ABIresearch®
REPORT
MARCH 30, 2017

THE ROLE OF AUTOMATION IN CYBERSECURITY

Dimitrios Pavlakis: Industry Analyst
Michela Menting: Research Director

TABLE OF CONTENTS

1. TO AUTOMATE OR NOT TO AUTOMATE?

1.1. INTRODUCTION AND OVERVIEW

Automating security processes is a concept that has been gaining traction recently with the enterprise market, both driving new innovative applications but also absorbing some much-needed upgrades. A first introductory example for automation that always comes to mind is automating malware detection and threat containerization—a task that has remained a mandatory component for companies worldwide over the course of the past decades. It would seem beyond comprehension for even novice technology users today to go through each and every security alert in their personal computers and manually rectify any security issues, separating legitimate from infected files. This would be not only unnecessary but also bothersome, not to mention highly prone to errors.

Fast-forward just a few years and machine learning (ML) is set to completely transform the cybersecurity landscape (among others). The raw power of both structured and unstructured data from endpoints, open-source repositories, user information, web-browsers, honeypots, networks, servers, security information and event management (SIEM) logs, and a myriad of other sources is made manifest in the form of highly sophisticated systems able to detect, investigate, analyze, and remedy a wide spectrum of security alerts. User and entity behavioral analytics (UEBA), multiple layers of neural networks, and deep learning for anomaly detection, incident response automation, and even predictive security forensics are just some of the novel applications of machine learning in cybersecurity.

Of course, it would be an understatement to mention that automation is highly tied to machine learning at this point in time. In fact, automation is just another step toward a highly functional artificial intelligence (AI) system in cybersecurity, one that will be able to match the cognitive aptitude of human security analysts and, in due time, possibly even surpass them. However, certain misconceptions still exist regarding the precise nature of automation, which harbors pitfalls and keeps new implementations from being realized on a broader scale.

These pitfalls usually concern the concept of “over-automation,” which is the danger that arises from either a) trying to automate certain security processes that should not (at least for now) be automated or b) facing difficulties in automating certain tasks which might complicate security operations and overall orchestration, especially in cases of a perceived threat. A perceived threat can be understood as a result of the false positive ratio (FPR) of a software solution which, among other results, might force an automated incident response function to:

- lock down legitimate users in an effort to block a foreign entity,
- mistakenly block IP addresses and add them to the blacklist,
- cut off necessary resources (e.g. blocking a required port in case of the perceived malware infection),
- cause authentication and encryption issues due to inappropriately handled certificates, or
- unnecessarily isolate sections of a network or lock down entire servers as a countermeasure.

The technological discourse regarding automation and machine learning resembles that of modern security measures: increasing surveillance/monitoring breadth and accepting a higher false positive rate as result of proactively teaching a system to respond faster to incoming threats. In the case of the cybersecurity objective, though, it is recognized that this tradeoff between decreasing reaction time (which is undeniably invaluable for IT) and a sharp increase in FPR as well as the aftermath (including system restoration) is not quite as balanced. To put it more simply, automation can cause more issues than it actually solves.

The talent gap is also a powerful factor. Analysts that feel comfortable or are even capable of working in cybersecurity environments are becoming quite rare, with job postings for related jobs increasing by 80% over the past few years. Governmental sources and world-leading organizations have different estimates on the exact number of security analysts required, but on average there are expected to be approximately 2.7 million unfilled cybersecurity positions by 2022. All of the above point towards one single realization: security automation might not be a choice in the years that follow, but a requirement. This report aims to present a comprehensive set of guidelines regarding automation for that not-so-distant future.

1.2. RESEARCH METHODOLOGY AND MARKET DATA

Many of the technical concepts presented in this report are, as expected, intertwined with one another. Thus, in order to prevent material repetition, some sections will include suggestions to read other parts first if needed. The methodology for authoring this report rests upon four major sources of information and includes both primary and secondary research.

First, cybersecurity and data analytics vendors were contacted and interviewed. Technical, business strategy, and market outlook insights were gathered according to their perspective of the future market, as well as their thoughts on both the automation and artificial intelligence aspects of the cybersecurity segment. Insights from this research are reflected in the form of suggestions or perspectives about current and future technologies or policies (*e.g.* judging whether an automated solution is affordable for most companies, where does interoperability matter the most*, etc.*).

Second, technology journals and articles from both cybersecurity and academic sources were analyzed in order to establish the most promising emerging trends, applications, pitfalls, and security issues. Insights gathered from this endeavor concern specific applications within the overarching automation perspective (*e.g*. how threat surface reduction is realized, why does TSL/SSL certification need to be properly managed, *etc.*).

Third, this report also endeavors to add to the automation and machine learning discourse by investigating certain hot topics that revolve around the aforementioned technologies and, in some cases, even polarize opinions in the scientific community. This is reflected, for example, in sections examining the conundrum regarding the replacement of the human element for autonomous systems, the attainability of the recommendations proposed by NIST, *etc.*

Finally, market data from other ABI Research technology verticals are utilized as a comparison point in order to establish a tangible connection to the current and future market outlook.

Touching upon the actual market sizing for security automation at this point in time is a rather elusive subject, since it involves only sections of overall technologies. For example, only a certain portion of a SIEM deals with automation, but it would impractical and unrealistic to separate the data aggregation or threat-monitoring functions of a SIEM and isolate only its automation element. Therefore, ABI Research posits that security automation can be framed between two adjacent markets: network security and orchestration, and machine learning applications.

Incident response revenues dominate by a minimum factor of three versus other technologies, projected to rise from US$14.7 billion in 2017 to US$28.3 billion in 2021, following a 17.2% CAGR increase. As will be discussed later on, incident response is a critical function for security operations and one that stands at the crossroads in the manual vs. automated dilemma. Incident response has not crossed the automated threshold yet, and it will take some time to achieve that milestone. While tasks like data gathering, data aggregation, and alert prioritization can be automated, incident response requires human security analysts more than ever (see section 2.5 for a more detailed explanation).

Additionally, Intrusion Detection/Intruder Prevention Systems (IDPS) are forecast to make an impressive jump to almost US$8.3 billion by 2021, with a CAGR of 23.2%; the highest in the studied technologies. Stable but almost static growth will be seen from the SIEM market, reaching from US$1.9 billion in 2017 to US$2.1 billion in 2021. While SIEMs generate perhaps the largest sum of data-rich resources, they are rarely considered by most mid- and lower-tier companies due to lack of IT budget. This will be further explained in conjunction with NIST's recommendations in Section 2. Table 1 and Chart 1 illustrate the data.

Table 1: Network Security and Incident Response Revenues by Segment
World Market, Forecast: 2015 to 2021

(Source: ABI Research)

Segment	Revenue	2015	2016	2017	2018	2019	2020	2021	CAGR 16-21
Incident Response	(US$ Millions)	10,800	12,660	14,790	17,320	20,320	23,886	28,233	17.5%
Firewall (NG & WAF)	(US$ Millions)	4,044	4,577	5,245	6,061	7,072	8,288	9,689	16.6%
Threat Management and Intelligence	(US$ Millions)	4,092	4,463	4,839	5,348	5,675	6,126	6,628	8.2%
Data Loss & Prevention	(US$ Millions)	2,200	2,800	3,600	4,600	5,700	6,947	8,298	23.2%
Intrusion Prevention & Detection Systems	(US$ Millions)	1,911	2,031	2,154	2,279	2,400	2,520	2,642	5.2%
Security Information & Event Management	(US$ Millions)	1,809	1,877	1,934	1,988	2,034	2,077	2,120	2.3%
Security Orchestration	(US$ Millions)	482	507	650	673	841	1,057	1,353	20.1%
Other Security (Integrated, Gateways)	(US$ Millions)	121	118	124	122	128	152	187	10.8%
Total	(US$ Millions)	25,459	29,033	33,335	38,391	44,169	51,053	59,150	15.4%

Chart 1: Network Security and Incident Response Revenue
Global Market, 2017: Segmentation by Technology

(Source: ABI Research)

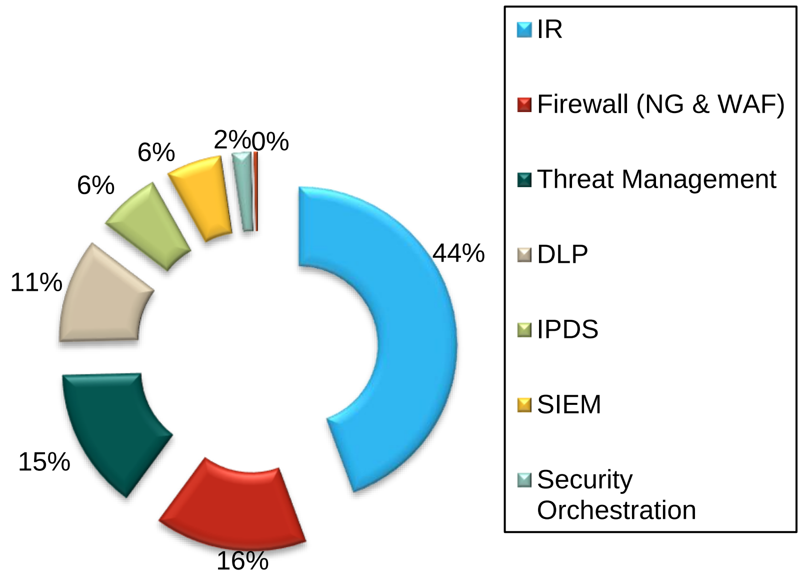


Table 2: Network Security and Incident Response Revenues by Region
World Market, Forecast: 2015 to 2021

(Source: ABI Research)

Region	Revenue	2015	2016	2017	2018	2019	2020	2021	CAGR 16-21
North America	(US$ Millions)	11,915	13,413	15,151	16,681	18,816	21,749	24,547	16.3%
Latin America	(US$ Millions)	1,833	2,119	2,450	2,860	3,357	3,880	4,555	21.1%
Europe, Middle East & Africa	(US$ Millions)	7,892	9,058	10,501	12,746	14,841	17,154	20,348	22.4%
Asia-Pacific	(US$ Millions)	3,819	4,442	5,234	6,104	7,155	8,271	9,701	21.6%
Total	(US$ Millions)	25,459	29,033	33,335	38,391	44,169	51,053	59,150	19.5%

Automation in cybersecurity is heavily driven by investment in machine learning and advanced analytics capabilities. The global market outlook for hardware, servers, infrastructure, Big Data platforms, business intelligence, and data analytics was estimated at approximately US$38 billion in 2016 and is expected to climb by a factor of three to US$96 billion by 2021, marking a 20.3% CAGR increase. As Big Data technologies (particularly Hadoop) are becoming more easily accessible, total revenues are expected to hit US$24.4 billion in 2016 and reach US$41.3 billion by 2021. Business intelligence and data analytics are also expected to make an impressive climb to US$30 billion in 2021, up from US$9.9 billion in 2016.

ABI Research posits that much of the data analytics and business intelligence offerings will also merge to be offered as additional services of existing SaaS, cloud management, or security services, thus lowering a portion of the total available market moving forward. However, the advantage of this scenario would be that the amount of companies making use of data analytics will increase considerably within a few years. Table 3 and Chart 2 below illustrate the data.

Table 3: Data Analytics, Big Data and Infrastructure
Global Market, Forecast: 2015 to 2021
Segmentation by Technology

(Source: ABI Research)

Segment	Revenue	2015	2016	2017	2018	2019	2020	2021	CAGR 16-21
Hardware, Servers & Infrastructure	(Billions)	21.5	22.7	24.4	26.8	29.5	34.1	41.3	12.7%
Big Data Platforms	(Billions)	4.6	5.5	6.9	8.5	11.5	16.4	24.7	35.0%
Business Intelligence & Data Analytics	(Billions)	7.9	9.9	12.0	14.2	17.7	22.9	30.0	24.8%
Total	(Billions)	34.1	38.1	43.3	49.5	58.7	73.3	96.1	20.3%

Chart 2: Data Analytics, Big Data, and Infrastructure
Global Market, Segmentation by Technology, 2015 to 2021

(Source: ABI Research)

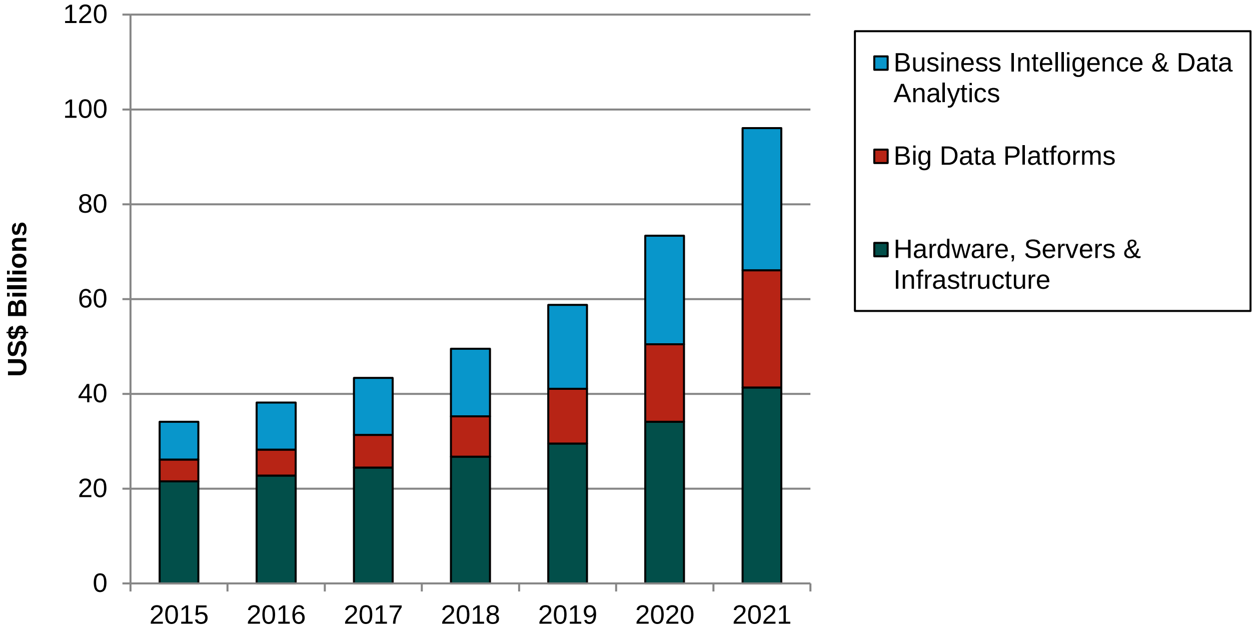


Table 4: Data Analytics, Big Data and Infrastructure
Global Market Forecast, 2015 to 2021
Segmentation by Technology

(Source: ABI Research)

Region	Revenue	2015	2016	2017	2018	2019	2020	2021	CAGR 16-21
North America	(US$ Millions)	15.9	17.6	19.7	21.5	25.0	30.4	39.0	22.0%
Latin America	(US$ Millions)	2.5	2.8	3.2	3.7	4.5	5.6	7.5	28.2%
Europe, Middle East & Africa	(US$ Millions)	10.6	11.9	13.7	16.4	19.7	25.2	33.6	29.6%
Asia-Pacific	(US$ Millions)	5.1	5.8	6.8	7.9	9.5	12.0	16.0	28.7%
Total	(US$ Millions)	34.1	38.1	43.3	49.5	58.7	73.3	96.1	26.0%

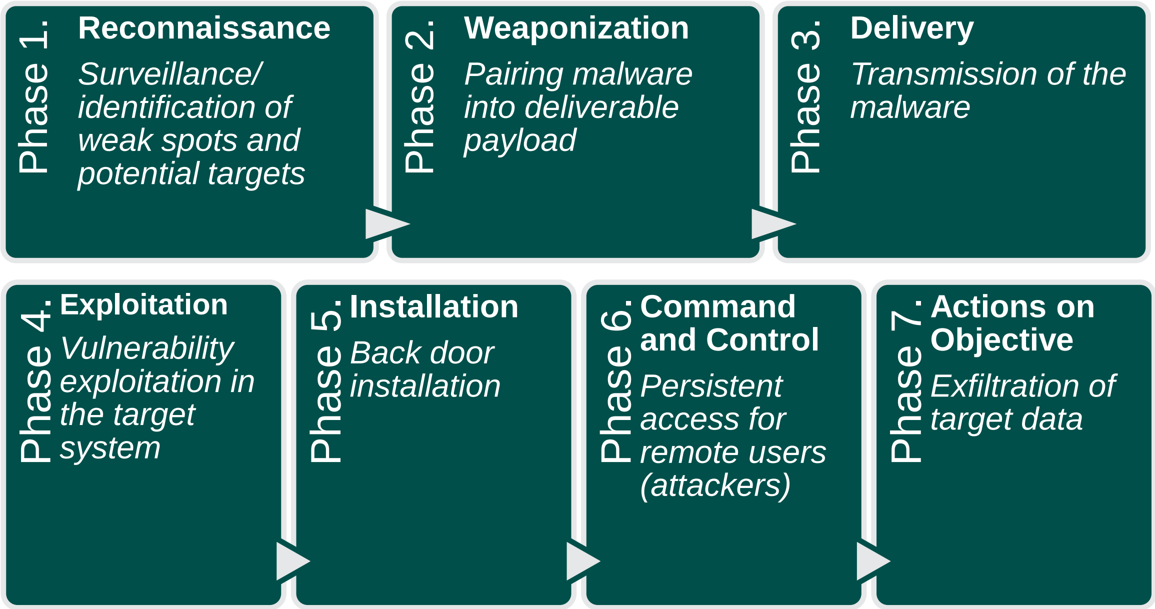
Although this report includes explanation of both network security and orchestration technologies as well as machine learning where needed, it is beyond its scope to elaborate upon these in depth. For research specifically geared towards security orchestration and machine learning in cybersecurity, please see ABI Research's reports entitled *Security Policy Orchestration and Automation*, (AN-1798) and *Machine Learning in Cybersecurity Technologies* (AN-2312).

1.3. NETWORK DISRUPTION AND ALERT FATIGUE

Evidence published in quite a few investigative reports regarding the Target data breach of 2013, including those by the U.S. Senate Committee on Commerce, Science, and Transportation, and Lockheed Martin, illustrates that there were crucial junctures where security automation took action. The phases of that particular attack are shown in the graph below.

Figure 1: Phases of the Target Attack
U.S. Senate Committee on Commerce, Science, and Transportation

(Source: U.S. Senate Committee on Commerce, Science, and Transportation)


Phase 1:
Reconnaissance
Surveillance/ identification of weak spots and potential targets
Phase 2:
Weaponization
Pairing malware into deliverable payload
Phase 3:
Delivery
Transmission of the malware
Phase 4:
Exploitation
Vulnerability exploitation in the target system
Phase 5:
Installation
Back door installation
Phase 6:
Command and Control
Persistent access for remote users (attackers)
Phase 7:
Actions on Objective
Exfiltration of target data

Cyber-forensics revealed that the first automated indications appeared, as expected, in the second and third aforementioned stages. Both FireEye and Symantec products reportedly detected something suspicious but, unfortunately, this also coincided with the failure of the security operations team to adhere to the automated response mechanisms of the FireEye Malware Intrusion Detection system and Symantec's Endpoint Protection, choosing to disregard indications pointing toward an attack. Some sources report that the security team disregarded triggered alerts believing they were just false positives. This realization forces the core question to become:

1) What would be the point of automating security if analysts treat certain automated processes like alerts from malware detection and incident response emerging from day-to-day operations as false positives;

2) while, on the other hand, allowing most automated alerts to be addressed accordingly by human analysts (including false positives) will result in security fatigue, network disruption, and overall disorder in daily operations?

Given the fact that most IT departments are understaffed, overwhelmed, and on a limited budget, the vital question now transitions to the following*: if automation is unavoidable, what should and what should not be automated*? The sections that follow will tackle this question from the point of view of the most crucial and emerging perspectives both in current and future applications in enterprise security.

It should be noted that cyber attackers are already employing automation and machine learning to empower their arsenal by purchasing and testing the most popular security hardware and software products freely available in the market, thus greatly increasing the sophistication and frequency of their attacks. They can then proceed to train their own tools to “crack open” said products, scout their target, prepare and launch their attack, and patiently await the result and the aftermath. Therefore, machine learning applications and automation in cybersecurity are seen by data engineers as a necessity, in an effort to fight fire with fire, rather than an option.

1.4. SCAP/NIST SPECIFICATIONS BURROW DEEPER INTO ENTERPRISE SECURITY

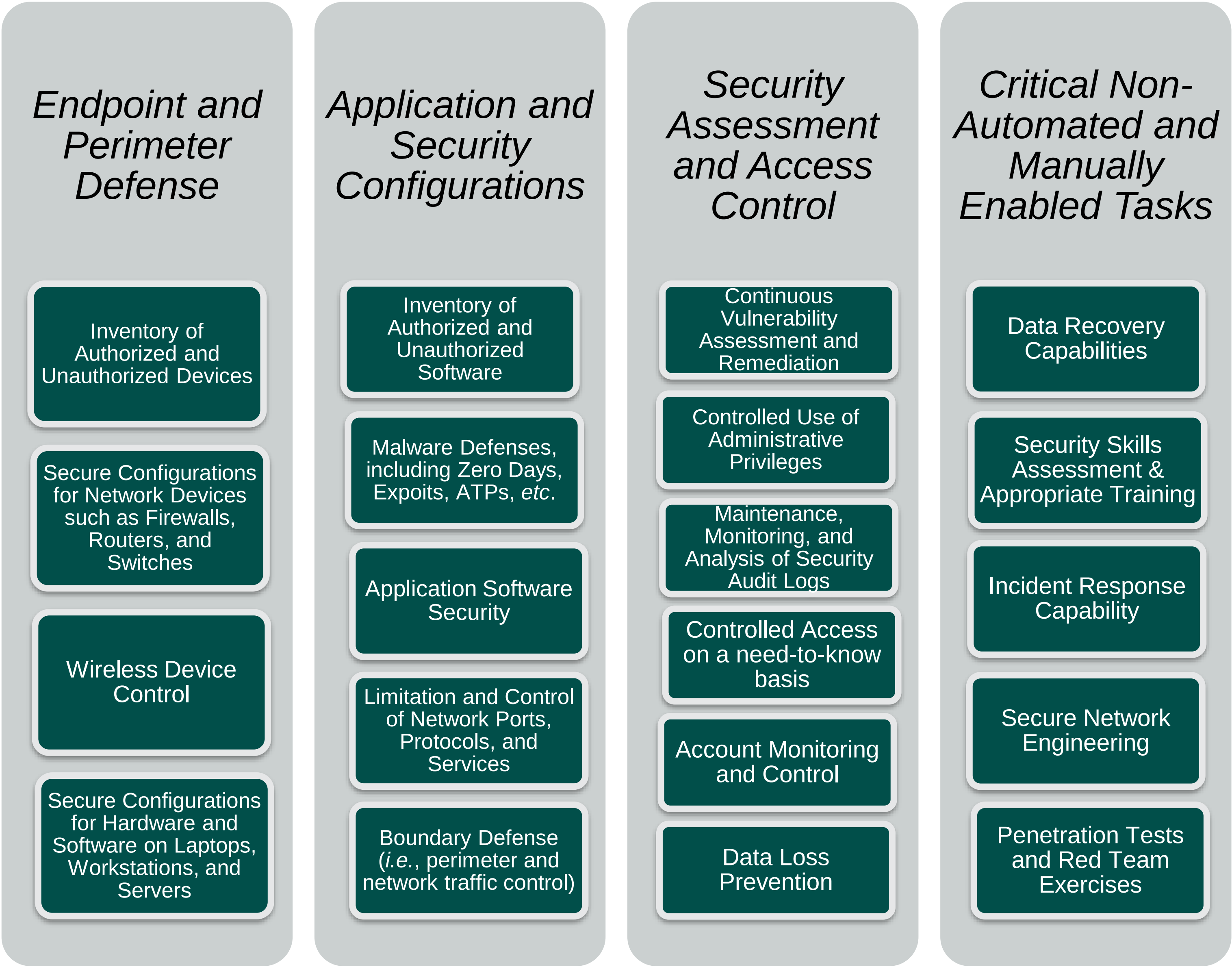
Sponsored by the National Institute of Standards and Technology (NIST), the Security Content Automation Program (SCAP) was created as a response to the massive attacks and data theft incidents of the past years. SCAP comprises different methods under one unified umbrella aimed at creating a standardized approach for enabling automated configuration and vulnerability management, security readiness measurement, vulnerability and patch checking, technical control and compliance activities, as well as policy compliance evaluation.

Through NIST and its partners, SCAP’s endeavor is to instigate a new wave of standardized security implementations in governmental and enterprise IT systems particularly targeted toward improving automated security through new industry standards and specifications. Standardization focuses primarily on achieving a higher human and machine-readable threshold with regard to format and security nomenclature originated by software flaws and vulnerabilities. An open-source tool called OpenScap has also been developed in order to assist developers in creating more advanced tools for managing system security and achieving greater levels of standards compliance.

A paper published by SCAP in 2011 entitled “Twenty Critical Security Controls for Effective Cyber Defense: Consensus Audit Guidelines (CAG)” examined 15 major security applications where automated measures could be applied and five where actual human cognitive effort is required to fine-tune the underlying applications and processes. For the reader’s convenience, and in order to better communicate the concept of automation, the critical functions have been categorized in four major categories in the graph below. The first three columns (starting from the far left) feature all functions that can be automated and the last ones are those that face difficulty crossing that threshold.

Figure 1: SCAP Critical Security Controls: Automated *versus* Non-Automated Solutions
Title 2

(Source: ABI Research)



Endpoint and Perimeter Defense

- Inventory of Authorized and Unauthorized Devices
- Secure Configurations for Network Devices such as Firewalls, Routers, and Switches
- Wireless Device Control
- Secure Configurations for Hardware and Software on Laptops, Workstations, and Servers

Application and Security Configurations

- Inventory of Authorized and Unauthorized Software
- Malware Defenses, including Zero Days, Expoits, ATPs, *etc*.
- Application Software Security
- Limitation and Control of Network Ports, Protocols, and Services
- Boundary Defense (*i.e.*, perimeter and network traffic control)

Security Assessment and Access Control

- Continuous Vulnerability Assessment and Remediation
- Controlled Use of Administrative Privileges
- Maintenance, Monitoring, and Analysis of Security Audit Logs
- Controlled Access on a need-to-know basis
- Account Monitoring and Control
- Data Loss Prevention

Critical Non-Automated and Manually Enabled Tasks

- Data Recovery Capabilities
- Security Skills Assessment & Appropriate Training
- Incident Response Capability
- Secure Network Engineering
- Penetration Tests and Red Team Exercises

First are the Endpoint and Security Configurations. These functions predominantly concern endpoint security, inventory of authorized devices and improved visibility over unauthorized ones, and software and hardware configurations for workstations as well as routers, switches, and other network devices.

Second in line are the Application Security and Perimeter Defenses. These are a) malware defenses, from Anti-Virus (AV) systems all the way to Advanced Persistent Threats (APT), zero days, including exploits from expired encryption certifications, *etc*.; b) constant monitoring for unauthorized software, persistent application security, as well as perimeter protection; and c) setting the stage for account access control by managing protocols, services, and network ports.

Third, and aided by the previous two function clusters, is system-wide Security Assessment and Access Control. This includes management of user access and administrative privileges, account monitoring, analysis and audit of security logs (which is intertwined with all other function clusters), Data Loss Prevention (DLP), and continuous vulnerability assessment and remediation.

Finally, but perhaps most importantly, are the critical functions that are more difficult to automate and for which organizations have to rely on actual, manual actions to complete objectives. These primarily include incident response capability (which is undoubtedly one of the key aspects fueling the flames in the manual versus automation dichotomy in cybersecurity), secure network engineering and penetration testing, data recovery capabilities, and skill acquisition and security training initiatives. Please note that Section 2 offers further arguments regarding the steps currently underway to add automation functions in the “manual” cluster.

One vital point that needs to be added is that SCAP, NIST, SANS, international cybersecurity entities, and IEEE strongly suggest that modification of defenses should be focused not only on combating existing threats but also preparing, investing, and anticipating future ones. This is certainly a worthwhile goal that does carry a fair amount of validity. It is also, however, a somewhat unattainable one. The majority of companies do not even have the budget to tend to most security alerts that are generated from a SIEM on a daily basis, let alone to begin preparing for any future attacks without knowing the precise nature of those attacks, or to justify a higher hypothetical ROI to management. At the very least, an achievable goal that ABI Research can recommend at this point is that organizations should put more effort in investing in flexible and customizable security orchestration solutions—particularly ones that will not be dragged down by secondary services (*e.g.* from third-party security vendors), lack of interoperability between software products, lack of transparency and network visibility, or burdensome customization.

2. MACHINE LEARNING APPLICATIONS DRIVING AUTOMATION ADOPTION IN CRITICAL FUNCTIONS

Possibly one of the most important insights to retain from the previous section is the actual need to perform a clear distinction between the functions that can and those that can’t be automated, and understanding the reasoning behind that decision. Functions like incident response, network engineering, skill assessment, and the full-on physical, digital, and social attacks associated with red team exercises have been excluded from the automation list for quite some time.

The advent of machine learning in cybersecurity, however, has breathed new life into IT processes, rendering some of the aforementioned functions a few steps closer to the automation objective. While SCAP’s report presents one of the most comprehensive examinations of automated processes, it is somewhat dated. After discussions with industry players and analyses of related technologies, ABI Research posits that the industry is actually a lot closer to the 85% cutoff point of autonomous processes than most organizations believe (*i.e.*, the main objective is to automate at least 70 to 85% of all processes). Below are some of the critical operations that actually made significant steps in that respect and are very likely to be excluded from the 4th cluster in Graph 1 (Section 1). Realistically, at least for the foreseeable future, 100% full-system automation cannot be achieved or expected. However, it may be practical, secure, and even attainable in some cases (*e.g.*, penetration testing).

2.1. DATA RECOVERY

Otherwise known as the dreaded path organizations are forced to tread as part of their disaster recovery plans, data recovery technology has hit a few bumps in the road in the past, but recent developments, along with more comprehensive policy-based storage management systems, are allowing it to join the ranks of automated functions. Modern rule- and policy-based management systems are now offered in most new software services and products and allow for very specific and detailed actions to be taken if certain technical conditions are triggered.

Automated data recovery systems could include, for example, instances of data replacement, server migration, more readily available restore points, continuous backups, more convenient database processing capabilities, and archiving. This is not to say that a complete autonomous data recovery process can take effect in every corruption or migration occurrence. It will still be dependent, of course, on the actual infrastructure, as well as the data engineers to create the rules and monitor the policies for the actual recovery process. However, a partially automated system can be achieved.

2.2. SKILLS ASSESSMENT AND TRAINING

In principle, training human personnel and creating and training an algorithm both have the same objectives, but must follow completely different paths. In the challenge where the task is to decipher whether a file is of legitimate or malicious origin, an analyst's cognitive processes would almost immediately begin by recalling previous instances of similar encounters. A machine learning algorithm would attempt to classify that file according to similarities it can detect in millions of other files it has encountered in the past. Both a human and the ML model require time and effort to produce significant results, but the skill assessment would still remain the same. Both need to learn how to identify a threat, follow historical data patterns, understand which areas of the system are affected, and make an intelligent decision as fast as possible.

Two key distinctions appear at this point. First, knowledge cannot be forced into an analyst's mind to learn the exact patterns needed to recognize future events. On the contrary, an ML model can achieve exactly that: learn a specific task (with varying success), with specific data, in a very specific pattern. Second, algorithms would produce erroneous results if the monitored entity (*e.g.* a potential incoming threat) does not fit any known patterns, or if the model has not been properly trained or has been trained with insufficient data. Analysts, however, can still go a few more steps ahead when the model does not know how to proceed and come up with rather ingenious, simple, and innovative solutions. For an algorithm, complexity falls to a secondary level. Even if some challenges are indeed quite simple for the human mind, an algorithm that has not been properly trained will not be able to perform at the same level.

In a sense, training an analyst or an algorithmic model from the bottom up requires "teaching" them the very same principles. What differs is the nature of the teaching method: the first requires practice and resources taught by analysts and engineers, while the latter requires machine-readable data and coding by the very same professions. Therefore, automating training of the human element will never be possible (at least not without neuroscientific intervention and electrical stimulation on the parts of the brain that hold the respective faculties!). However, automated skill assessments would be much more easily attainable and could involve similar work being performed by international standardization bodies or benchmarking entities.

2.3. PENETRATION TESTS AND RED TEAM EXERCISES

As mentioned previously, cyber-attackers are already employing machine learning to empower their arsenal by purchasing and testing popular security products, making penetration testing as relevant as ever. However, both blue team and red team pen testers can rest assured that their jobs will not be replaced by a machine anytime soon. Automation is but one aspect of this highly complicated and multidisciplinary job (especially for red teams) and one that cannot accurately be replaced.

2.4. NETWORK ENGINEERING

Network engineering, like penetration testing, can only be enhanced—not replaced—by automation. Machine learning usage can be empowered with analyses from AV systems, IP/ID systems, SIEMs, traffic flow analyses from switches, server capacity, and network systems. These appliances can provide additional insight and assist IT personnel in engineering said network. From a practical standpoint though, this is a task that cannot be automated. Modern automated systems, however, can certainly harvest machine-readable data from all different system technologies, including SIEMs (*e.g.* HP's ArchSight), AV systems (*e.g.* Symantec), Analytics and SIEM-like services (*e.g.* Splunk), ID/IP systems (*e.g.* PaloAlto) and convert it into applicable threat intelligence, guiding the process of future network engineering endeavors. Parts of this process can also apply to network privilege management and the influx of new IoT devices and the "Network of Things" (as NIST postulates), which includes the creation of technical, administrative, and physical standards.

2.5. INCIDENT RESPONSE

Perhaps most important of all manual tasks which has repeatedly hit barrier after barrier is incident response. More specifically: the development of an *intelligent* incident response (IR). This primarily includes the ability to recognize and estimate the ramifications of its actions to day-to-day operations and adjust accordingly without shutting down critical operations or constantly (and erroneously) isolating endpoints, user accounts, or even servers, while at the same time containing or partially dealing with the threat at hand.

Incident response has been a purely human-focused task and thus one that carries a great deal of human error—not from lack of intelligence or domain knowledge, but from lack of time to review, analyze, and remediate incoming threats. A survey of 259 employees conducted by SANS in 2014 showed, among other findings, that 62% of respondents reported insufficient time to review and practice procedures; 60% reported a lack of resources, tools, and technologies; and 61% admitted to receiving additional assistance from internal, non-IR staff in order to deal with surge needs. The need for autonomous IR becomes, therefore, quite apparent. While this subject merits an investigation on its own, this report will mention some of the most important challenges below and in the section that follows (Section 3).

2.5.1. DATA INPUT

IR automation is highly dependent upon automated detection and threat prioritization capabilities. This includes any data that can be generated from the system but primarily is used from SIEM logs, AV software, and host/network ID/IP systems.

2.5.2. INTEROPERABILITY AND SCALABILITY

The aforementioned data diversity is, in turn, dependent upon the interoperability and scalability capabilities of a chosen automated solution, as it would need to be able to gather logs from literally every corner of the system. A security product that is unable to leverage both AV and SIEM data at the same time, or that does not support a specific IDPS vendor product, should be faced with skepticism.

2.5.3. NETWORK VISIBILITY

Network visibility needs to be one of the higher priorities for most critical functions, not just for IR. It is not uncommon for companies to stay "in the dark" when it comes to gathering intelligence from across their systems. Even managing the number of devices in their network or containing unauthorized devices becomes a challenge. That being said, it will be very difficult to achieve any data granularity, not to mention reduction of false positive rates, without granting automated systems increased network visibility.

2.5.4. CLEARLY DEFINE AUTOMATION NEEDS

According to the second revision of NIST's "Computer Security Incident Handling Guide," automation is required in order to "*perform an initial analysis of the data* [recorded by security software] *and select events of interest for human review.*" This includes correlational analysis and relates more closely to SIEM operations. In fact, the process mentioned above describes a semi-automated solution in which the product *produces* an output for a human to review and act upon. It does not *remediate* or tackle a situation without, or with limited, human supervision.

For some companies, the desire for automated processes can easily be achieved by a SIEM. In fact, modern SIEMs from tech giants like Intel, RSA, IBM, and HP do quite well in that respect, and are continuously improving their offerings. Other companies, however, are aiming toward fully fledged automated solutions that touch upon the boundaries of what AI can offer at this point in time, including real-time analytics, predictive prevention system, automated IR, behavioral pattern analytics, and highly sophisticated self-learning ML algorithms.

2.5.5. OVERCOME INFORMATION SHARING CHALLENGES

Part of the task for IR teams is to effectively communicate and exchange threat intelligence information with third parties. In some instances, this may also include governmental and law enforcement agencies, or even other companies and organizations. If an automated system unwittingly made data freely available or communicated sensitive information to unauthorized parties, this would place an organization in a nightmarish position.

2.5.6. DID WE SKIP A STEP? THREAT DETECTION AND AUTONOMOUS IR

It should also be mentioned that malicious and suspicious behavior is still quite difficult to detect, and a task that continues to surpass even modern security products. It would be easier to automate an internal task. In fact, most internal tasks can be automated since the organization has complete control over the processes, data format, and involved players. Attempting to automate something like threat detection and remediation, which does not follow the same rules but rather ever-changing variables, will require a completely different process.

On the other hand, IR specialists are under severe time constraints to manually address all security alerts. Out of thousands or millions of alerts per day (depending on company size), only a small fraction of about 1.5% to 2% will constitute a threat, and out of that number only less than half will actually be adequately monitored by security personnel. At its best, autonomous IR would essentially be an exercise in compromise, which might require an additional validation phase through a pair of human eyes. The objective that organizations wish to see achieved in the following years is to be able to transfer at least 75% of the human IR tasks to a combination of ML/automation.

2.6. LEVERAGING THE EXTENSIVE DATA-GATHERING CAPABILITIES OF SIEMS

There really wouldn’t be a discussion regarding automation without investigating the role of SIEMs and the chief function they offer: a constant influx of data logs. While SIEMs are indeed a nice boost to enterprise IT, they have not changed significantly over the past decade. Most changes concern enhancement of their analysis abilities, further modification of product interfaces, a refinement of the data logs collection process, and adjustment of interoperability capabilities. Nonetheless, SIEMs are still as relevant as ever given their unique capacity for monitoring, detection, and incident response capabilities across a multitude of hosts, which are without a doubt critical functions for most mid- to top-tier organizations.

Automation can also have an application in the interoperability aspect of SIEMs. This is due to the two-way directionality feature which must characterize interoperable solutions; *i.e.*, other applications are able to successfully send data to SIEMs and they, in turn, are able to digest that data and send back further instructions. This feature can easily be a part of a larger package solution which can automate communication between different software applications, devices, servers, and hosts. The latter may also hold a higher applicability factor in cloud-based SIEMs, in which certain processes can be automated to gather insights (anomaly detection, attack pattern, *etc.*) from other hosts and organizations within the same cloud cluster and automatically incorporate them as part of their own incident response in preparation of a similar occurrence.

This will not only provide fresh data and insights in anticipation of similar threats, but it will also empower IT professionals when attempting to correlate different events or even justify their correlation analysis to management and operations. As with most other aspects in enterprise cybersecurity, speed is of the essence: an automated response using the aforementioned technique would be a lot faster than tasking an analyst to perform this analysis, justify it, and then proceed to adapt to an attack inflicted on another host. This also proves to be a thorn in the attackers' side, since their attack will not only fail on other hosts, but adequate preparation can yield important findings for a forensics investigation later on, or even from an anticipatory honeypot and tracing for the most adventurous organizations.

Cloud-based SIEMs are another important market that needs to be considered. Running off-premises, cloud-based SIEMs may seem like a noteworthy automation approach, and it can be—at least in theory. However, implementers need to consider that automating data collection may also have an adverse effect on bandwidth capacity, cloud storage, and other related costs, which may actually produce the opposite effect than the streamlined approach desired by IT professionals.

One can imagine how impractical it would be to implement an inappropriately tuned automated SIEM function which collects huge amounts of unneeded (or at the very least "somewhat needed") data, which will:

- increase workload for security analysts,
- slow down performance by performing data cleaning and wrangling for all that unnecessary information,
- increase all costs related to cloud-based services attached to this function, and
- fail its cost/benefit result by draining much needed resources and human attention.

This can (and should) be addressed by:

- employing further ML modifications to the actual automation process,
- configuring the precise nature, objective, and lifespan of the data logs that are required, and
- fine-tuning the interaction between existing SIEM operations (at least those that are considered of the highest operational value and critical to the enterprise) and the proposed automation process.

2.7. THE DOUBLE-EDGED SWORD OF ENCRYPTION

Encryption can be a double-edged sword. Machine learning tools and network monitoring solutions claim to aggregate all incoming and outgoing traffic. However, even network traffic from within the company (*e.g.*, VPN, *etc.*) is encrypted by default, which might leave some solutions missing a vital portion of their own data-gathering capacities. An automated process within a SIEM solution that has not addressed this issue will have a continuous blind spot over certain (but still critical) portions of the network.

This can be addressed by: a) choosing a third-party provider that guarantees constant evaluation of their clients' systems, allowing for full visibility across the organization (although note that definitions vary greatly on what "guarantee" actually entails); or b) having the SOC personnel create, maintain, and update custom-fitted rules for this exact purpose. While the first option will undoubtedly save time but add to the overall cost, the latter will have the exact opposite effects, with the added benefit of giving the security analysts a chance to investigate for themselves any new issues that will undoubtedly arise. The latter approach is a semi-automated solution—a compromise—between overall cost, security, time, and resources.

This problem becomes apparent further down on SIEM operations. Another humble (at least from a statistical point of view) but vital function of a SIEM is to perform a cross-correlational analysis between massive sets of data pouring in from every edge of the organization (and even outside the organization if external sources are needed). This analysis will perform the traditional but potent relationship evaluation between multiple variables. However, missing certain key variables and performing even a simple correlational matrix analysis will not only prove less than fruitful, but also attract the analysts' attention toward the wrong conclusion. Translated into SOC terminology, this means "loss of time and resources hunting down false positives." Blind spots across the system will also produce blind spots in the actual analysis which, in turn, can render even the application of a supposedly "modern" SIEM rather useless.

3. AUTOMATION OBJECTIVES IN CYBER SECURITY

This section will examine more specific uses of automation in cybersecurity, as well as other processes that contribute to security indirectly (*e.g.* formatting and machine-readable data for machine learning, SIEM integration, encryption optimization, *etc*.). Greater focus will be given toward enterprise settings, mostly due to their apparent lack in budget and resources, but the following applications apply to governmental systems as well.

3.1. AUTOMATED PATCHING ACROSS MULTIPLE SERVER INSTANCES

Patching is possibly one of the paramount applications of automation and one that is currently considered a primary candidate for every organic implementation for new clients entering the machine learning and automated solutions space. As expected, the three key motivations behind this approach are: a) tackle potential malware infections across the system which could escalate due outdated firmware, b) decrease manual patching time and free up engineers' time, which in turn contributes to c) conform with regulatory compliances and state-of-the-art requirements. This extends both to day-to-day operations when understaffed IT departments must sustain patch management across hundreds or thousands of servers, as well as crisis instances when an attacker malware payload is infecting system nodes.

During crisis mode, when patching must be done across thousands of servers to fix an issue that should not have existed in the first place, security analysts must work with software engineers to patch vulnerabilities and keep servers running up-to-date firmware to contain the damage. At this point, it is truly a numbers game. The victim organization will try to protect as much ground as possible before attackers make it deeper into the system, something that can arguably be performed faster in a well-structured, automated fashion.

One cannot expect an automated solution to simply appear when the company is in crisis mode and just "fix" everything. It is not a plug-and-play panacea solution—far from it. Like machine learning, automation requires elaborate design, testing, and evaluation prior to such an event so that they could handle a crisis when it appears.

3.2. SYSTEMS ENGINEERING AND OPTIMIZATION

If implemented correctly, automation can truly shine in systems engineering by streamlining performance, testing, and optimization. Systems engineering is a multifaceted approach that involves considering a broader perspective of the digital roadmap of an organization and continuous testing and evaluation of both hardware and software components. Due to the interdisciplinary nature of the work involved, systems engineers are required to perform multiple tasks including network security, system architecture, digital management, software integration, and overall system testing and optimization. This can be achieved by:

- automating certain daily tasks from the systems engineers' workload, which will reduce overall mistakes;
- decreasing the bulk of the mundane work, allowing the rest of the key tasks (which require manual attention) to be easier to tackle;
- greatly assisting in the mandatory testing, analysis, and evaluation of newly configured software;
- saving time in network infrastructure optimization (*e.g.* configuring user access privileges, firewalls and security benchmarking, resource allocation and detection threshold, data encryption); and
- reducing resources spent on standardizing data formatting between multiple operating systems, SIEM logs, devices, *etc.*

3.3. EXPEDITE FORMATTING AND MACHINE-READABLE DATA

Standardizing data formatting in systems engineering is a fundamental aspect of automation that deserves its own analysis. This addresses the frustration of trying to tackle data and system logs from multiple sources, each with a completely different data format. User personal devices, operating systems, third-party services, applications, SIEMs and network data, web browsers, databases, and Big Data programming languages and repositories all use different data formats: from the JSON Javascript format, to .tar for Unix, to the .xls format necessary for the Excel sheets required for management personnel, and many others.

Any programmer that has dabbled in more than one programming languages knows first-hand the difficulty of constantly trying to convert and homogenize data from completely different sources and make them work as part of a greater cross-platform analysis. Standardization in any form is a well-known issue among many technology segments and one that is expected to become even larger in the fast-approaching interconnected IoT environment. Data formats available from one single device or program are not meant to be readable to other software products. Most of them are actually meant to be readable by humans rather than machines, and it is humans that have to manually perform all mandatory transformations in order to involve more people or programs in a greater data consumption process.

Automated data formatting is another crucial area of application and is tied to the specific needs of every organization, and must be tailored according to the architecture and software dependencies. As will be discussed later in automated data collection, automated data formatting should not encompass each and every application, nor should it be applied round-the-clock. This will not only be a significant drain on resources but will also result in unnecessary effort if the analysts and engineers do not actually require the processed data. Rather, it should be chosen as a part of a greater response chain anticipating analysts' needs instead of blindly throwing data (no matter how uniform they are) toward them. This automated formatting application is not just limited to analytics. Performing analysis and having access to the greater picture is crucial to an SOC team aiming to correlate different events together. However, it also applies to conforming to machine-readable compliances which can expedite the communication between SIEMs, SOC analysts, encryption, and certification management (this will be addressed in a later section of this report).

3.4. SECURITY ALERT STREAMLINING

One of the most palpable applications of security automation rests within the hundreds of security alerts that each security analyst might be forced to process every year. Security vendor Hexadite has estimated that approximately 92% of companies are forced to deal with an estimated 15,000 alerts per month. A single security analyst has the capacity to deal with an estimated 10 alerts per day.

Once again, it is a numbers game, and one that human analysts cannot hope to tackle alone at this moment in time (let alone the myriad of new IoT threats approaching on the horizon). Automating a great deal of security alerts will:

- reduce the overall attack surface by allowing IT to "cover more ground" much faster,
- allow analysts to be able to invest more time dealing with more threatening issues,
- reduce the amount of mistakes performed by overworked employees,
- provide a more streamlined and tailored approach to IT security instead of blindly allocating human talent to deal with issues when they can be avoided in the first place,
- assist SOC teams to oversee, correct, and adapt the automation process depending on their organization's needs, and
- empower analysts and engineers to perform much better on more important tasks.

Instead of forcing security analysts to spend more time constantly monitoring security alerts, matching different patterns together, and deciding which ones require their immediate attention, an automated alert prioritization process can also be implemented. In this application, the need for machine learning becomes apparent as past SIEM logs, external threat intelligence, and other enterprise-generated machine-readable data (also addressed in other sections in this report) can be easily fed into sophisticated machine learning models in order to classify, predict, and categorize any potential incoming threats hiding behind security alerts.

This can also allow security teams to create organization-tailored alert classification profiles. With the use of the supervised learning algorithms, even certain SIEM alerts can be retrofitted to appear categorized according to pre-customizable variables depending on the organization's needs. Similarly, unsupervised algorithms and deep learning can be used to identify "category-agnostic" or, in simple terms, unknown threats that cannot fit any supervised classification objectives or are highly dependent on large amounts of unstructured data. So what would this configuration mean for companies in practical terms?

For one, it will depend on the security infrastructure and the threats that each company is most concerned about. For example, companies that have suffered from APTs in the past will have their SOC teams searching for completely different patterns than those that have dealt with ransomware or Remote Access Trojans (RAT). Second, similar monitoring patterns and restrictions can also be observed depending on the more operational factors such as encryption compliance and certification, constant patching requirements, API, and third party client dependencies.

Third, IT needs (and budgets!) fluctuate constantly and are tied to overarching organizational success. Companies need to take into account that in the event of a limited budget, they should be able to customize automated processes depending on their current needs. It should be noted that while some security vendors, including SIEM and UEBA vendors, offer only some client-side customization on their respective products, others can only offer provider-side services (*i.e.* as part of an additional paid service).

3.5. TSL/SSL CERTIFICATION MANAGEMENT

Effective management of cryptographic keys is one of the highest priorities for security operations and one where automation is gaining traction. Although automated Transport Layer Security and Secure Sockets Layer (TSL/SSL) certification management is not a new idea, many organizations have suffered in the past from miscalculations in related automating processes. While having that extra "S" at the end of HTTP is a crucial addition to the IT arsenal, providing some extra protection layers, it does come with its own set of responsibilities.

Internet-based organizations in particular have had a hard time dealing with issues caused by expired certificates or certificates that might require some unconventional configurations to reinstate them to an operation state, leaving companies with easy-to-target faults and vulnerabilities (not to mention forcing analysts to fix those issues manually).

Some organizations (*e.g.* Amazon Web Services) can tackle this problem through APIs and with the use of Elastic Load Balancing (ELB), which can increase the vulnerability tolerance by allowing traffic to be routed across multiple instances. ELBs come in two different forms, the Classic and the Application Load Balancers, tackling application and network, or advanced application and content traffic information.

However, if the security industry has learned a valuable lesson, it is that automated processes handling configuration and renewal for TSL/SSL certificates must be achieved in tandem with the points regarding data formatting discussed in a previous section of this report. In order to achieve a truly centralized management threshold (including certificate management), further compliance with machine-readable policies has to be enforced, allowing easy integration and digestion of data not only in the examined instance (*i.e.*, certificate management) but also as part of the SOC team. Additionally, this will not only save precious time for security analysts (and web developers alike), but also allow a smooth transition for future implementations and further TSL/SSL configurations when a new expiration, restriction, dependency, or general certification challenge inevitably emerges during the next renewal cycle.

It should be mentioned that depending on the nature of the issue, this could occur as late as one year or as early as one month. Companies with fewer dependencies or general needs might argue that they cannot justify for automated process to take effect once every year. Other, more HTTPS-focused organizations with more frequent TSL/SSL certification and compliance needs can certainly gain much more from such an automated process.

3.6. PERMISSION CONTROL AND PRIVILEGE MANAGEMENT

A general rule in IT security is that the amount of access that a specific user or a group of users (*i.e.*, group cluster with similar needs) can achieve in a company network is highly correlated to a sizable increase in overall threat surface. This, of course, coincides with a substantial increase in the likelihood of falling victim to both external and internal attacks (including a higher probability for more insider threat instances). Research interviews with security vendors have revealed that higher- ranking employees present most of the barriers when it comes to restricting access.

In fact, employees that usually stand higher in the corporate ladder are more resistant to change, especially when the desired result is a sizable decrease in user permission. Evidence of this was also found in another security research regarding the usage of Multifactor Authentication (MFA) in the workplace. Interestingly, these higher-ranking employees do appear to be in favor of reducing threat surface when similar measures are enforced on the majority of other employees. The irony, of course, is that managers and directors are the second-most desired target group for external attacks (with the first being IT managers and personnel and most C-level employees).

Thus, permission management is another application that automation can improve. The main benefit of such a solution would be not assigning or increasing access privileges for users that make such a request. Rather, its vital function has the exact opposite purpose: to trim down unnecessary privileges or relatively unused ones and either replace them or remove them altogether. Purposefully shutting down unused access privileges has two major advantages.

First, it can reduce overall potential threat surface. Second, it provides a major boost when it comes to cyber-forensics, SIEM correlational analyses, or any historical-based investigation. In addition, it can streamline any existing behavioral-based analytics approaches like UEBA by limiting the number of factors that might cause the algorithms to derail an employee's normality curve and flag them as an insider threat. This could include, for example, accessing certain types of resources that might not be considered normal behavior for an individual, but due to extended privilege access, such an event would still be plausible. From an analyst's point of view, this may be cause for an alert dismissal due to its possible classification as a false positive—which loops the discussion back at the first section of this report and touches upon the alert fatigue versus security disruption conundrum.

3.7. EMPOWERMENT, NOT REPLACEMENT, OF THE HUMAN ELEMENT

The replacing versus empowering of the human element conundrum is perhaps the primary source of confusion and controversy in the entire machine learning, automation, and robotics discourse. The question in the minds of many professionals revolves around "intelligent" automation impacting job availability. While this answer undoubtedly varies according to the end-market in question, in the case of cybersecurity, all findings point toward the answer being an emphatic "no" or at the very least "not for now."

First, ML and automation hope to bridge the ever-increasing cybersecurity gap. The lack of security analysts, along with their somewhat lower retention rate and combined with the overall lower reported job satisfaction among security professionals (not to mention the multidisciplinary training required) are some of the primary factors contributing to this phenomenon. In addition, this is the exact opposite trend of the sharp demand curve that can be found in enterprises regarding the frequency and intensity of cyberattacks on literally every market sector. This makes the talent gap issue even more pressing in the foreseeable future. As far as cybersecurity is concerned, automation and machine learning are not even close to replacing human talent, but rather are filling the gaps.

Second, human analysts do possess an incredible capacity for adapting to situations, making use of advanced learning mechanisms (something cognitive computing attempts to copy), and employing extraordinary out-of-the-box thinking. However, as mentioned in earlier sections of this report, humans are prone to errors, especially when confronted with cognitive overload (a phenomenon that can occur quite often when being part of a SOC team). An automated program would be much more efficient at going through millions of data logs or performing such tasks (even on a massive scale), but would be at a prodigious disadvantage in the following examples:

- if not properly trained (using *e.g.* deep neural networks), it would not be able to decide whether a new file is legitimate or infected or identify unknown threats;

- it will not be able to proceed to appropriate incident response (without disrupting operations by unnecessarily raising quarantine or shutting off ports) if not specifically programed to do so; or
- it will not be able to thoroughly investigate an alert pattern unless it actually has the capacity to do so and has direct access to a wide spectrum of machine-readable data.

Third, the exponential increase in cyberattacks over the last few years has not only left IT severely outnumbered, but also raised the processing, awareness, and capacity threshold required to keep enterprises running smoothly. Thus, when it comes to cybersecurity automation, ML is a natural evolution of technology and primarily aims at alleviating some of those issues. ML is meant to empower the human element, not replace it. A glimpse into the future of cybersecurity will see enterprises being completely transformed by AI and automation including:

- the ability to automatically harvest, filter, and digest data from a wider market, including the consumer and even academic spectrum, depending on current and potentially future organizational needs;
- the incorporation of Natural Language Processing (NLP) and chatbots as part of everyday, round-the-clock issue reporting, with the ability to automatically investigate said issues and adapt perimeter and network defenses accordingly without the need for authorization by administrators;
- AI capable of predicting insider threats and proactively begin monitoring specific endpoints;
- the ability for security systems to learn what appropriate incident response actually means to the unique needs of a company.

Most importantly, enterprise security is highly dependent upon the actual presence and operational status of administrators and security personnel. Currently, the industry has achieved some major strides in allowing systems to learn from user input over time regarding file legitimacy and the normalcy adjustment of the UEBA curve. However, in the future, automated security systems will emerge that can detect a multitude of patterns and learn what a holistic and appropriate incident response actually means on a 24/7 scale, not 8 to 9 hours per day for five days a week.

4. VENDOR SECTION

This section of the report will present a representative sample of vendors from each of the technologies discussed in the report: SIEM, SOC, AV, Machine Learning, UEBA, Automation, and Orchestration. Companies featured include SIEM vendors McAfee (Intel) and Splunk; ML and AV vendor Symantec; Automation and Orchestration vendors Ayehu, Hexadite, and Hexis Cyber Solutions; as well as ML and UEBA security vendors Gurucul, Trudera, and Vectra Networks.

4.1. AYEHU

Ayehu operates with the Automation-as-a-Service business strategy and particularly focuses on IT process automation. Although the company is not active in other related automated processes like threat detection (which is more a concern of a SIEM vendor), they offer a great deal of flexibility and granular analysis thanks to their fully customizable rules-based engine which is part of their eyeShare flagship product. Ayehu focuses on the main aspect where companies require some additional back-up: intelligent assistance to their security analysts. Ayehu posits that security analysts need to maintain full control on detection and incident response, but also allow them to scale both their individual talents and IT security readiness. As much of the competition, they have developed their tools so that even a junior admin without any coding experience or programming background can make use of it.

Last year (February, 2016), the company also introduced its next generation IT automation and orchestration solution as a Software-as-a-Service platform for security operations. It features an intelligent machine learning solution driving decision support, dynamically created rule-based recommendations, and more sophisticated correlation analyses for both fully or semi-automated workflows. Most importantly, the SaaS platform includes support for hybrid deployments across on-premise, private, or public cloud environments.

4.2. GURUCUL

Gurucul operates using advanced UEBA and identity analytics and is one of the leading players that place the “unknown threat” at the cornerstone of modern cybersecurity research endeavors, as well as peer group monitoring. The company has undergone a significant transformation since 2009, developing their own risk analytics models, Big Data and cloud analytics platform, auditing software, and intelligent role models.

One of the core elements of the solution is its ability to compare individual behavioral patterns against other organizational groups for signs of deviation and act upon these insights if there is a high enough probability of fraudulent behavior or insider threat. In its effort to minimize false positive rates, Gurucul strives to collect an abundance of data from multiple data sources, including identity management platforms (*e.g.*, Oracle, RSA, CA, Sailpoint), privilege access management (DELL, Centrify, CyberArk), SIEMs (McAfee, IBM, Splunk), *etc.*, while also attempting to limit user access and privileges as a means to decrease the overall threat surface. An additional closed-loop DLP application along with risk score alert analysis for SIEM have also been added in its software products. With its solution based on Hadoop and over 200 attribute features available for analysts to tinker with, organizations can achieve fairly customizable solutions according to their security needs.

4.3. HEXADITE

Israel-based startup Hexadite operates in the automated threat detection and response market. For a company that reached US$10 million in early 2016, it has already managed to integrate a significant number of security products in its interoperability portfolio, including AV software (*e.g.*, Symantec, SOPHOS, Kaspersky), endpoint protection (FireEye, RSA, TrendMicro, Cylance), network-based detection (Cisco, Palo Alto), and SIEM vendors (QRadar, ArcSight, Splunk).

Hexadite's main value proposition is the ability to greatly optimize threat detection, alert prioritization, and incident response based on their client's existing security infrastructure investments. The company specializes in bringing faster threat intelligence results, gathering data from any available sources, and partially automating incident response. Most importantly it helps companies cover more ground and deal much faster with security alerts they encounter, and allows them to be more autonomous regarding their own security infrastructure, relieving any ongoing dependence on external cyber security personnel to help them make better sense of their own data.

4.4. MCAFEE

Intel, empowered by a host of McAfee threat defense products, offers one of the most comprehensive SIEM solutions available. Its main components include McAfee's Active Response and ePO (Policy Orchestrator), enterprise security management advanced threat defense, and threat intelligence and data exchange layer. As most SIEMs, the strongest aspects in its skillset relate to data log correlation, data aggregation and correlation, normalization, retention and workflow analysis.

Furthermore, if McAfee has something to contribute it is it's threat detection capabilities. What makes Intel's SIEM solution to stand out is its ability to not only prioritize incoming security alerts through continuous monitoring, but also allow clients to customize their security infrastructure over time through actionable intelligence. This is also reflected by McAfee's database and application data monitors, event receivers, and the risk analysis and historical correlation product called Advanced Correlation Engine. It is the latter that provides real-time identification and scoring of threat events on a rule- or risk-based logic. As mentioned in earlier sections of this report, McAfee also makes strides towards receiving a top mark in interoperability. Its integrated list of partners include automation and orchestration vendors (*e.g.* Ayehu, Phantom), risk and compliance (CyberArk, Beyond Trust), UEBA vendors (Gurucul, Niara), along with application and database security, incident response and forensics, as well as authentication and encryption vendors.

4.5. SYMANTEC

Symantec is perfectly positioned (both market-wise and technology-wise) to instigate a wave of change, and is the perfect transition from traditional signature-based systems into high-end machine learning solutions. The company thrives on end-point security with exceptional anti-zero-day capabilities and multilayered device protection, and offers network monitoring as part of its monitoring process. Symantec is known for deploying cutting edge endpoint technologies (entitled EDR, endpoint detection and response), which, when combined with its ML methods, are focused on preventing exploitation and high-ranking anti-malware support. Due to its ever-growing virus repositories, Symantec is able to easily triangulate and intercept incoming threats with a high mutation rate, as well as instantly kill memory exploitation, prevent ransomware attempts, and enhance network intrusion prevention.

Symantec makes use of nine different approaches based on nine different stages of attack incursion, infection, and exfiltration: 1) network firewall stops malware from entering network traffic; 2) application, file, registry control is funneled through blacklisting/whitelisting points; 3) memory exploitation and zero-days are blocked with popular software being prioritized; 4) reputation analysis leverages community ratings (potentially through sentiment analysis) to determine risk; 5) machine learning algorithms are aimed at detecting potential threats; 6) virtual machines search for hidden malware in custom packers; 7) AV protection activates; 8) behavioral monitoring blocks files exhibiting suspicious behavior; and finally, coming full circle, 9) intrusion prevention systems activate, subverting the threat past the network firewall.

4.6. TRIPWIRE

Tripwire is one of the leading security and compliance solutions providers, focusing in the enterprise, governmental, and industrial markets with clients that include Fortune 500 companies worldwide. Tripwire’s main value proposition is helping organizations detect and deal with incoming cyber security threats through a variety of custom-fitted products. These products include monitoring and prioritization of threats (Log Center), file risk assessment and reconciliation (File Integrity Manager), and perhaps its signature, flagship products, the continuous configuration and compliance assessment (Compliance Manager) and the policy, integrity, and remediation management suite (Tripwire Enterprise). Responding to the emerging threats starting to appear as a result of the device influx and proliferation of industrial IoT systems, Tripwire is attempting to position itself as a leading contender for cyber-security in the constantly evolving smart city technologies, as well as critical city infrastructure.

4.7. SPLUNK

Splunk is positioned in the client-side, real-time UEBA market, but also holds a comfortable position as a security vendor offering more cost-effective solutions. Its flagship product, Splunk Enterprise System, is actually a sophisticated SIEM-like product that makes use of both supervised and unsupervised learning methods, which can also be deployed as a software or as a private, public, or hybrid cloud service. Splunk makes use of both traditional rule- and threshold-based systems for the detection of known threats, as well as more sophisticated unsupervised machine learning algorithms and peer group analysis to detect unusual behavioral patterns. Two important aspects that elevate Splunk's solution are convenient and dynamic data pattern investigations, as well as the ability to teach the system to prioritize detection and investigation according to specific types of behavior.

4.8. TRUDERA

Trudera is positioned in hybrid-based training (client-side and vendor-side) UEBA with hybrid learning algorithms (supervised and unsupervised). Trudera's flagship product called Sentinel is specifically geared to a) lower false positive rates and b) comprehend (or at least learn to comprehend) interdependencies of workloads, applications, and systems. Sentinel is an innovative approach to validate and segregate critical applications to be delivered from different geo-locations. Trudera aims to employ cyber-forensics as part of its UEBA by linking together different events.

Trudera strives to take a proactive approach, make a connection between all those potentially unnoticed events (which have to be selected out of a pool of a million others), predict that a threat is (possibly) on the horizon, and perform the necessary counter-measures before the threat manifests in enterprise systems (*i.e.*, closely monitoring the victims of the previous attacks, raising defense security in appropriate servers, running further checks and updates, quarantining suspicious files, potentially lowering threshold required for security alerts, *etc.*).

4.9. VECTRA NETWORKS

Vectra Networks offers one of the most comprehensive ML solutions in the behavioral and user analytics segment with a sophisticated monitoring process. The company does not use signature-based detection systems but instead uses a purely behavioral analytics approach. As such, it specializes in UEBA, using solely real-time network data traffic analysis in order to flag events or files against endpoint agents, NetFlow, SIEM, and data logs. Vectra Networks is also one of the few companies in the segment striving for a blindspot-free coverage.

Vectra Network's software solution is constantly monitoring user network traffic (packet capture, cache, and metadata), syslogs, and SMTP (simple mail transfer protocol) to scan incoming emails and respond accordingly when a pattern that appears similar to specific clusters of cyber-threats is detected. Among others, this includes detecting threats like ongoing botnets, RATs, port scans (and port sweeps for other devices), outbound DoS, firewall backdoors, sub-OS rootkits, Kerberos servers, ransomware patterns, suspicious updates or transfers, and quite possibly even FTP bounces (file transfer protocols) used to obfuscate an attacker's position.

Published March 30, 2017
©2017 ABI Research
South Street
Oyster Bay, NY 11771 USA
Tel: +1 516-624-2500
www.abiresearch.com