IJRSET 2018 Volume 5, Special-Issue-1 Pages: 55-59

ISSN 2394-739X

International Journal for Research in Science Engineering and Technology

APPLICATIONS OF BIG DATA ANALYTICS - NETWORK SECURITY

[1] Dr. J. Thirumaran, [2] M. Karthik Prabu
[1, 2] RVS Kumaran Arts and Science College,
[1, 2] Ayyalur, Dindigul, 624 801.

ABSTRACT: The term Big Data refers to large-scale information management and analysis technologies that exceed the capability of traditional data processing technologies. Big data is surprising in several ways. As a technology, it does not fall into any obvious category, but instead spans several categories. The objective of the paper is to describe the promise and potential of big data analytics in security. The paper describes the nascent field of big data analytics in security, discusses the benefits, outlines an architectural framework and methodology, describes examples reported in the literature, briefly discusses the challenges, and offers conclusions.

Keywords: [Big data, Analytics, Hadoop, Security, Framework, Methodology]

1. INTRODUCTION

The term Big Data refers to large-scale information management and analysis technologies that exceed the capability of traditional data processing technologies. Big Data is differentiated from traditional technologies in three ways: the amount of data (volume), the rate of data generation and transmission (velocity), and the types of structured and unstructured data.Human beings now create 2.5 quintillion bytes of data per day. The rate of data creation has increased so much that 90% of the data in the world today has been created in the last two years alone. This acceleration in the production of information has created a need for new technologies to analyze massive data sets. This paper describes how the incorporation of Big Data is changing security analytics by providing new tools and opportunities for leveraging large quantities of structured and unstructured data.

2. BIG DATA ANALYTICS

The process of analyzing and mining Big Data – can produce operational and business knowledge at an unprecedented scale and specificity. The need to analyze and leverage trend data collected by businesses is one of the main drivers for Big Data analysis tools. The technological advances in storage, processing, and analysis of Big Data include (a) the rapidly decreasing cost of storage and CPU power in recent years; (b) the flexibility and cost-effectiveness of data centres and cloud computing for elastic computation and storage; and (c) the development of new frameworks such as Hadoop, which allow users to take advantage of these distributed computing systems storing large quantities of data through flexible parallel processing. These advances have created several

differences between traditional analytics and Big Data analytics.

3. BIG DATA ANALYTICS FOR SECURITY

This part explains how Big Data is changing the analytics landscape. In particular, Big Data analytics can be leveraged to improve information security and situational awareness. For example, Big Data analytics can be employed to analyze financial transactions, log files, and network traffic to identify anomalies and suspicious activities, and to correlate multiple sources of information into a coherent view. Data-driven information security dates back to bank fraud detection and anomaly-based intrusion detection systems. Fraud detection is one of the most visible uses for Big Data analytics. Credit card companies have conducted fraud detection for decades. However, the custom-built infrastructure to mine Big Data for fraud detection was not economical to adapt for other fraud detection uses. Off-the-shelf Big Data tools and techniques are now bringing attention to analytics for fraud detection in healthcare, insurance, and other fields.

4. NETWORK SECURITY

In a recently published case study, Zions Bancorporation8 announced that it is using Hadoop clusters and business intelligence tools to parse more data more quickly than with traditional SIEM tools. In their experience, the quantity of data and the frequency analysis of events are too much for traditional SIEMs to handle alone. In their traditional systems, searching among a month's load of data could take between 20 minutes and an hour. In their new Hadoop system running queries with Hive, they get the same results in about one minute. The security data warehouse driving this implementation not only enables users to mine meaningful security information from sources such as firewalls and security devices, but also from website traffic, business processes and other day-to-day transactions.10 This incorporation of unstructured data and multiple disparate data sets into a single analytical framework is one of the main promises of Big Data.

5. BIG DATA ANALYTICS WITH MACHINE LEARNING

Machine learning in isolation doesn't mean that employing big data analytics to IT security operations will be successful, though. To provide significant change towards improved security requires layering the big data gathering with human intelligence; analysing and feeding back into the system on threats, alerts and other anomalous activity, teaching the system what is a threat and what isn't. The human intelligence input helps machine learning systems to learn, and very quickly you have a system in place that understands what a false positive is and what is a genuine, real threat that needs to be auctioned. As time goes by, the system continues to adapt. Using big data analytics with machine learning in combination with human intelligence provides a new self-learning solution to the problem of sophisticated attacks and advanced threats. Polymorphic malware that would normally evade signature-based security technologies can be detected and stopped with a combination of advanced analytics, machine learning and human expertise.

6. NETWORKING TRENDS IMPACTING SECURITY

By 2020, there will be more than 4 billion global internet users, 26 billion networked devices and connections, and global IP traffic will grow three-fold, reaching 2 Zetabytes, so states Cisco's VNI Global IP Traffic Forecast report. The data volumes we're seeing are growing exponentially. This is in part being driven by the internet of things (IoT), with the numbers of connected devices such as smart sensors rising towards 50 billion by 2020. The other big trend to affect how we operate online, in addition to the IoT, is the increasing popularity of software defined networking (SDN) and network function virtualisation (NVF). More and more companies are taking advantage of the benefits of replacing

individual routers, firewalls and switches with virtual machines. While the move brings benefits for dynamically provisioning network services and streamlining operations, the switch to using virtual images that interact with each other for routing, firewalls or session border controllers, rather than individual appliances, may also increase the security risks to the network from a single compromised device. Making changes to a corporate network to allow for SDN and NFV, without addressing and changing security will leave an organisation vulnerable to an attack.

7. PROTECTING NETWORKS WITH BIG DATA ANALYTICS

Software-Defined Networking (SDN)-based controllers and Big Data analytics within and about the data network itself are tools designed to provide a comprehensive overview of each and every network, which allows network administrators to detect more threats when compared with the capabilities of threat detection from a single access point. For example, many hospitals use behaviour analysis software to prevent the misuse of patients' personal information by using software that detects abnormal network behaviour to identify employees who may be leaking patient information. Big Data analytics enables network administrators to predefine policies and actions of the controller to reduce maintenance workload and ensure secure network operations. Preset rules can ensure that suspicious traffic is imported to the security centre and eliminated, as appropriate.

Another benefit of Big Data analytics is its ability to process large amounts of data quickly to generate real-time results. It analyzes network security attacks and potential risks immediately, which prevents security breaches.

8. BIG DATA ANALYTICS IN CYBER SECURITY

At the core of this approach stands improved detection – and that is where big data analytics comes into play. Detection must be able to identify changing use patterns; to execute complex analysis rapidly, close to real time; to perform complex correlations across a variety of data sources ranging from server and application logs to network events and user activities. This requires both advanced analytics beyond simple rule-based approaches and the ability to run analysis on large amounts of current and historical data – big data security analytics. Combining the current state of analytics with security helps organizations improve their cyber resilience.

9. BIG DATA SECURITY ANALYTICS: A NEW GENERATION OF SECURITY TOOLS

As the security industry's response to these challenges, a new generation of security analytics solutions has emerged in recent years, which are able to collect, store and analyze huge amounts of security data across the whole enterprise in real time. Enhanced by additional context data and external threat intelligence, this data is then analyzed using various correlation algorithms to detect anomalies and thus identify possible malicious activities. Unlike traditional SIEM solutions, such tools operate in near real time and generate a small number of security alerts ranked by severity according to a risk model. These alerts are enriched with additional forensic details and are able to greatly simplify a security analyst's job and enable quick detection and mitigation of cyber attacks.

10. Key features distinguish big data security analytics from other information security domains

10. 1: SCALABILITY:

One of the key distinguishing features of big data analytics is scalability. These platforms must have the ability to collect data in real or near real time. Network traffic is a continual stream of packets that must be analyzed as fast as they are a captured. The analysis tools cannot depend on a lull in network traffic to catch up on a backlog of packets to be analyzed.

It is important to understand that big data security analytics is not just examining packets in a stateless manner or performing deep packet analysis. Although these are important and necessary, it is the ability to correlate events across time and space that is a key differentiator of big data analytics platforms. This means the stream of events logged by one device, such as a Web server, may be highly significant with respect to events on an end-user device a short time later.

10. 2: REPORTING ANDVISUALIZATION:

Another essential function of big data analytics is reporting and support for analysis. Security professionals have long had reporting tools to support operations and compliance reporting. They have also had access to dashboards with preconfigured security indicators to provide high-level overviews of key performance measures. Once again, both of these existing tools are necessary but not sufficient to meet the demands of big data.

10. 3: PERSISTENT BIG DATA STORAGE:

Big data security analytics gets its name because the storage and analysis capabilities of these platforms distinguish them from other security tools. These platforms employ big data storage systems, such as the Hadoop Distributed File System (HDFS) and longer latency archival storage. Back-end processing, meanwhile, may be done with MapReduce, a well-established computational model for batch processing. While MapReduce is highly resistant to failure, it is at the cost of I/O-intensive processing. A popular alternative to MapReduce is Apache Spark, a more generalized processing model that utilizes memory more effectively than MapReduce.

10. 4: INFORMATION CONTEXT:

Since security events generate so much data, there is a risk of overwhelming analysts and other infosec professionals and limiting their ability to discern key events. Useful big data security analytics tools frame data in the context of users, devices and events.

Data without this kind of context is far less useful, and can lead to higher than necessary false positives. Contextual information also improves the quality of behavioral analysis and anomaly detection. Contextual information can include relatively static information, such as the fact that a particular employee works in a specific department. It also includes more dynamic information, such as typical usage patterns that may change over time. For example, it may not be unusual to have a large volume of queries on a data warehouse on Monday mornings, as managers run ad-hoc queries to better understand events described in their weekly reports.

10. 5: BREADTH OF FUNCTIONS:

The final distinguishing characteristic of big data security analytics is the breadth of functional security areas it spans. Of course, big data analytics will collect data from endpoint devices; that is any device that is connected to a TCP or IP network via the Internet. This includes anything from laptops and smartphones to Internet of Things devices. In addition to physical devices and virtual servers, big data security analytics must attend to software-related security. For example, vulnerability assessments are used to determine any possible security weak points in the given environment. The network is a rich source of information and standards, such as the Cisco-developed NetFlow network protocol, which may be used to gather information about traffic on a network.

CONCLUSION

Big data delivers an interactive analytical capability that can be quickly implemented to access and analyze huge amounts of data, up to trillions of rows. It has a broad set of analytical capabilities and consequently it has broad areas of application, beyond the normal range of most analytical BI products. We recommend that companies looking for an analytical platform, especially those that gather, accumulate and analyze large amounts of data, consider it as an option.

The goal of Big Data analytics for security is to obtain actionable intelligence in real time. Although Big Data analytics have significant promise, there are a number of challenges that must be overcome to realize its true potential. The following are only some of the questions that need to be addressed:

1. Data provenance: authenticity and integrity of data used for analytics. As Big Data expands the sources of data it can use, the trustworthiness of each data source needs to be verified and the inclusion of ideas such as adversarial machine learning must be explored in order to identify maliciously inserted data.
2. Privacy: we need regulatory incentives and technical mechanisms to minimize the amount of inferences that Big Data users can make. CSA has a group dedicated to privacy in Big Data and has liaisons with NIST's Big Data working group on security and privacy. We plan to produce new guidelines and white papers exploring the technical means and the best principles for minimizing privacy invasions arising from Big Data analytics.
3. Securing Big Data stores: this document focused on using Big Data for security, but the other side of the coin is the security of Big Data. CSA has produced documents on security in Cloud Computing and also has working groups focusing on identifying the best practices for securing Big Data.
4. Human-computer interaction: Big Data might facilitate the analysis of diverse sources of data, but a human analyst still has to interpret any result. Compared to the technical mechanisms developed for efficient computation and storage, the human-computer interaction with Big Data has received less attention and this is an area that needs to grow. A good first step in this direction is the use of visualization tools to help analysts understand the data of their systems.

REFERENCES

[1]. Alperovitch, D. (2011). Revealed: Operation Shady RAT. Santa Clara, CA: McAfee.

[2]. Bilge, L. & T. Dumitras. (2012, October) Before We Knew It: An empirical study of zero-day attacks in the real world. Paper presented at the ACM Conference on Computer and Communications Security (CCS), Raleigh, NC.

[3]. Bryant, R., R. Katz & E. Lazowska. (2008). Big-Data Computing: Creating revolutionary breakthroughs in commerce, science and society. Washington, DC: Computing Community Consortium.

[4]. Camp, J. (2009). Data for Cybersecurity Research: Process and "whish list". Retrieved July 15, 2013, from http://www.gtisc.gatech.edu/files_nsf10/data-wishlist.pdf.

[5]. Cugoala, G. & Margara, A. (2012). Processing Flows of Information: From Data Stream to Complex Event Processing. ACM Computing Surveys 44, no. 3:15.

[6]. Curry, S. et al. (2011). RSA Security Brief: Mobilizing intelligent security operations for Advanced Persistent Threats. Retrieved July 15, 2013, from http://www.rsa.com/innovation/docs/11313_APT_BRF_0211.pdf

[7]. Dumitras, T. & P. Efsathopoulos. (2012, May). The Provenance of WINE. Paper presented at the European Dependable Computing Conference (EDCC), Sibiu, Romania.

[8]. Dumitras, T. & D. Shou. (2011, April). Toward a Standard Benchmark for Computer Security Research: The Worldwide Intelligence Network Enviornment (WINE). Paper presented at the EuroSys BADGERS Workshop, Salzburg, Austria.