machine learning can be supervised (inferring a function from labeled training data) or unsupervised (developing and modifying the behavior model without owning a previous model) through constantly analyzing available data. Supervised machine learning can identify a threat almost immediately without knowing anything about the threat or environment; unsupervised machine learning needs to learn the local context of what is "normal" in order to identify what is not normal.[9]

MACHINE LEARNING
One aspect of artificial intelligence: Algorithms and processes "learn" by being able to generalize past data and experiences to predict future outcomes.

Integration with Workflow

Automated workflows enable quick and informed decision making and eliminate the potential panic factor experienced by security analysts when they get too much security-related data that is not connected, correlated or pre-analyzed. This lack of context can cripple or delay response in a high-pressure situation.

Workflows adhere to a process and enable dependable decision making. During an incident investigation, the requirements of each threat change as the situation unravels. Because the workflow reacts to both new data detected through automation and input from the security analyst, it is imperative to integrate both sources into a workflow in critical situations.

By using workflows and automated analytics, organizations can improve the experience for security analysts and transform the environment from reactive to proactive. Automated analytics should handle the tasks outlined in Table 1 on next page.

[9] "The Expanding Role of Data Analytics in Threat Detection," October 2015, www.sans.org/reading-room/whitepapers/analyst/expanding-role-data-analytics-threat-detection-36362

ADVICE

Threat levels can change rapidly, and new events can change the threat severity. Decisions need to be based on the very latest information. Integrated workflow, with new threat data delivered in near real time, is critical because attackers can damage the environment within a short time.

Table 1. Top 10 Industries Represented

Tasks to Automate	How It Helps
Automate manual tasks, such as IDS alert reviews.	Focusing on events rather than logs and alerts provides visibility into multiple actions that can constitute a single event.
Correlate large volumes of security and operational data from disparate sources, regardless of original format.	Knowing what is good and comparing that to unknown and unapproved behaviors through automated correlation and machine learning provides greater accuracy.
Present that data in a way that is relevant to the organization.	This saves time in searching through threats that don't matter because the organization has no vulnerability to those threats.
Assist with vulnerability management.	New threats may involve new vulnerabilities the organization didn't know about. When vulnerability data becomes part of the workflow, all potentially impacted systems can be patched or updated.
Provide a single view for detection and response, with ability to pivot and drill down easily.	With so many disparate tools, being able to click on a pivot chart, link or icon for automatic drill-down is a critical time-saver for understaffed teams.
Continuously improve attack prevention.	This is the sweet spot of machine learning: Once a new threat action is detected, future instances are automatically blocked and vulnerabilities are continuously monitored.

Integrating and automating workflow as much as possible helps teams assign, triage, investigate and remediate incidents in a quick and efficient manner.

Attackers leverage different techniques to infect desktops, laptops, servers and mobile devices, including watering holes, phishing and spearphishing scams, and malicious websites that infect user endpoints through their browsers.[10] Because so many organizations have a mobile workforce and BYOD policies, it is no longer sufficient to rely on only traditional, signature-based products for endpoint threat protection. With so many attack vectors focusing on endpoints, organizations are turning to endpoint behavioral analytics.

Detect the Unknown

When faced with progressively advanced, custom malware and zero-day attacks, organizations need advanced analytics for detecting attack behaviors (rather than signatures) embedded in their endpoint monitoring solutions. Endpoint intelligence should be able to detect malware activity and malicious behavior tied to zero-day attacks. It should also detect other indicators of compromise, including the following:

- Local security override host access attempts
- Unauthorized local accounts
- Misconfigurations
- Abnormal process activity
- Multiple firewall changes
- Other changes in the endpoint ecosystem

Intelligence should also apply to hosts running workloads in the cloud, including cloud-specific threat and vulnerability information, and prevent attacks on those systems as well. Intelligence can also be applied on the network. For example, it could look for:

- Traffic patterns between machines that shouldn't talk to each other
- Port-to-port activity
- Encrypted packet activity
- Outbound activity

[10] "Next-Gen Endpoint Risks and Protections: A SANS Survey," March 2017, www.sans.org/reading-room/whitepapers/analyst/next-gen-endpoint-risks-protections-survey-37652

Consume, Consolidate and Correlate

The framework must ingest and analyze an overwhelming amount of network and security device data generated by endpoint, cloud and networking systems, and third-party intelligence providers. The data must then be parsed and made available for real-time automated threat detection to aid security analysts in making decisions quickly and with more accuracy.

Security event correlation (or rule-based correlation) is the most well-known and used form of data analysis. Event correlation refers to the task of creating a context by revealing relationships between different events received from various data sources. A context can be bound by time, heuristics and asset value.

Look for the Unknown

Another approach to use in conjunction with event correlation is anomaly-based correlation, which requires a static profile of your environment to establish a baseline. After you have a baseline, the SIEM system can identify activity patterns that deviate from the baseline, alerting on potential security incidents. Profiling an environment typically generates multiple baselines, such as the following:

- **Traffic rate baseline—**Average events per second (EPS) and peak EPS per day of the week
- **Network baseline—**Protocol and port usage per day of the week
- **System baseline—**Monitoring average and peak CPU and memory usage; average number of running services; and user login attempts per day of the week

This data is often correlated with other sources collected from logs or a SIEM system looking for patterns and anomalies.

Compare Against the Known

Detection tools based on unsupervised machine learning can complement other controls already in the organization that provide information on what is known about users, their behaviors, applications, systems and usage habits. They can also be used in conjunction with assessment and vulnerability management systems to ensure baseline protections are in use and determine whether those systems are deviating from those baselines. Rather than administrators telling the system what to learn, it should automatically learn the baseline state of the environment and alert on what is abnormal—down to the level of a specific identity. The behavior of that identity can be mapped across physical locations, endpoints, and mobile and wearable devices, allowing quick containment.

Never Stop Learning

In addition to rapid event detection, correlation and response, security analysts must be able to predict future trends based on past and current behavior. This is where security analytics and machine learning really shine. Analytics and machine learning provide security analysts and responders with more data and historic information from the environment over a long period of time, helping them identify the trends and emerging patterns to accurately predict future indicators and attacks. External threat intelligence feeds should provide a lot of value here. Input from the real-world threat landscape always provides a big advantage for the security data analysis model.

Integrate with SOC

ADVICE

Using analytics and machine learning, analysts create and work from baselines that help identify new and abnormal patterns for faster detection and remediation of known and unknown threats.

Many SOCs—31 percent, according to the SANS 2017 SOC survey[11]—have started integrating threat intelligence into their full operations to provide their security and operations teams with current, relevant information on the latest threats in the real world. When complementing that insight with asset information and vulnerability intelligence, automated risk-based analysis (and computer-aided human analysis of threats) becomes much more achievable. All these forms of analysis should be packaged and used together to help categorize and prioritize alerts and help analysts and defenders focus on high-priority threats instead of chasing false positives.

Here are some practical ways intelligence-driven security improves the speed and efficacy of detection and response:

- Logs inside the network and at endpoints provide immediate visibility into the activity in the internal network.
- Analytics from different data sources, including third-party intelligence platform providers, improves the odds of finding unknown threats instead of analysts having to guess and manually search for indicators.
- Automated analytics and correlation helps administrators take informed actions.
- Workflow integration makes it easier for analysts and responders to work as a team—to share the same information, monitor thoroughness of response and detect future similar attack methods. This augments and teaches signature-based security devices through behavior analysis in the network and on the endpoint.

[11] "Future SOC: SANS 2017 Security Operations Center Survey," May 2017, www.sans.org/reading-room/whitepapers/incident/future-soc-2017-security-operations-center-survey-37785, p. 19.

Establish Requirements

Basic requirements for an intelligence framework should include integrating intelligence feeds, centralizing the data into a single view for security, response (even operational IT groups) as needed, and automating and integrating as much of this workflow as possible for all of those involved in detection, investigation, response, remediation and reporting.

Here are some basic requirements to consider when setting up or expanding your intelligence and analytics programs:

- Know who will need intelligence data, how they will use it and for what purposes.
- Know what you are going to monitor and how; monitor the network and endpoints continuously and exhaustively.
- Include the ability to perform full packet capture on the network and behavior-based threat detection and analysis on the network and hosts.
- Include sophisticated analytics techniques that can ingest huge amounts of data, such as network flow and network traffic, in near real time.
- Ensure the ability to detect, analyze, report and educate on malicious behavior on never seen before malicious software by relying on the actual behavior or hashes of the executables and comparing it against external threat intelligence.
- Include workflow automation among security, response and operational groups so security analysts can spend more time investigating emerging threats and less time on routine tasks that can be automated.

Using this type of approach, organizations move from a reactive security operations framework to a proactive one, based on risk scoring, threat management, continuous monitoring, automated workflow, and more informed incident response and threat hunting capabilities.

Increase Effectiveness, Reduce Stress

A proactive security operations framework streamlines day-to-day security processes and increases security effectiveness by instituting a defense-in-depth model with analytics and machine learning that detects advanced threats at each stage of an attack. This operational framework functions best under one platform that reduces all the security alerts to actionable intelligence rather than overwhelming security analysts with myriad, seemingly unrelated, alerts. Sharing of knowledge-based expertise will immensely improve the response of security analysts in terms of both efficiency and the correctness of their judgment. Moreover, global threat intelligence sharing through advanced persistent threat (APT) and threat intelligence portals and integration of those sources in SIEM systems provides unique, proactive insights into the motives and intentions of attackers in the real world. This enables organizations to proactively adjust policies, invest in security planning and predict future indicators of compromise or future attacks.

A proactive operations framework functions best under one platform that reduces all the security alerts to actionable intelligence rather than overwhelming security analysts with myriad, seemingly unrelated, alerts.

Look to the Cloud

According to the 2017 SANS survey on CTI, organizations are integrating intelligence feeds into their defense and response systems through vendor-provided or home-written APIs, threat intelligence platforms and intelligence service providers.[12] And most are using their SIEM systems for data processing.

Because most organizations are short on security and response skills, and because it takes massive resources to correlate so much intelligence and threat data, it makes sense to use cloud-based services to provide these capabilities. Lack of trained staff was cited as the top inhibitor to CTI satisfaction in SANS 2017 CTI survey, followed by lack of funding and lack of time to implement new intelligence processes.[13] This is another reason cloud-based intelligence and analytics make good sense.

Advantages of using a cloud or hybrid cloud approach as an alternative to a dedicated appliance include the following:

- Central management for all data sources deployed globally
- Simplified and faster deployment
- Automatic updates and upgrades
- Compliance with mandates such as PCI and HIPAA
- Cost reductions
- Scalability
- High availability and cross-regional backups

[12] "Cyber Threat Intelligence Uses, Successes and Failures: The SANS 2017 CTI Survey," March 2017, www.sans.org/reading-room/whitepapers/analyst/cyber-threat-intelligence-uses-successes-failures-2017-cti-survey-37677, Figure 10.

[13] "Cyber Threat Intelligence Uses, Successes and Failures: The SANS 2017 CTI Survey," March 2017, Figure 12.

Enable Human Analysts

Even though the trend is moving toward a fully automated ecosystem, human input is still needed. The existing machine learning technology still needs to be supervised and cannot work as a fully autonomous system. Security solutions powered by unsupervised machine learning are plagued by a high number of false positive alerts.

When MIT's Computer Science and Artificial Intelligence Lab (CSAIL) developed a system called AI2, it created an adaptive cyber security platform that uses machine learning and the assistance of expert security analysts to improve over time.

Figure 2 illustrates a more generic reference structure using data science and machine learning.

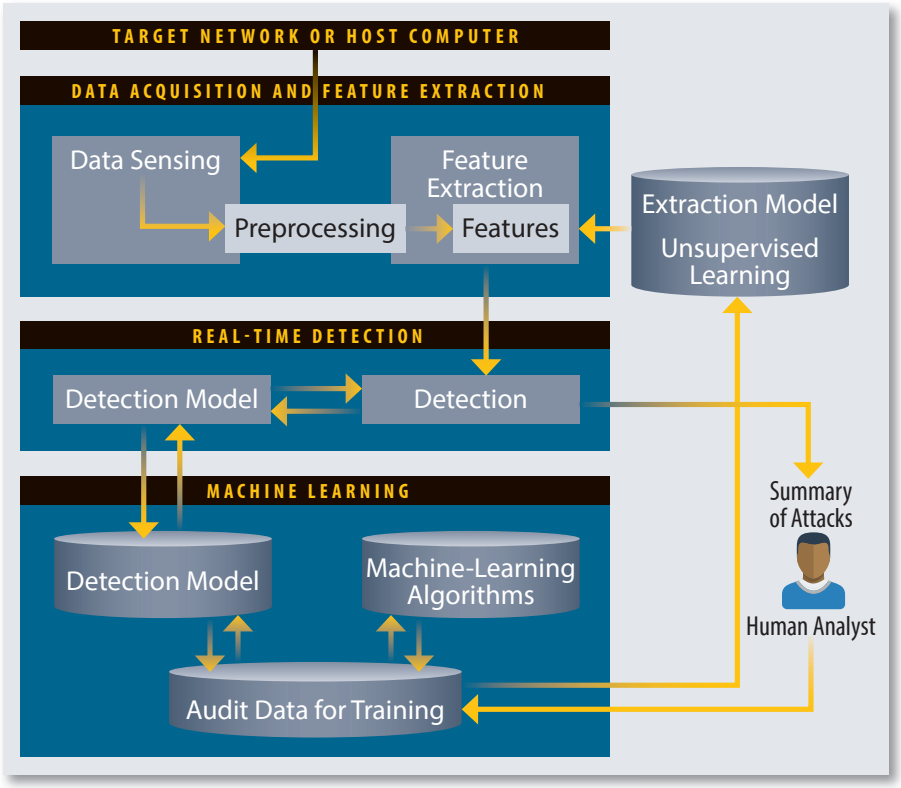

Figure 2. Machine Learning and Human Interaction[14]

With the assistance of humans, CSAIL was able to predict 85 percent of cyber attacks while reducing false positives by a factor of five.[15] It does so by first correlating all the security and intelligence data it can, which is unsupervised learning, then making recommendations to the analyst under a model called supervised learning. The program also includes a strong feedback loop for continued supervised learning.

[14] "The Expanding Role of Data Analytics in Threat Detection," October 2015, www.sans.org/reading-room/whitepapers/analyst/expanding-role-data-analytics-threat-detection-36362, Figure 4.

[15] "System predicts 85 percent of cyber-attacks using input from human experts," MIT News, April 18, 2016, http://news.mit.edu/2016/ai-system-predicts-85-percent-cyber-attacks-using-input-human-experts-0418

Conclusion

Security analytics and machine learning offer the advantages of increased visibly and the ability to react quickly and efficiently to emerging threats, while also helping close the skills gap. Current SIEM solutions are helpful, but they are not configured to process the volume of intelligence organizations are consuming. Other solutions use multiple tools that security analysts have to process and correlate manually. Under these circumstances, it's no surprise that threats are getting through and even persisting.

Automated security analytics and machine learning can reduce the stress on analysts, who are used to dealing with a flood of unconnected data from multiple sources. Rather than focusing on a log-only approach to deriving threat data, organizations should strive toward automated intelligence and machine learning to provide additional layers of behavioral analysis on the network and the endpoints. Such a shift would enable them to connect the dots between actions and provide a full-picture view of activity for analysts, instead of analysts searching through all this data themselves.

Automating the process and integrating it in the workflow will help security analysts focus on important activities, such as improving risk management, secure development, threat hunting and more. Threat intelligence from external sources is also necessary because it aggregates the latest data from threats in the real world and then applies that knowledge to the organization being protected. Organizations can also leverage the cloud for analytics: It is more cost effective, can scale quickly, and makes it much easier and faster to deploy new resources.

About the Author

Ahmed Tantawy is a member of the GIAC Advisory Board and a SANS analyst. He currently holds GIAC Penetration Testing (GPEN), GIAC Web Application Penetration Tester (GWAPT) and Offensive Security Certified Professional (OSCP) certifications, as well as the HP ArcSight Administrator and Analyst certificates. Ahmed works primarily as a security operations engineer. His experience includes working on enterprise security information and event management (SIEM) solutions and other enterprise security products, as well as leading a security operations center team in the financial sector. In addition, Ahmed has experience as a penetration tester and with ensuring PCI DSS compliance.

Sponsor

SANS would like to thank this paper's sponsor:

RSA®