ISSA JOURNAL

November 2016
Volume 14 Issue 11

Beware the Blockchain

PRACTICAL APPLICATION AND USE OF CRYPTOGRAPHY

The Race against Cyber Crime Is Lost without Artificial Intelligence

By Keith Moore – ISSA member, Capitol of Texas Chapter

This article discusses the current shortcomings of the cyber world and how artificial intelligence and machine learning can help you keep your networks safer and more secure.

Abstract

Malware is constantly changing to stay a step ahead of best-of-breed security tools. In the never-ending game of catch-up, we have finally reached a point where traditional network protection tools are no longer enough to keep us safe. New technologies must be explored.

In this new era of computing, artificial intelligence is being used to super charge human intelligence in threat detection, evidence gathering, and remediation. From combating polymorphic malware to helping prioritize analyst workloads, this article discusses what techniques are being used and will be used in the future to finally get ahead of malware.

The cybersecurity landscape over time is analogous to Aesop's fable, "The Tortoise and the Hare." One participant, malware, runs at full speed with no consideration of the race at hand. The cyber defense industry, or tortoise in this scenario, consistently develops defense tools in a slow and steady fashion. While the fable is about consistency prevailing over haste, there is a wrinkle in the cyber landscape that makes it slightly different. In the story, the tortoise and the hare started the race at the same time. In the real world, the hare was given a six-year head start.

The first major "wild" virus, ELK Cloner [7], was developed in 1981. The first antivirus tools, developed independently by McAfee and G Data Software [6], did not arrive until 1987. To address the growing issue of network security, the firewall was created to supplement antivirus toolkits in 1988 by Digital Equipment Corporation. By this time, there were already more than 1,738 unique malware samples lurking in the wild [12]. Cyber defense tools started off well behind the curve.

Of course, since 1987 there has been significant progress in the cyber defense industry. Antivirus companies have consistently developed more signatures, and countless innovations have been made on the firewall. With the invention of packet detection tools, application-level firewalls, and SIEM solutions there are more ways now to protect a network than ever before. The problem is, it is not enough. Malware and exploit kits have been racing ahead at a hare's pace, with a tortoise's consistency. From APTs and spear phishing to ransomware and remote access Trojans (RAT), the dangers of the cyber world have been rapidly outpacing any security tools designed to stop them.

Malware is racing ahead and existing security tools cannot keep up

Looking solely at executable malware, the patterns over the past year are absolutely frightening. Twenty-seven percent of the malware ever created [11] was developed in 2015, and more malware had been developed from January to July of 2016 than in the totality of 2015. According to AV-TEST [3], over 390,000 new malicious programs are registered on a daily basis. The use of domain generation algorithms (DGAs) and polymorphism make malware much more difficult to detect and have led to 78 percent [9] of security analysts no longer trusting the efficacy of antivirus tools. This is a well-founded concern, which was reinforced by the *2016 Verizon Data Breach Investigation Report*. According to the report [18], "99 percent of malware hashes are seen for only 58 seconds or

less. In fact, most malware was seen only once. This reflects the fact that hackers have automated the process of creating new malware variants and underscores the conclusion that automated threat detection and response is absolutely critical moving forward.

Polymorphism and DGAs are significant problems, but distrust in detection and prevention tools only represents part of the challenge analysts face. Another prominent issue is the overwhelming amount of traffic that each security tool creates. As industry moves into a true era of "Big Data," the aggregation tools created to handle the torrent of data generate massive amounts of data themselves. The common big data security tool of today, the SIEM, proves this by the number of heuristic-based alerts that it generates. According to a Damballa analysis [5], the average US business fields 10,000 security alerts per day. These alerts then need to be analyzed, prioritized, and investigated. Many of them end up being false positives, significantly hampering analyst productivity.

In fact, according to Sumologic [1], over 50 percent of analysts cited too many false positives as a significant detractor of SIEM use. The number of false positives delivered within that tidal wave of alerts is significant, as it brings to light another cyber analyst problem—the shortage of personnel to handle the load. According to a Peninsula Press analysis [15] of Bureau of Labor Statistics data, there were over 209,000 unfilled cybersecurity positions in the United States alone in 2015. That number has only risen in 2016, as the market is growing quickly. Demand is expected to rise nearly 53 percent by 2018, indicating that we are in the midst of a serious cybersecurity staffing shortage. With continued network growth due to the number of endpoints coming online (expected to double by 2020 [14]) and enhanced bandwidth, reliance on human scalability will not be enough to alleviate this crisis.

To summarize: more dynamic and polymorphic malware is being created, and heuristic-based security tools are generating too many alerts for the limited analyst workforce to deal with. The steady cyber industry needs a new game plan in order to compete in the race against malware.

Game changer: Artificial intelligence

So, in order to compete with the racing automated capabilities of malware, new technology must be employed. It must be capable of detecting polymorphism, minimizing false positives, and optimizing analyst response.

Enter Artificial Intelligence (AI), the paradigm shift that will revolutionize the cybersecurity industry. It is capable of acting as a human analyst, but tirelessly at machine speed. Unfortunately, AI summons negative connotations in the minds of most, more similar to the *Terminator* movies' Skynet than Pandora's ability to recommend songs. The Skynet comparison could not be further from the truth. Artificial intelligence as it is used today can be more accurately characterized as *Intelligence Amplification,* or IA. Intelligence amplification focuses on improving the efficiency of people, not on replacing them.

In fact, the use of algorithms to amplify the intelligence of humans is completely contrary to the initial goals of artificially intelligent systems. In the first era of artificial intelligence, which came in the 1950s, the goal was to replicate the processes of a human brain on computing devices. Predictions were made claiming that AI would be crowned the chess champion of the world, that AI would discover breakthrough mathematical formulas, and that artificially intelligent systems would completely overtake the workforce in a matter of one to two decades. Unfortunately, these bold predictions never came to fruition in the expected time frame. These excited researchers had failed to accurately predict the challenges they would face, notably the limited computational power available. This brought on an AI winter of sorts, or a lesser-researched period in AI during the 1970s. This winter lasted until early in the 1980s when the second wave of AI began.

The second wave of AI narrowed the scope of what artificial intelligence was intended to do by focusing on building "expert systems." Instead of trying to replicate all of human knowledge like their predecessors, the goal of AI 2.0's expert systems was to perform singular, domain-specific tasks functionally in similar fashion to how a human might perform them. Unlike AI 1.0, there were fantastic successes in this era, like Xcon, developed by DEC [19]. Xcon helped to pair customer needs with the right DEC product, and was estimated to save over $25M per year by reducing technician errors. While it ran on fairly simple heuristics, it was astoundingly more productive than the state of the art at the time. Howev-

Polymorphism

Polymorphism is the capability of certain files to replicate themselves with slight code mutations while keeping the original functionality of the code intact. For a malicious file, polymorphism increases the chances of infecting a system because signature-based detection tools do not immediately recognize it. These tools miss the file because the payload is often encoded or encrypted, and this part of the code looks different on every replication. Machine learning techniques are capable of recognizing these files for two reasons:

- These files often have high entropy (randomness) or other anomalous features that indicate they are highly compressed or seemingly random
- The unencrypted decryptor used within the file can often be associated with other known malware types

Think of polymorphic malware as you would the flu. You can get a flu shot, but that is not a 100 percent guarantee you will not contract the flu. What the shot does is train your body to recognize certain variants of the flu virus. There are still certain mutations that your body will not recognize that can still make you ill.

er, Xcon and similar systems were not without their limitations. AI 2.0 systems quickly proved too difficult to maintain, and were still prone to making extreme errors that caused catastrophic mistakes in critical work flows. These deficiencies, coupled with a downturn in the AI hardware market, resulted in a second AI winter that lasted until the late '90s.

This brought the world to the artificial intelligence of today, or AI 3.0. Often synonymously interwoven with the term *Machine Learning*, the third wave of AI is specifically focused on creating intelligent software that can adapt to new data to better predict, optimize, classify, or understand the behavior of a specific system. Thanks to innovations brought on by Moore's law, physical compute has finally resulted in artificially intelligent systems capable of surpassing human capabilities consistently on the domain-specific level. Complete human intelligence and capacity have yet to be replicated, but many tasks today can be done significantly faster and better by a machine than by a human expert. For example, it is now nearly impossible for a human to beat a well-trained computer in chess; social media feeds can customize themselves to display content that they believe you are interested in; and self-driving vehicles are using artificial intelligence to safely traverse roadways. We have finally reached a point where artificially intelligent systems are capable of amplifying human intelligence.

Luckily, there are numerous opportunities for intelligence amplification in the cyber landscape. By applying advanced algorithms like Bayesian reasoning, deep-learning neural networks, natural language processing, and more, it is possible to automate and predict many of the processes that a cyber analyst might initiate while running threats to ground. With the automation of standard processes, analysts can then supervise their ever-increasing network volume and focus on making key decisions.

AI and polymorphic malware

Let us begin with the first major challenge—handling adaptive polymorphic malware and detecting malicious intent. Think of the techniques a human analyst would go through in order to identify malicious files. A person would begin by looking at every file moving through his network, either via proxy logs, firewall logs, machine syslogs, or a network cache. Instead of analyzing every file, which would take a significant period of time, he would inspect the unusual or "anomalous" ones. These could be files that were downloaded less frequently, those downloaded from non-whitelisted websites, or anything else deemed by a particular analyst as suspicious. From there, each file might be hashed and compared against an existing threat repository. After that, the analysis must be taken offline and dug into on site.

The onsite work of file investigation is slightly more laborious. Each anomalous file worth digging into is usually run through an antivirus tool at this point. It has already been established that antivirus tools are not trusted by security experts. However, in this case, it is necessary to drive home exactly how bad they can be by digging into a series of OPSWAT reports that came out in 2015. According to the January 2015 report, APTs were present on 0.7 percent of all devices [2]. According to that same report in August of 2015, that number had skyrocketed to 4.4 percent [17] of all devices, noting that "even though the [anti-malware] product may detect a threat, further actions may be required to remediate the threat, which is not always apparent from the anti-malware product alone." The four percent jump in only seven months indicates that either malware advanced exponentially, or the tool was behind the curve to begin with and was not able to detect malicious APTs.

In order to get around this, analysts often choose to detonate files in a sandbox environment, as sandboxes are marketed as effectively identifying zero-day malware. The truth is, they are far from a silver bullet [13], as has been called out by cyber researcher Christopher Kruegel. As usual, malware has significantly outpaced any sandboxing advancements, using various techniques to identify sandbox environments and avoid detonation. These techniques, while enumerated in a white paper published in early 2016 by the SANS Institute [10], include delaying execution, sandbox diagnosis, and human activity monitoring. In fact, Symantec believes that 16 percent of malware [8] samples are "virtual machine aware," meaning it is unlikely they will be captured by a sandbox.

So how does machine learning solve these problems? Unlike sandboxing environments, machine learning techniques can automatically analyze files statically, significantly reducing the amount of time required when compared to sandboxing or offline analyst-driven analysis. Instead of looking for sig-

Domain generation algorithms (DGAs)

A domain generation algorithm is a code component within malware that generates a large number of domains for the malicious file to try and access. These generated domains are then used as a bridge to existing command and control servers that simply need to register a single one of the many generated domains in order to gain control over all enslaved devices. DGAs are so beneficial to hackers because they avoid storing a domain list within a malware file, making it much more difficult for security tools to easily identify malicious command and control domains.

DGAs first ingest a seed component as an input. This seed can be as simple as a few integers stored within the file or as complex as dynamically pulling information (such as the price of the USD versus the Euro) off the Internet at a given time. This seed input is fed into the domain generation function and results in the output of seemingly random domain strings to which .com, .net, or .ru can be appended (resulting in many domains that might look something like this: sxy38rwaqyt7.com). The malware file then reaches out to each of these seemingly random domains to eventually bridge back to its dedicated control server and determine if it should perform any actions.

natures or analyzing file execution, machine learning techniques can break files into millions of static pieces and conclude the likelihood that they could comprise malware. To do this, advanced math and data science techniques are employed. Whether via a simple strings analysis, or by looking at more complex features like file entropy or headers, every file is first deconstructed into what is referred to as a "feature set" for an algorithm to look at. The algorithm can then derive correlations between the feature set and generate even more data to inspect.

Once a feature set, termed the file's "DNA," has been generated, the true machine learning work begins. By training and learning from the DNA of millions of known malware files and millions of known benign files, different algorithms are able to pinpoint all of the behaviors and characteristics of files most frequently associated with malware. These algorithms use techniques like *Tree-Based Bayesian Reasoning* and *Deep Learning* to analyze files and create predictions. These analyses easily surpass the capabilities of a human expert because they evolve to understand the file DNA in ways that even the most intelligent of people are unable to do. In order to understand why that is the case, it is important to look a bit further under the hood.

Bayesian reasoning explained

Bayesian reasoning is a statistical concept postulating that the probability of an event occurring can be defined by conditions relating to that event. For example, if the age of a person is related to how likely she is to attend an educational institution, information about a person's age should help to more accurately assess the probability that she attends school. In machine learning, Bayesian reasoning is often used to automate the creation of complex probabilistic trees that can more accurately assess the likelihood of certain events occurring. Think of them as extremely complex Boolean decision trees that help to understand data relationships. In malware detection, these trees can extend to thousands of "branches" to help better classify a file as malicious or benign. A simple Bayesian malware detection relationship might look like this: if a file is packed, meaning it contains high levels of encryption and compression, it is 50 percent more likely to be malicious than benign. However, that relationship can then be further defined by factoring in other DNA features, such as certain function calls contained within the file, file entropy, header content, and more. By adding these features and additional "branches," the algorithm is much more accurately able to differentiate between malicious files and benign files.

File entropy

File entropy, at a high-level, is the measure of randomness present in a file. It is calculated by incrementing the number bytes in a file and identifying how frequently they appear relative to file size. For example, a file that contained all zeroes would have a very low entropy score whereas a file that contained all 1 byte values would have a very high entropy score. Entropy was originally designed for use as a methodology for determining how much compression could be applied to a file, with each file getting a score between 0 and 8. The closer a file is to an entropy score of 8, the less that it can be compressed. Today, it is more often used as a tool to recognize packed and compressed files, which generally contain higher amount of randomness.

Deep neural networks explained

Where Bayesian reasoning primarily optimizes complex if-this-than-that rules when building tree-structures, neural networks add another level of complexity. Instead of branching at Boolean logic, neural networks implement a neuron architecture, much like the human brain, when building models. These neurons accept input data and then use embedded computational functions to decide if and when to pass information to the next layer. A deep neural network is a network that contains many hidden layers of neurons and most often many neurons per layer. For malware detection, as the DNA of every file is passed through these hidden layers, analysis occurs helping to transform the data and identify if the different pieces distinguish the file as more likely to be malicious or benign. Because of all of these transformations and complex neuron firing decisions, neural networks are very difficult to explain at a high level. However, they have proven highly effective in solving constrained problems—like malware detection.

An easier approach to understanding how machine learning malware detection works is to compare it to something simple—like Lego building blocks. Think of a Lego pirate ship. It is the scourge of the Lego seas, and Lego villagers fear it for all of the damage it might do to their Lego town. The Lego pirate ship is malware in this scenario. If a person were to take all of the pieces that comprise this Lego pirate ship, he could use them to create numerous other Lego structures—a Lego house, for example. These pirate ship pieces could actually be put together to create a fairly convincing house. That house would not be without flaws, however. There would still be specific components that do not quite fit in perfectly, such as the ship helm or the pirate flag. When hidden within the house, they might be difficult to notice as outliers. But if the house were to be broken apart, and every piece laid out individually, it would be obvious that they are out of place. With careful inspection and consideration of these outliers, a well-trained mind could look at all the individual Lego bricks and conclude that they are in fact looking at a malicious pirate ship, not a house at all!

Machine learning malware detection takes the same approach, except it is capable of analyzing the millions of pieces that can make up a file's DNA, not just the hundreds of Legos that make up a pirate ship. These complex algorithms are able to identify the malware lurking beneath the seemingly benign facade.

Unlike the Terminator, this technology is not science fiction. Malware detection algorithms have already moved

far beyond the academic world and are impacting industry. These out-of-the-box machine learning tools are being used to significantly enhance the security of critical endpoints at financial institutions, retailers, technology companies, and governments. Companies like SparkCognition, Cylance, Invincea, and a few others are employing machine learning antivirus at over 99 percent effectiveness with minimal false positive generation. The techniques combat polymorphism and domain generation algorithms because of the complex techniques innate to the machine learning process. Instead of looking for a hash or a heuristic match, pattern detection and correlation are employed to identify similarities to other known malicious files. Even in the case of handling packed and obfuscated files, machine learning techniques can quickly identify and block any malicious intent present on a client endpoint. Best of all, these techniques learn malicious patterns over time as new file types and threats are discovered. Using semi-supervised and reinforcement techniques that are able to learn from user input, these algorithms can train themselves to better recognize new malware variants, maintaining pace with ever-evolving malware mutations.

AI and making analysts more efficient

The second major risk faced by the cyber industry is the rising tide of security alerts coupled with the insufficient number of analysts in place to handle them. In order to address these problems, there is only one reasonable path forward—make the workforce more efficient. The most obvious way to do this is through better filtering and prioritization of threat streams. In existing heuristic-based security systems like SIEMs, Snort, and firewalls, filtering and prioritization capabilities are fairly limited. Most alerts—from an advanced persistent threat to a blacklist hit—are generated with equal weight. This is not sustainable moving forward. Instead, artificial intelligence can be used to augment human analysts and prevent them from being overwhelmed.

First, it is necessary to understand what an analyst must currently do in order to analyze and prioritize threats. Once an alert has been triggered in a security tool, most experienced analysts will rely on prior-knowledge to gain a bit of context. If only given a small number of alerts, an analyst can quickly identify that a downloaded file matching the signature of a known Trojan is more important than a blacklist hit from an employee's web browser. However, when inundated with a large number of alerts, it is very difficult for inexperienced analysts to even know where to start. Most prioritize by alert type, but this methodology is not always foolproof [4]. They can then use other tools to dig deeper into each alert and run it to ground. Finally, if experience and existing tools are not enough, an analyst can then call on an even more robust tool—the Internet. According to LookingGlass Cyber Solutions [16], outside of traditional security tools, the most valuable resource a researcher can have at his or her disposal is a search engine. Using Google or Bing, the two most popular search engines, an analyst can build a hypothesis that an alert is legitimate or a false positive. However, this conclusion is open to interpretation, as human error is possible. This process can also take hours or days for some alerts to be investigated.

Artificial intelligence algorithms can automate this process for an analyst and help to present only the relevant information on which a decision can be made. Not only are these algorithms capable of ensembling all of the disparate data sources, but they can do so in real time. For example, algorithms are capable of looking at a simple heuristic match, like an anomalous domain accessed, and quickly diagnosing its level of malice. They can analyze the domain, the user agent, the port used, the geo-location, and a number of threat feeds

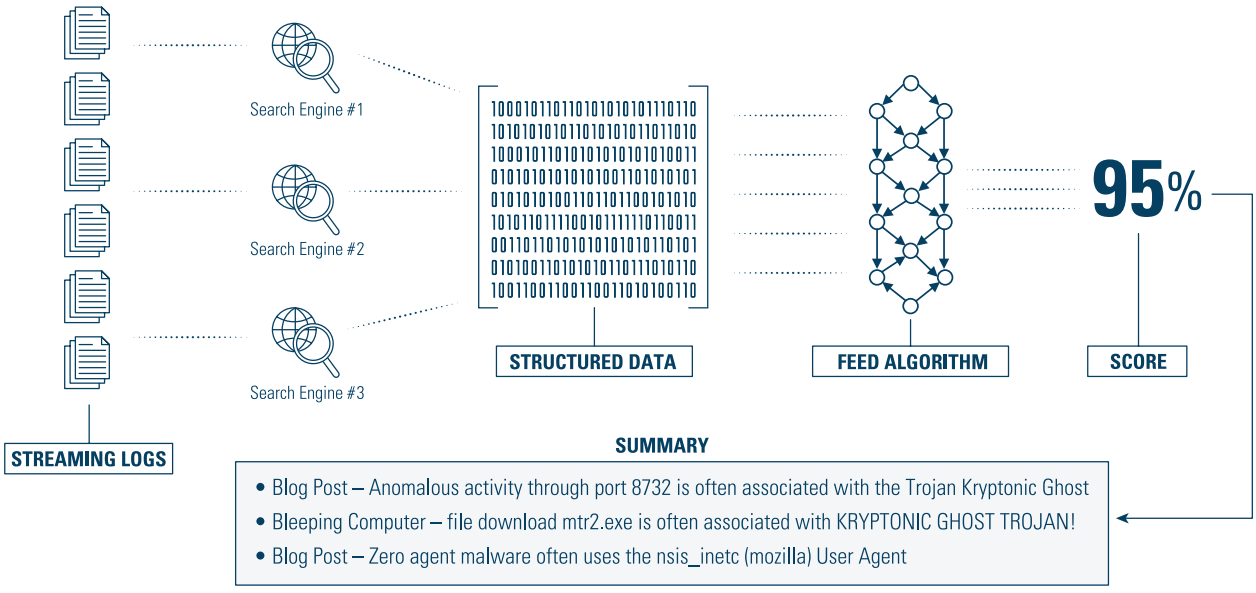

Figure 1 – Automated event research, prioritization, and summarization using natural language processing

to determine if the activity was really threatening or just another false positive. Separate algorithms can then compare this type of event to other events, creating a total "risk score" for each one, helping an analyst to prioritize her workload.

Machine learning algorithms can also automate analysts' research to help them better understand network activities. Using a new-to-security technology called Natural Language Processing (NLP), it is possible for algorithms to read the Internet. For example, if an employee anomalously used port 8888 to access a site in China, NLP algorithms could automate dozens of Google queries around that type of activity. They could then read through the hundreds of returned pages to diagnose if the observed behavior is more likely to be malicious or benign. How this works technologically is quite unique. Unlike file analysis, where there is a very specific set of features that can be created, the Internet is not so orderly—it is a mess of content in thousands of varying formats. Machine learning algorithms prefer to look at highly structured data, so the very first step required in order to perform cyber research is to structure the content retrieved in each query. This is done by using techniques like TF/IDF, word vectorization, sentiment analysis, and more. Once structured, all content can be run through algorithms like those described for file analysis to predict how likely it is that each document, paragraph, or sentence is referring to malicious or benign behavior. This information can then be cross-referenced with the original query and scored by confidence to define if the behavior exhibited in a specific network is correlated with what is being read and classified on any given webpage.

To provide a more visceral example, a user agent containing "() { :;}; /Bin/Bash" does not necessarily strike fear into the heart of a junior analyst. However, after investigating this anomalous user agent on the Internet, an NLP-based security system will quickly recognize that this pattern is associated with the Shellshock exploit and flag it as extremely malicious.

Automated research capabilities and risk assessment brought about by use of natural language processing in security can significantly help to automate the prioritization of threats. In the hours to days it might take a human analyst to perform this research, an algorithm with a decent amount of bandwidth (i.e., Gb/s) can complete all of the research and scoring necessary in less than a second. Better yet, it can do it for every alert that is ever generated by any security component on site in real time.

But the scoring and prioritization alone are not good enough to be considered beneficial intelligence amplification in security. The automated threat research mentioned above provides no value at all unless the security professional responsible for making decisions believes in it. That is why these algorithms have been designed to go a step further than simply scoring and prioritizing content. As can be viewed in figure 1, machine learning and natural language processing algorithms are also capable of distilling all relevant content—from geo-location of IP accessed, to the three paragraphs on the Internet that best represent the given risk score—into something easily interpreted by an average cyber practitioner. This allows analysts to quickly make the right decisions while relying on exhaustive research and robust analysis.

A world with AI – Moving forward

Racing ahead at a hare's pace, malware has come a frighteningly long way since ELK Cloner. Threatening files have evolved to the point where they are no longer traceable using traditional signature-based tools and mutate themselves on every replication. The security tools in place to identify these threats now have analysts facing more alerts than ever before, and there is no good way to handle them all.

Luckily, our tortoise, the defense industry, has advanced rapidly as well. Artificial intelligence algorithms including machine learning and natural language processing provide

the weapons necessary to hunt polymorphic malware and persistent threats. Without AI, we are surely too far behind malware to ever hope to catch up. With it, there is hope that artificially cognitive systems can amplify the protective capabilities of every cyber defense in the world. It is time for a paradigm shift in how we go about security. As malware becomes automated, intelligent, and aware, so must cybersecurity.

References

[1] "Achieving Continuous Intelligence with Advanced Security Analytics." Sumologic Inc., November 2015. Accessed August 30, 2016 – https://www.sumologic.com/wp-content/uploads/2015/11/Achieving-Continuous-Intelligence-with-Security-Analytics.pdf.

[2] "Antivirus and Compromised Device Report: January 2015." OPSWAT. January 23, 2015. Accessed August 30, 2016 – https://www.opswat.com/resources/reports/antivirus-and-compromised-device-january-2015.

[3] "AV-TEST – The Independent IT-Security Institute." *Malware Statistics & Trends Report*. August 25, 2016. Accessed August 30, 2016 – https://www.av-test.org/en/statistics/malware/.

[4] Downing, Larry. "Target Says It Declined to Act on Early Alert of Cyber Breach." *Reuters*. March 13, 2014. Accessed August 30, 2016 – http://www.reuters.com/article/us-target-breach-idUSBREA2C14F20140313.

[5] Dunn, John E. "Average US Business Fields 10,000 Security Alerts per Day, Damballa Analysis Finds." *CSO Online*. May 14, 2014. Accessed August 30, 2016 – http://www.csoonline.com/article/2154861/average-us-business-fields-10000-security-alerts-per-day-damballa-analysis-finds.html.

[6] "G Data Presents Security Firsts at CeBIT 2010." G DATA Software AG. February 18, 2010. Accessed August 30, 2016 – https://www.gdata.pt/central-de-imprensa/reportagens/detalhes-de-noticias/article/1532-g-data-presents-security-first.

[7] "History of Viruses." NIST. March 10, 1994. Accessed August 30, 2016 – http://csrc.nist.gov/publications/nistir/threats/subsubsection3_3_1_1.html.

[8] "Internet Security Threat Report." April 2016. Accessed August 2016 – https://www.symantec.com/content/dam/symantec/docs/reports/istr-21-2016-en.pdf.

[9] Keane, Jonathan. "Is Anti-Virus Enough? Security Professionals Say Preventative Measures Are Much Stronger." Digital Trends. July 06, 2015. Accessed August 30, 2016 – http://www.digitaltrends.com/computing/anti-virus-isnt-enough-security-professionals-say-preventative-measures-are-the-future/.

[10] Keragala, Dilshan. "Detecting Malware and Sandbox Evasion Techniques," SANS. January 16, 2016. Accessed August 30, 2016 – https://www.sans.org/reading-room/whitepapers/forensics/detecting-malware-sandbox-evasion-techniques-36667.

[11] Korolov, Maria. "27% of All Malware Variants in History Were Created in 2015." *CSO Online*. January 29, 2016. Accessed August 30, 2016 – http://www.csoonline.com/article/3027598/cyber-attacks-espionage/27-of-all-malware-variants-in-history-were-created-in-2015.html.

[12] "Lower Security Risks and Costs by Minimizing the Time to Protection," Trend Micro. January 2009. Accessed August 23, 2016 – http://www.techdata.com/techsolutions/Softwareconnections/files/april10/TREND_ENT-SECWhitepaper.PDF.

[13] Messmer, Ellen. "Malware-Detecting 'Sandboxing' Technology No Silver Bullet." *Network World*. March 26, 2013. Accessed August 30, 2016 – http://www.networkworld.com/article/2164758/network-security/malware-detecting--sandboxing--technology-no-silver-bullet.html.

[14] Middleton, Peter, Thilo Koslowski, and Anurag Gupta. "Forecast Analysis: Internet of Things - Endpoints, Worldwide, 2015 Update." Gartner. December 15, 2015. Accessed August 30, 2016 – https://www.gartner.com/doc/3178626/forecast-analysis-internet-things-?__hstc=83621449.89080f868519c6d129b88865722cb413.1472137138827.1472487100704.1472575273394.4.

[15]Setalvad, Ariha. "Demand to Fill Cybersecurity Jobs Booming - Peninsula Press." *Peninsula Press*. March 31, 2015. Accessed August 30, 2016 – http://peninsulapress.com/2015/03/31/cybersecurity-jobs-growth/.

[16] Shames, Alyssa. "Three Tools Every Security Analyst Needs," LookingGlass Cyber Solutions Inc. October 15, 2014. Accessed August 30, 2016 – https://www.lookingglasscyber.com/blog/threat-intelligence/three-tools-every-security-analyst-needs/.

[17] "Top Anti-Malware Vendor Market Share and Device Security Report." OPSWAT. August 21, 2015. Accessed August 30, 2016 – https://www.opswat.com/resources/reports/anti-malware-market-share-security-august-2015 - anti-malware-vendor-market-share.

[18] "Verizon's 2016 Data Breach Investigations Report." Verizon Enterprise Solutions. April 2016. Accessed August 30, 2016 – http://www.verizonenterprise.com/verizon-insights-lab/dbir/2016/.

[19] Winston, Patrick Henry, and Karen A. Prendergast. *The AI Business: The Commercial Uses of Artificial Intelligence* (Cambridge, MA: MIT Press, 1984).

About the Author

Keith Moore is the Director of Product Management at SparkCognition, responsible for the development of both their IoT and cybersecurity product lines. He specializes in applying advanced data science and natural language processing algorithms to complex data sets. He previously worked National Instruments as a analog-to-digital converter and vibration software product manager. He may be reached at kmoore@sparkcognition.com.