BIG DATA ANALYTICS AND CYBERSECURITY:
Three Challenges, Three Opportunities

Presented by

the Armed Forces Communications and Electronics Association Cyber Committee

December 2017

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This white paper by the Armed Forces Communications and Electronics Association (AFCEA) Cyber Committee provides recommendations on the applications of big data analytics and data science generally to the cybersecurity domain. It examines ways in which big data can be used to improve predictive analytics and to detect anomalous behavior that may be indicative of cybersecurity problems such as exploits or attacks. This paper also examines the special challenges the security of big data environments pose given the enhanced value of information that is made part of and subject to analysis within such environments. In addition, it discusses the implication of the use by foreign intelligence services and cyber criminals of big data analytics in the exploitation of large databases and repositories, e.g., the data extracted from the U.S. Office of Personnel Management (OPM).

Overall, this paper recommends research and development the government and private sector can conduct regarding ways in which big data analytics can secure complex networks and environments. It also recommends enhanced, enterprise-level security regarding big data environments. Finally, it recommends stronger efforts by the Intelligence Community to understand how adversaries may be using big data analytics to understand the United States and craft courses of action that affect national interests.

SCOPE

This white paper discusses ways both the government and the private sector can use big data analytics to improve predictive analytics relating to cybersecurity problems. It pays special attention to the challenges associated with large databases as well as data environments comprising many smaller databases, particularly those with national security importance that might be exploited and subjected to big data analytics by adversaries. In addition, it addresses the need to secure data environments used for national security purposes in which big data analytics tools are used to support high-value decisions. Given the potential big data analytics hold for the analysis of large data sets, it also discusses the use of such tools by foreign adversaries and cybercriminals who may wish to understand and intervene in large-scale national security, political, diplomatic and economic developments in a manner that affects the interests of the United States.

PROBLEM STATEMENT

Complex networks comprising traditional information technology systems, the Internet of Things, including critical infrastructures, and multiple cloud environments present cybersecurity analysts with difficulty in detecting, preventing and mitigating sophisticated cyber exploits and attacks. Such complex networks may be difficult to characterize in terms of either baseline behavior or the anomalous behavior that may or may not be indicative of a cybersecurity problem. This white paper describes ways in which big data analytics can be used to improve our understanding of baseline and anomalous behavior caused by cybersecurity problems generally and in complex, dynamic networks in particular.

Progress in this area is vital as networks requiring effective defense become more complex. The use of technical tools to compare anomalous and baseline behavior in the detection of cybersecurity events has proved difficult given the complexity of contemporary networks and challenges of differentiating real cybersecurity issues from other issues such as unusual but legitimate user activity; changes in network topology, particularly in information technology/operational technology (IT/OT) systems; and information technology malfunctions. This challenge, known as the false positive problem, results in the presentation of too many indications, most of which are not indicative of a cybersecurity problem, or too few indications, masking dangerous cyber exploits and attacks.

Organizations that rely on information have seen amplification—perhaps by orders of magnitude—in the value of the information they must protect. Work done in the 1980s by Michael Porter of the Harvard Business School identified the concept of "information intensity." Porter and others argued that organizations that produce information (e.g., Dow Jones) or use information to coordinate the creation and production of their products (e.g., Walmart, Boeing) are highly information intensive. The theft or destruction of the information on which these organizations depend can be fatal. At the same time, the exploitation of such valuable information can be of significant value to the organizations capable of such exploitation.

Photo credit: Solarseven, Shutterstock

Porter's work in the 1980s, however, should be considered in light of the value that big data analytics adds to what had been largely individual facts. Big data analytics allows Walmart and other retailers to understand and predict the demand for the goods on its shelves and to engage in pricing strategies that consider customer behavior, competitor activities, supplier pricing and other factors. Big data analytics helps pharmaceutical companies structure complex research activities and interpret complex results. Financial services firms use big data analytics to understand future markets; governments use big data to understand subjects as diverse as agricultural production and the demographics of populations that need assistance in the wake of natural disasters. As a result, the information intensity of many organizations has risen in proportion to their use of big data analytics, as has the value of the information aggregated and analyzed using such tools, raising considerably the stakes in the effective cybersecurity of this information.

The recent breaches of OPM and other large data environments also call into question the way adversaries are using their big data analytics to understand large-scale demographics at a macro level and other trends associated with the millions of people whose records have been exposed. Adversaries may also use big data analytics to gain significant insight into U.S. national security decisions, the economy and even political dynamics.

Understanding how adversaries might use these tools is important to understanding the implications of such breaches and to anticipate the use of this information by adversaries. Such use of exploited information could include efforts to analyze trends in U.S. research and development in critical and sensitive industrial and technology areas, or even to spot trends in the behavior of persons granted security clearances. As a result, the U.S. national security and intelligence communities should seek to understand how foreign intelligence services and cybercriminals may be using their big data analytics to amplify the benefit they gain through the exploitation of U.S. computer networks.

APPROACH

The white paper defines specific problems and challenges associated with:

- securing large-scale data environments;
- using big data analysis to improve the cybersecurity of complex networks; and
- understanding how big data analytics can be used by adversaries to enhance the value of information they exploit.

It suggests lines of research and development as well as best practices that should be employed by government and the private sector, particularly in support of environments important to national security.

To develop this white paper, AFCEA Cyber Committee members:

- met with leading cybersecurity research and development authorities within the federal government and industry;
- held discussions with the National Counterintelligence and Security Center (NCSC) to understand the intelligence and counterintelligence value of information gleaned by and from big data environments;
- developed draft findings and recommendations;
- drafted a presentation to the Cyber Committee;
- refined the findings and recommendations;
- developed a stakeholder engagement approach to present the findings and recommendations; and
- presented the results to stakeholders.

THE CHALLENGE OF COMPLEX NETWORKS

Even as computer networks[1] are becoming more complex, so too is their composition. Newer and emerging networks combine "traditional" information technology, including business applications, media, analytic and other corporate functions with the Industrial Control Systems and Supervisory Control and Data Acquisition (SCADA) systems that collected data from and manage today's manufacturing, energy, transportation and other infrastructures. Such networks can change frequently as new devices are installed, subnetworks established and older equipment decommissioned. More networks than ever are cloud-based; support to industrial infrastructures from cloud-based networks is becoming a reality. The desire to move information processing workloads among clouds, known as orchestration, as well as the scale and flexibility of today's cloud infrastructures are creating computer networks with baselines difficult to characterize and in which anomalies caused by cybersecurity exploits and attacks can be difficult to prevent, detect and mitigate.

Such networks pose operational and management challenges as well. Responsibility for the management and cybersecurity of such networks may be distributed and possibly uncoordinated. This diffuse responsibility can be accompanied by a lack of an integrated view of network operations and behavior. A power grid connected to factories, homes, offices, hospitals, schools and government facilities may encompass devices managed by both the electrical power grid operator and the myriad organizations, large and small, that the grid serves. In the near future, smart grids are likely to be connected to smart roads that will mediate access, traffic flows and energy use. It will be difficult enough to characterize baseline behavior of such networks, more difficult to ascertain anomalous behavior because of cybersecurity attacks and exploits, and equally difficult to manage cybersecurity across such a disparate environment. Cybersecurity managers may be overwhelmed by problems associated with endpoint protection on networks subject to disparate management and characterized by dynamic topologies.

Big data analytics may, at least in part, hold the key to meeting this challenge. The Intelligence Community has built an understanding of the challenge big data poses as unrelenting increases in the volume, velocity and variety of information with which it must contend. The community's struggles with the global data

environment may yield lessons valuable for the present inquiry. However, perhaps no more useful or simpler definition of big data analytics is available as the one in the Webopedia, which notes:

Big Data analytics is the process of collecting, organizing and analyzing large sets of data (called Big Data) to discover patterns and other useful information. Big Data analytics can help organizations to better understand the information contained within the data and will also help identify the data that is most important to the business and future business decisions. Analysts working with big data basically want the knowledge that comes from analyzing the data.[2]

This definition says as well:
High-Performance Analytics Required
To analyze such a large volume of data, Big Data analytics is typically performed using specialized software tools and applications for predictive analytics, data mining, text mining, forecasting and data optimization. Collectively, these processes are separate but highly integrated functions of high-performance analytics. Using Big Data tools and software enables an organization to process extremely large volumes of data that a business has collected to determine which data is relevant and can be analyzed to drive better business decisions in the future.[3]

The Webopedia adds this caution, however:

> ***For most organizations, Big Data analysis is a challenge. Consider the sheer volume of data and the different formats of the data (both structured and unstructured data) that is collected across the entire organization and the many ways diverse types of data can be combined, contrasted and analyzed to find patterns and other useful business information.[4]***

This definition is rich in ways pointing to the promise of big data analytics to help meet the challenge of securing complex networks, though the cybersecurity of such networks will require other components as well, including new management models that create collective responsibility for the cybersecurity of interconnected and interdependent networks. It notes that big data analytics can help discover patterns of useful, relevant information; can help drive better business decisions; and can be used to find patterns and other useful business information.

Work to apply big data analytics to the challenges of cybersecurity has been taking place for several years. In 2013, authors Tariq Mahmood and Uzma Afzal surveyed the use of big data analysis in cybersecurity and noted:

> ***Analytics can assist network managers particularly in the monitoring and surveillance of real-time network streams and real-time detection of both malicious and suspicious (outlying) patterns.[5]***

Photo credit: Macrovector, Shutterstock

The authors make a useful distinction between "malicious" and "suspicious (outlying)" patterns of behavior, demonstrating awareness of the need to identify activity that may or may not be indicative of a cybersecurity problem but that still warrants investigation. In other words, the authors point to the need to minimize "false positives" if big data analytics are to be useful to cybersecurity.

Several companies are tackling the challenge using big data analytics poses to improving cybersecurity. IBM is combining its QRadar Advisor used as a security enterprise information management (SEIM) tool with the intelligence built into its Watson supercomputing platform to integrate numerous sensors, unstructured data and network security incidents to create a more predictive environment. Blue Coat and Symantec are creating cloud SEIMs designed to detect and manage cybersecurity incidents in cloud environments.

Perhaps even greater promise may lie in the adaptation of GE's Predix platform, which is designed to move the analytics and management of industrial systems to the cloud. GE's rationale for creating this platform is compelling; the firm notes:

> ***Investment in the Industrial Internet of Things (IIoT) is expected to top $60 trillion during the next 15 years[6] …(a)nd by 2020, over 50 billion assets will connect to the Internet.[7]***

Other Predix features are worth noting. Predix operates at the edge, meaning that it has insight into and can manage endpoint devices while providing cloud-level efficiency. The platform is designed explicitly with a view toward the creation and deployment of new analytic applications; its Predix Machine allows for the collection and analysis of data from myriad endpoints, providing a starting point for additional analytic technologies. In addition, the Predix platform itself allows for both high-level integration of connected environment and their segmentation to improve security and privacy. Overall, the Predix model argues for a combination of both big data analytics and the use of more modern cloud architecture for complex networks that include industrial systems also known as operational technology.

Government technology advances also are promising. The Defense Advanced Research Projects Agency's Big Mechanism program "aims to develop technology to read research abstracts and papers to extract pieces of causal mechanisms, assemble these pieces into more complete causal models, and

Photo credit: Tommy Lee Walker, Shutterstock

reason over these models to produce explanations."[8] Although Big Mechanism is intended to support cancer research, its ability to address "causal models" might help mitigate the challenges associated with false positives, i.e., the problem of identifying too many examples of anomalous behavior without clear causality by a cyber exploit or attack.

Other government efforts, particularly those undertaken by the Networking and Information Technology Research and Development (NITRD) Program, also may offer promise. The Big Data Interagency Working Group (BD IWG) focuses on research and development to improve the management and analysis of large-scale data. The group's purpose is to develop the ability to extract knowledge and insight from large, diverse and disparate sources of data, including mechanisms for data capture, curation, management and access. The AFCEA Cyber Committee recommends this effort be given the additional charter to help identify anomalous behavior in the complex networks the nation must secure.

Despite these advances and programs, significant hurdles remain. Today's complex network environments are managed on a distributed and sometimes multi-enterprise basis. Agreement will be required to gain access to the data generated across these interconnected networks that comprise these environments. This agreement will require both data transparency and strong assurance—related to provenance and accuracy—as well as anonymity to protect personal identifiable information (PII), proprietary information and other sensitive information.

THE RISING VALUE OF INFORMATION

The role of information within enterprises is changing; it is growing more important and helping shape the view of cybersecurity. The importance of information can be viewed as an enterprise's "information intensity." In the general economy, information—and by extension, its security—is recognized as an essential aspect of corporate strategy and, more importantly, as an enterprise's overarching value proposition.

The concept of information intensity reflects the recognized value of information. This concept has existed for decades, but it gained currency in the 1980s and has grown in importance through the present day. Two types of information intensity were defined in the 1980s, and both are vital to today's enterprise: product information intensity and value chain information intensity.[9]

Product information intensity measures the extent to which a product is information-based (i.e., information-as-product), which is increasingly the case in today's global economy in general and in the United States and other advanced economies in particular. Any business that provides information-for-value (e.g., financial reporting and transactions, media and social networking) delivers one or more products that comprise principally or solely information. For such enterprises, the security of the information they employ and provide affects materially the value of the product they convey to their customers. Their value proposition can exist and thrive only to the extent cybersecurity and information assurance relating to provenance, processing and delivery are present.

Value chain information intensity is the extent to which information contributes to the production and delivery of non-information products. Global supply chains for the manufacture of aircraft, for example, rely on a complex web of information ranging from specifications and test data to pricing and delivery schedules. Every element of this information is crucial to production. In fact, many of the processes used in manufacturing are information-technology controlled, enhancing the level of information intensity on which these products and their value chains rely. A cybersecurity failure in these value chains can result in faulty parts, dangerous industrial operations, loss of intellectual property and non-delivery of the product as promised.

Linked to value chain intensity is the extent to which many physical products such as airliners are characterized by an increasing proportion of information technologies. Today's Boeing Dreamliner, for example, uses computer-based fly-by-wire technologies to control critical flight systems. It possesses Internet-based architectures for other systems ranging from avionics to passenger entertainment subsystems. In many ways, the Dreamliner is a computer around which someone designed an airplane. In Boeing's own parlance:

> ***The 787 Dreamliner, the world's first e-Enabled commercial airplane, combines the power of integrated information and communications systems to drive operational efficiency, enhance revenue, and streamline airplane maintenance.***[10]

Boeing also notes:

> ***These tools promise to change the flow of information and create a new level of situational awareness that airlines can use to improve operations. At the same time, the extensive e-Enabling on the 787 increases the need for network connectivity, hardware and software improvements, and systems management practices.***[11]

The importance of the concept of information intensity is not new. Compelling work by Michael E. Porter and Victor A. Miller in 1985[12] described the value of information in both information-as-product and in value chains. The authors defined the concept of manufacturing information and distribution systems (MIDS), noting that "an information intensive MIDS will generally bring value to a company if it adds high value to the product."[13] In today's world, such systems are of vital importance.

Whether an enterprise delivers information itself as a product or provides products that rely on information to empower and mediate their value chains, cybersecurity clearly bears directly on information intensity, corporate strategy and the value proposition an enterprise delivers. Indeed, the cybersecurity of information-intensive products is intrinsic to the value of those products and rises, therefore, to the level of a corporate strategic issue.

Recent research increases the importance of the concept of information intensity as well as intensifies the urgency of focusing on cybersecurity. For example, research provides powerful evidence about information-intensive businesses that produce information-as-product: These businesses should use information technology to disaggregate their production for the purpose of efficiency, just as value chain information-intensive manufacturers are building global IT-enabled value and production chains.[14] Such disaggregation is an important component of corporate strategy designed to take advantage of regional and local specialization and cost structures.

Photo credit: Elnur, Shutterstock

At the same time, securing the IT infrastructures involved is essential for every aspect of development, production, integration and delivery. Indeed, in these cases, the ability to provide effective cybersecurity is an essential enabling element of strategy. It can even be a competitive discriminator vis-à-vis competitors for which product quality, for example provenance and test data, and the integrity of information can be enhanced by cybersecurity.

The publication of Porter and Miller's work perhaps came too early for the application of the term "big data" used frequently today. Had the term been in vogue in the 1980s, Porter and Miller might have added information analysis value. This term describes the ability of today's analytic tools to aggregate many types of data from many sources—heterogeneous data in a homogeneous environment—to create decisions of significant value. Some examples are deciding which products to offer to specific consumers at specific prices and times, how to deploy valuable medical research and development resources, what crop futures the market might expect, or the likely progression of a dangerous epidemic. Tools applied from disciplines such as business intelligence, enterprise resource management and data mining considerably amplify the value of information.

Overall, it's no surprise that the rise in the importance of information—and the need to secure it—is followed closely by the attempts globally to steal intellectual property, to gain illegal access to information-as-product, and to enter value chains and achieve the ability to damage the information on which those chains rely.

ADVERSARIES' USE OF BIG DATA ANALYTICS

The development of large data warehouses and the creation of enterprise-level data repositories fundamentally has created information targets for our adversaries that are different than the information environment in the past. Information technology and today's modern computer networks that combine myriad streams of information in motion with enormous, complex, structured and unstructured referential databases or information at rest change the stakes enormously for both those defending today's information infrastructures and those seeking to exploit them. In the past, exploitation was subject to requirements for key facts or essential elements of information relating to military, diplomatic, political or economic plans and specific actions. In the past, those conducting exploitation wanted to understand force deployments, to gain access to orders to "attack at dawn," and to gain insight into a competitor's diplomatic or economic strategy.

But the environment has changed. Large-scale databases are being rebuilt as data warehouses that are designed explicitly for the use of big data analytics. A cogent explanation of this revolution is provided by Health Catalyst, which explains the need for the creation of today's large-scale information environments as follows:

> ***To effectively perform analytics, you need a data warehouse. A data warehouse is a database of a different kind: an OLAP [online analytical processing] database. A data warehouse exists as a layer on top of another database or databases (usually OLTP databases). The data warehouse takes the data from all these databases and creates a layer optimized for and dedicated to analytics.***[15]

Serving these information environments are data centers and architectures unimaginable in the recent past. Currently, the Lakeside Technology Center is the world's largest data center, described as a 1.1-million-square-foot multi-tenant data center hub owned by Digital Realty Trust … and today it is one of the world's largest carrier hotels and the nerve center for Chicago's commodity markets, housing data centers for financial firms attracted by the wealth of peering and connectivity providers among the 70 tenants.[16]

Such data warehouses and their associated infrastructures beg the question every sophisticated adversary must ask: Assuming we can gain access to them, what kind of big data analytics can we use to exploit these environments and what might we learn?

That large-scale data environments have become targets is indisputable. The breach of the Office of Personnel Management resulted in the compromise of some 21.5 million records.[17] A breach in 2015 of a voter database in the United States compromised information contained in some 191 million records,[18] while a compromise of Yahoo's user base may represent the largest data breach in history.[19]

All the breaches are interesting because of their sheer size, but they also represent data of significant variety. The Yahoo breach, for example, may have exposed financial information, as well as the affiliations and professional interests—and, in some cases, the professional work—of numerous prominent Yahoo members.

What might an adversary do with information environments this large, this varied and this rich? One might speculate that a foreign power could look for patterns in U.S. research and development, seek to understand important and valuable trends in the state of technology development for aerospace systems, as well as monitor clinical trials of new pharmaceuticals. Using big data analytics, other countries —and potentially cyber criminals—might attempt to understand and predict important economic developments. With such information in hand, they might even attempt to preempt and alter these developments in ways favorable to their interests and inimical to U.S. interests.

Even greater risks may lie ahead. Newer platforms, such as the one being developed by C3 IoT, hold the potential to monitor, analyze and predict the behavior of industrial and infrastructure-based information systems. According to C3 IoT, its new platform provides a complete IoT-platform-as-a-service solution that enables the rapid design, development, deployment and operation of enterprise IoT applications. With an elastic cloud architecture capable of handling data sets growing by hundreds of terabytes per day, the platform generates tens of millions of predictions every day for more than 20 full-scale production deployments worldwide.[20]

Such tools will allow operators to optimize resource allocation and performance, but they could give adversaries the means to do the same. The use of such platforms could help an adversary understand—and possibly alter—the use of resources in industrial and infrastructure systems in the United States and other countries.

Big data analytics adds substantial value to information. It amplifies the value of individual facts by allowing them to be integrated into large-scale data environments and to help find patterns useful to decision makers, whether those decision makers are military leaders, bankers, aerospace CEOs or leading researchers. At the same time, many of the big data analytics tools in use today are available internationally. Given access to big data environments, there is little that would impede the use of these tools by adversaries, and we should expect that this use is, in fact, taking place today.

RECOMMENDATIONS

USING BIG DATA TO SECURE COMPLEX NETWORKS

Predictive analysis is one of the keys to securing large complex networks. Predictive analysis relies on the collection of a large volume of data, the normalization of the data set, correlation of data to related data sets and security requirements to reduce false positives, and analytical tools and software. Every device, every sensor and every application on the network generates some sort of log, used generally for maintenance and troubleshooting. The same information plays a significant role in the security of networks. Big data analytics platforms have been designed to cope with the scope of data generated by the complex networks deployed today and will deploy in the future. We are challenged, however, in applying these platforms to detecting, preventing and generating courses of action regarding malicious activity.

The AFCEA Cyber Committee recommends increasing the emphasis on the kind of research being conducted by IBM, work that applies artificial intelligence to all data collected by modern security information and event management systems and includes all log data generated by network devices as well as endpoints deployed in the industrial infrastructures to be secured. Particular attention should be paid to the development of joint research efforts among companies such as IBM, which already is contending with this challenge, GE and C3IOT. These latter companies are pioneering advances in understanding the data generated throughout complex information technology/operational technology networks, including networks that may be managed via cloud infrastructures.

If possible, government research and development organizations, including the Defense Advanced Research Projects Agency, the Homeland Security Advanced Research Projects Agency, and the various National Laboratories of the Department of Energy, should undertake a joint effort to support the development of such technologies. The current Grid Modernization Lab Consortium, comprising DOE's National Laboratories,[21] might represent a starting point for gathering support for such an effort. The Department of Energy's Office

of Electricity Delivery and Energy Reliability[22] could also help define such an effort and provide a challenge to the department's National Laboratories and other national research and development centers.

The efforts underway at DAPRA (e.g., Big Mechanism) and supported by the NITRD should be tasked with helping to identify anomalous behavior in complex networks and to improve identification of the causality associated with cyber exploits and attacks.

Given the progress made by the private sector, the government should allow the private sector to help shape and conduct such research. The AFCEA Cyber Committee is prepared to support the government and its private sector and academic partners in the development of Terms of Reference to facilitate such research.

PROTECTING BIG DATA ENVIRONMENTS

The collection and storage of data and information presents another major challenge to the security of big data analytics environments. The OPM hack illustrated that a single adversary can move easily across the enterprise to steal information. While we consider the network and endpoint device security the first line of defense, big data environments are still vulnerable once an adversary is inside the network. It is prudent that all data must be encrypted during transmission, at rest and, if possible, during processing. An additional layer of security may be the virtualization of the analytical platform.

We also might ask what in our big data environments is not visible to adversaries? Existing technologies can hide the enterprise big data environments from everyone except for users with the appropriate credentials that can be authenticated by the servers' supporting given big data environments.

Such an approach should be combined with effective, enterprise-level identity governance. This method would enhance the security features used to protect a given big data environment with more rigorous control over the provision of access to such environments. The governance of identity across enterprises should be considered as well for critical infrastructure sectors, an effort the various Information Sharing and Analysis Centers might undertake with support from the Department of Homeland Security. The AFCEA Cyber Committee recommends the government departments and agencies, critical infrastructures owners and operators, and other stakeholders employ these approaches.

THWARTING ADVERSARY BIG DATA

Concern is rising that the United States is losing billions of dollars as well as terabytes of data in intellectual property each year. Thwarting adversary big data analytics is a significant challenge given the pervasive availability globally of big data analysis tools. We might consider the scenarios to which such tools might be applied, including the combination of exploited data. For example, an adversary might co-mingle OPM data with information taken from financial data breaches. Such an effort could provide adversaries a means to determine which government employees are in financial distress. The new Department of Defense Insider Threat Program report is attempting to forestall the exploitation of DoD staff members by adversaries seeking leverage against victims.

Several approaches are worth exploring. Digital markers—additional data that can be used to track data transmission and transactions—for sensitive data might be used to spot an adversary's use of big data analytics against exploited data. Such an approach depends, however, on adversaries conducting their big data analytics on devices connected to the Internet, which may not be the case.

Another approach might be to create mechanisms by which sensitive data self-destructs when it meets other sensitive data to which it is not ordinarily exposed. It should be noted that research in this regard appears nascent. Perhaps the best investment would be in understanding more clearly how adversaries intend to employ big data analytics in ways that affect U.S. national interests. The AFCEA Cyber Committee recommends the Intelligence Community look as carefully as possible at foreign countries' policies, doctrines and concepts of operations regarding:

- areas in which their national interests intersect those of the United States;
- approaches they are using or contemplate using to affect U.S. national interests;
- the kind of exploited information they would need to affect U.S. national interests using these approaches; and
- the kind of big data analytics tools they would use to derive meaningful intelligence to guide their actions.

The Intelligence Community could use the results of this inquiry to inform efforts by the Department of Homeland Security to emulate adversaries' big data efforts to understand the United States and affect its national interests and to help prepare the nation to detect, mitigate and overcome such efforts. The AFCEA Cyber Committee believes such intelligence could be of immense importance to national and homeland security, recommends such an approach and is prepared to support development of an initial Terms of Reference.

CONCLUSION

Big data analytics can help the United States manage government operations and critical infrastructures with a level of efficiency only glimpsed vaguely in the past, and the private sector is already developing the tools and models from which our national security can benefit. At the same time, these tools enhance the value of the information that adversaries may wish to steal, and these tools are available to our adversaries if they wish to employ them.

The AFCEA Cyber Committee urgently recommends that we apply big data analytics to the security of our government and critical infrastructures and undertake the research necessary to speed the development of these tools.

Given the value of big data environments, we also should look at ways to enhance cybersecurity at an enterprise level. In addition, urgent attention should be paid to the way other countries are using big data analytics to understand the United States and craft courses of action that affect our national interests. The AFCEA Cyber Committee stands ready to support these efforts.

NOTES

1 This paper uses the term “computer network” to denote integrated information systems that transmit, process and receive information in motion (transaction information) and information at rest (referential information). Waze is such a network, combining crowd-sourced transactional information relating to road conditions with referential geospatial data to produce recommended routing. Such systems are also characterized by common transmission and processing formats (e.g., TCP/IP), Internet Protocol (IP) addresses and a variety of interoperable data standards.

2 http://www.webopedia.com/TERM/B/big_data_analytics.html
Accessed April 23, 2017.

3 Ibid.

4 Ibid.

5 Tariq Mahmood and Uzma Afzal, 2nd National Conference on Information Assurance, 2013, Rawalpindi, Pakistan.

6 GE Estimates: https://www.ge .com/digital/press-releases/GE-Announces-Predix-Cloud-Worlds-First-Cloud-Service-Built-Industrial-Data-Analytics.

7 IDC report: December 2012 - THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East—https://www .emc .com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf.

8 https://www.darpa.mil/program/big-mechanism.

9 Department of Information Technology & Operations Management, Florida Atlantic University and Jim Quan, Department of Information Technology & Operations Management, Florida Atlantic University.

10 http://www.boeing.com/commercial/aeromagazine/articles/qtr_01_09/article_05_1.html.

11 Ibid.

12 “How Information Gives You Competitive Advantage,” Harvard Business Review 63, no. 2 (1985).

13 Ibid.

14 Sunil Mithas and Jonathan Whitaker, “Is the World Flat or Spiky? Information Intensity, Skills, and Global Service Disaggregation,” Information Systems Research 18 no.3 (2007), 237–259, available at http://terpconnect.umd.edu/~smithas/papers/mithaswhitaker2007isr.pdf.

15 https://www.healthcatalyst.com/database-vs-data-warehouse-a-comparative-review.

16 http://www.datacenterknowledge.com/special-report-the-worlds-largest-data-centers/worlds-largest-data-center-350-e-cermak/.

17 https://en.wikipedia.org/wiki/Office_of_Personnel_Management_data_breach.

18 http://uk.reuters.com/article/us-usa-voters-breach-idUKKBN0UB1E020151229

19 http://uk.businessinsider.com/yahoo-hack-by-state-sponsored-actor-biggest-of-all-time-2016-9?r=US&IR=T.

20 http://c3iot.com/products/c3-iot-platform/.

21 https://energy.gov/under-secretary-science-and-energy/grid-modernization-lab-consortium.

22 https://energy.gov/oe/services/technology-development/smart-grid.

The AFCEA International Cyber Committee White Paper Series

www.afcea.org/committees/cyber