IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

Review Paper on Security Intelligence with Big Data Analytics

Chiquita Prabhu[1], Omkar Neogi[2], Kriti Shrivastava[3], Neha Katre[4]

Dwarkadas J. Sanghvi College of Engineering, Mumbai, India[1, 2]

Assistant Professor, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India[3, 4]

Abstract: Unfortunately no one is immune to security threats that are exponentially increasing in cost, impact, extent and complexity. Though there are traditional security approaches, they aren't competent enough due to abundant information and lack of tools that don't allow them to gain insight into the information to obtain knowledge about unknown threats. To resolve this problem as well as detect weak signals of threats that hide behind the noise of huge data in an organization, we have security intelligence platform that is a Big Data solution. This aim of this paper is to summarize why Big Data analytics is used for security intelligence, the ideal requirements of a platform designed for security purpose, how Big Data analytics is helpful compared to traditional approaches and the study of various platforms developed for security intelligence using Big Data analytics. Lastly we have a comparative study between the two most popular security intelligence platforms and then we discuss about how Big Data could be a dominant name in the field of security if it overcomes certain challenges.

Keywords: Anomaly, Beehive, Big Data, Big Data analytics, Intelligence driven security, IBM QRadar.

I. INTRODUCTION

The marriage between technological advances and information society conceived the notion of data. As time passed by and access to technological miracles became easier and economical, the amount of data produced increased and from then onwards the influx of data is still increasing to the point that today we live in the petabyte era. This huge amount of data being generated and stacked was later called the Big data. The Big data analytics is nothing but tools required to work within big data.

As the amount of data keeps increasing, concerns regarding its security keep knocking at the user's door. Security can be defined as protection of assets. Security as a paradigm is not just restricted to data. The idea of security was born long before we can trace its life time. Security transformed from the cave men using Guard dogs; to protect their belongings which later transformed to moats around castles, and then we had security guards for individual houses, and later home security systems, and finally cryptography which is used to protect data.

Big data is popular due to its multifarious facets like being prescriptive, diagnostic, predictive and descriptive and hence could be a perfect candidate to fight for data security.

The following two data breaches shook the world and caused people to even lose trust in data security functionalities in place at the highest level of organizations.

Number 1: The WikiLeaks leaks on NSA spying: It was revealed to the world that the US government agency NSA spied on everything from phone calls to emails and even tapped the optical fiber lines passing through the American borders. This led to a massive outcry amongst internet users and raised awareness about the importance of protecting data.

Number 2: The successful attacks on Apple's personal storage cloud iCloud on the 31st of October, 2014 defaced the organization. This brute force and phishing based attack led to users losing faith in the security of cloud based services. A few months later, another exploit called the iDict was made available to the world through GitHub, which was a dictionary based attack and could be used to hack into iCloud accounts. The maker claimed that this attack was made available to the world just to display an obvious vulnerability of the security of iCloud and that he/she had not actually tested it on any accounts. But whether any precious data was stolen or not cannot be ascertained.

Such attacks have only exposed the fact that if security at some of the biggest organizations like Apple can be overcome so easily, it can't be much more difficult for attackers to target data at smaller organizations.

The problem with attack response is that there is a huge time difference between the attack starting to spread and it being detected. By the time it is detected, it is already too late. It is difficult to define the exact set of characteristics which characterize an attack that can be used for detection. Till date we haven't been able to trace the course of lifetime of a small bug hiding incognito in a program evading every test, thus causing a loophole in the program concerned and hence encouraging attackers to develop vulnerability patches of anti-virus signatures which would result in automatically generated exploits based on the patch and then giving rise to the competition between attacks and the remediation measures introduced by the security community.

To understand this pattern or life cycle we need an in depth knowledge of data accumulated from various sources, collected independently using sensors.

DOI 10.17148/IJARCCE.2015.411100

This is where Big data analytics comes into the picture.

Big data analytics can be defined as the management of information at a large scale that surpasses traditional data processing techniques.

Big data can be distinguished from the conventional set of data using three parameters: The huge volume of data that it has, the rapid rate of data generation it shows and the variety of data that it encompasses i.e. structured and unstructured formats.

As the amount of data is increasing by the day it is creating a tidal wave of data that contains an amalgamation of redundant as well as sensitive information. However the advancement in big data analytics has opened the doors to access sensitive data and violate privacy. Hence it is necessary to safeguard chunks of sensitive data against abuse.

Big data analysis can metamorphose security analytics in the following ways:

(a) It can accumulate data from various internal organizational sources as well as external sources to make a consolidated view of the required data into something called as a vulnerability database.

(b) It helps perform an in-depth analytics on the data using security intelligence hence uncovering unique patterns that could be the source of many security issues. (Anomaly detection)

(c) It provides a one dimensional view of all the related information.

(d) It provides real time analysis of streaming data and uses previous results as a feedback to the system as a whole. [5]

To list a few Threats that can be countered using big data; Fraud detection which could include credit card frauds, and identity frauds, insider trading attacks, targeted intellectual property attack, zero day attacks; these are some of the trending attacks and hence new Big data tools are now being developed to manage security issues as they are efficient in cleaning, maintaining and querying data in discontinuous, incomplete and noise affected formats; This is done by developing situation based intelligence in the form of log records.

II. WHY BIG DATA ANALYTICS FOR SECURITY

The present day security data sets are incompetent for revealing answers to many challenging problems. To obtain a proper judgment on security matters we need to gain more value from the already accumulated and analysed data and this requires better interpretation of the current as well as impending complexities related to the data.

The revolutionary changes in the types of security attacks can now be tamed using big data and big data analytics. Big data analytics can be used to study transactions, log files as well as network traffic to recognize anomalies and suspicious activities as well as present a one dimensional view of the combined data set.[4]

One of the main reasons why big data can be used with security is that an organization needs to maintain its traditional security information for a longer period of time to analyse the data. Historical study of this data has the perspective of uncovering unacknowledged security vulnerabilities and identify breaches in security over a period of time. This is possible using big data analytics as the storage cost has decreased tremendously and while the traditional data warehouses retain data for a specific period of time, big data can go about maintaining a piece of data indefinitely. Hence allowing us to find patterns in data so as to unearth possible security attacks. [4],[3]

Secondly, Data sources not conventionally adopted for security can help an organization judge assets and elements that need to be protected and observed. Traditional security tools mostly deal with structured data as they are rigid by nature but big data analytics deal with both structured and unstructured data eliminating the need of any predefined data formats and hence we can carry out analysis on information observed in sources like email, social media, business documents and web content. [4],[1]

Thirdly, various analytical operations need to be performed to unravel security related insights from large data sets and this will require more processing time. Big data tools like Hadoop and NoSQL databases have increased the processing speed of complicated queries and unique pattern identification within the large blocks of information. Asynchronous analysis needs to be performed in real time and once this analysis is complete, these results are fed as a feedback to the real time element to deliver an optimal solution over time. [4],[3],[1]

The sophistication and visibility required to identify and protect systems against cyber attacks are absent in traditional security technologies. Due to this they are only able to solve a part of the problem as the traditional tools lack the intelligence of perception that the attackers use to skirt through those defenses and camouflage with the background noise of an organization’s workflow. The intelligent security approach with big data and analytics could help build up a system that fills the gaps left behind by the traditional systems.

III. BASIC MODEL FOR INTELLIGENCE DRIVEN SECURITY

To extract knowledge from tons of data gathered, efficiently manage threats, to apply security intelligence to drive decision making; companies need to take big data approach for security management.

To take care of all the fields mentioned above the Big data driven security system should essentially provide for:

1) A scale out infrastructure that will respond to dynamic advances in the technological environment and evolving data risks. The infrastructure should be able to provide both physical and economical support to the stacking amount of data in an organization; retaining both current and historical records. The infrastructure needs to be flexible to entertain new environments as well as unique threats.[12],[8]

2) The system should contain analytical and visual tools that support event identification with the help of previously available information. These tools should be

DOI 10.17148/IJARCCE.2015.411100

able to reconstruct suspect files and provide for automated testing of those files. It should also enable full reconstruction of log as well as network information about a session to determine the trail of suspicious activity.[12],[4],[8]

3) The system should provide for threat intelligence that would compare the consolidated data with external environment to look for similarity in anomalous activities to detect threat indicators and appraise organizations about what to look for.[12]

4) The system should be able to reduce repetitive tasks related to investigating a particular threat i.e. conducting the same test using three different tools and hence reducing the number of steps involved overall.[12]

5) The system should be able to map the monitored data with critical issues handled by some third party and hence be able to determine sources of threats; this can be done by assessing behaviour and risk models and not just threat signatures.[12]

6) The system needs to eliminate results with false positive and noise and only provide information on data that are the biggest risk factors in the enterprise and the reason why it is so.[12]

7) The system should emulate a predictive model hence providing early warning against attacks therefore transforming the entering security system from a passive defence model to an active defence and prevention model.[12]

8) When high risk activity is detected the system should enable additional user authentication and blocking of data transmissions.[12]

9) The system should allow security analysts to access a centralized repository of security related information hence updating them about threat patterns.[12],[8],[4]

IV. HOW IS BIG DATA ANALYTICS HELPFUL?

With international headlines being inundated with major security breaches enterprises are turning towards security intelligence with big data analytics approach to address burgeoning problems of fraud, insider attacks and advanced persistent threats, because this approach provides with:

A) Base Line establishment:
Here the organization understands its information ecosystem, becomes aware of what needs to be defended, which enables it to make decisions about formulation of risk profiles and addressing the anomaly detection. It also provides answers to questions like who are the specific targets within an organization, what applications and information need strict attention due to presence of sensitive information and what is the difference between normal and abnormal behavior within the ecosystem.[12],[13]

B) Identify advanced persistent threats:
Organizations gain knowledge on the trail of security breach set up by attackers, incognito within a system which evade most of the testing phases. The organizations get a fair idea on which application have already been compromised or are susceptible to attacks, which particular source (external or internal) may be the inception of attack and helps identify low profile network components that might be the reason behind certain security attacks.[12],[13],[5]

C) Limit Insider Threats:
Enterprises either gain evidence or are warned against employees of the same organization who are capable of stealing intellectual information, compromising organizations system or perform actions that might be injurious to the entire organization. They get notified about what data is being lost and by whom, who among their employees is technically capable of attacking the system and who is showing abnormal behavior during usage of system data.[12],[13],[5]

D) Foresee System Hacking:
Security intelligence with Big data help organizations identify attacks from groups or entities that find loopholes within the system and act on it hence propagating negative image about the organization. In addition to this it helps monitor agnostic intentions of entities towards business practices as well as enterprise’s actions that might trigger increased impact of risk of attack.[13],[12]

E) Defence against cyber attacks:
The organization is alerted about an impending or currently occurring attack by criminal enterprises or government funded colonies. This way the organization gains information about the origin of the attack, the hacking weapons used and person using them, symptoms of attacks if any; so that these symptoms can be used for predicting future attacks.[12]

F) Attenuate fraud:
The enterprise is appraised of new and existing fraud systems which could challenge the company’s regulations and cause significant loss of sensitive data. Big data gives knowledge about how to identify fraudulent activity, compromised regulations that could lead to fraudulent activity, and if fraudulent activities have any specific pattern that can be identified or anticipated.[12]

V. SECURITY INTELLIGENCE TECHNOLOGIES

To stand tall against all cybercrimes we need effective threat detection and investigation services. An ideal security team requires a system that is flexible enough to garner data from various security specific domains which can unleash potential security threats and risk and alert analysts in seconds, hence trying to truncate the problem in the nib. Certain models that provide extensive features as an integrated security system are as follows:-

A. The RSA Security Analytics:
RSA security analytics is security software developed by EMC2; it has transformed security management from conventional log-centric approach to one with exceptional workflow ethics and hence stands tall against various other solutions, the RSA security analytics solution basically provides with:

1) Comprehensive infrastructure visibility:
This has the capacity to collect all of security related data at large and from various sources.

DOI 10.17148/IJARCCE.2015.411100

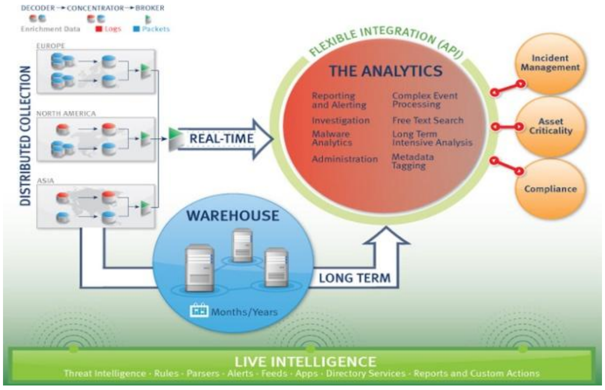

Fig1.Working of RSA Security Analytics.[15]

This allows analysts to access data concerned with security, threats and malicious activities in a single consolidated form. The architecture also provides real time analytics as well as historical data query. Therefore this architecture allows real time threat altering, investigation analysis along with historical retention and data archiving.[1],[3],[5]

2) Agile analytics:
This security platform provides rapid investigation tools to analysts along with intuitive tools presented for rapid analysis that helps in detailed study of a data set hence allowing the augmentation of decision making. This uses signature free analytics to detect malicious users and activities depending on the end users connected to the infrastructure. This also has provisions to replay and recreate situations; exactly how they happened.[3],[5]

3) Actionable Intelligence:
The threat intelligence provided by RSA helps analysts obtain optimal solutions by implementing current feeds of threat information. The intelligent system built into the architecture is bound by rules, reports and watch lists. This allows analysts to gain insights into a problem and accordingly prioritize response or counter measures.[3],[5]

4) Optimized Incident Management:
The platform includes functionalities which enable security teams to access and monitor threats by streamlining the activities related to preparedness and response through workflow systems to clearly define and activate responses if threats are realized and tools to access and monitor the currently open processes. This, along with third party tools and the RSA platform is an industry leading service to allow security teams to manage and respond to security incidents. In turn, the responses to incidents improve and the learning derived from each experience of responding to incidents increases, leading to improved reactions in the future.[1]

B. The Beehive Model
Attacks against intellectual property or physical system of high value are called Advanced Persistent Threats (APT). Unlike mass spreading malwares, such as Trojans, viruses and worms APT attackers are low and slow. Low mode means maintaining a low profile in networks and slow mode means persistently existing in a system for a long period of time. [2],[4] APT is one of the most serious technological threats in the information environment, this is mostly done by attackers to obtain financial gain, to embarrass or blackmail customers, to poison data or carry out illegal insider trading and destroying an organization’s reputation. APT by nature is sophisticated and diverse and are carried out by highly skilled hacktivists and Big data analysis approach can very well manage such threats.[2],[4]

Beehive is a behaviour profiling technology for APT developed at RSA labs. RSA labs have concluded that an attacker’s behaviour is never subtle; when the attacker plans to steal sensitive data he compromises his actions and deviates from the norm. APT attacks are multi staged attacks as it includes exploitation, objectives, command and control and lateral movement, each activity of the attacker provides an opportunity to identify the changing course of the usage behaviour in a system and hence consolidating such independent activities unearth proofs of stealthy attacks and intrusion.[2],[4]

Anomaly sensors are used in Beehive to identify behavioural deviations. Each sensor is responsible for tracking a particular activity concerned with the user in the enterprise network. For example a sensor might keep track of the sites that a user visits hence understand any unusual connections the user might have with a third party or it might monitor anomalous access patterns by keeping a count of machines a user logs into, the sensor might also keep an eye on the regular working hours of a user to flag usual as well as unusual activity and look for illegal sinks of data transmission between the internal and external sites.[2],[4]

If any one of the sensor is triggered it shows the presence of a single unusual activity whereas if many sensors are triggered at once it shows the presence of multiple unusual activities. This helps analyst produce behavioral reports according to attacker’s working ethics and hence develops an investigation and response plan.[2],[4]

The name beehive comes from the numerous sensitive elements (sensors) that work together to achieve a common goal emulating the real world bee scenario as each bee plays a different role hence cooperating with each other to maintain the hive.[4]

C.IBM QRadar Security Intelligence Platform.
IBM QRadar is a solution developed by the company IBM itself. It is a technology that has been built using Big data capabilities to align itself with the evolving threats and restrict any attacks before they can happen. It is responsible for uncovering hidden connections within massive amounts of security information with the help of analytics that reduce security events to a set of prioritized incidents.

This security intelligence platform is designed from ground up that could deliver benefits of next generation security systems by expanding visibility into network and user application activity which would help provide actionable intelligence to security breaches within an enterprise.[6],[7]

Features that make QRadar an ideal approach to combat advanced threats:

DOI 10.17148/IJARCCE.2015.411100

A. Interoperability and Scalability.
It provides for unified architecture which collects, stores, analyses, queries logs, threats, vulnerabilities and risk-related data. It also has a purpose built database that provides scalability and tunes performance by allowing a scan through millions of data in seconds. [6],[7],[10]

B. Pre- and Post exploit insights.
This gathers information and prioritizes it; it also helps detect suspicious behavior if any, within the network of the organization to recognize a breach. [6],[7],[10]

C. Anomaly detection capabilities.
This solution creates a baseline plan of activities currently happening and identifies any deviation from the norm by comparison of data. It can define which deviations are meaningful so a sure shot breach is found. [6],[10]

D. Real-Time correlation and analysis.
This solution maps massive data sets in real time which allows for accurate attack detection.

E. Reduced false positives.
QRadar quickly detects security breaches and de-prioritizes innocuous events which reduce the time analysts spend on investigating real compromises. [6],[7],[10]

F. Forensic capabilities.
To ease the burden off the security and network staff so that they can rapidly access the source and impact of a breach the QRadar provides for a console view of the entire enterprise data. [7],[10]

G. Flexibility.
Here you can easily add new data sources, tune and create analytics, create new reports and user views which provides for an effective approach to defend against advanced attackers. [6]

VI. COMPARATIVE STUDY

As we have studied some security intelligence model, they look quite the same to the naked eye but when we delve into the specifics we can actually see the parameters that differentiate the RSA security analytics and IBM QRadar.

TABLE I

Parameters	**RSA security analytics**	**IBM QRadar**
Scalability	Linearly scalable	Scalable in all dimensions
Text Analysis	Free analysis provided	Easy analysis provided
Anomaly detection	Capable on detecting advanced persistent threats, fraud, insider threats as well as unknown threats	Capable of detecting advanced persistent threats, fraud and insider threats
False Positive	May be present	Not present
Flexibility	flexible	Flexible
Added Features	No added features present.	Contains spread sheet tools to allow users explore and collect data.
User Interface	High visual browser based interface	Graphical Front end tool for visualizing and exploring data.
Efficiency	Very efficient as mobility from one security tool to another is not required.	Not very efficient as two platforms are required for accurate functioning.
Querying speed	High	High
Forensics	Comprehensive view of data	Console view of data
Parameters	**RSA security analytics**	**IBM QRadar**
Malware analytics	Provides for four distinct malware investigation techniques	Provides for customized malware tactics.
Interface Type	Open	Closed
Architecture	Distributed computing architecture	Unified architecture
Archiving	Supports large data archiving.	Deletes data after a specified period of time.
Sources	Network logs, SIEM, System alters and application records.	DNS transactions, emails, documents, social media data.
Incident Investigation	Present with auto generated remedy	Not present.

VII. CONCLUSION

To obtain actionable intelligence in real time is the sole aim of big data analytics. But still Big data analytics has to overcome a number of challenges to show its true potential and stand true to its significant promise that it holds in the field of security. Certain challenges that must be addressed to make Big Data analytics a prominent name in security are:

(1) Data Provenance: Big Data is an accumulation of information from a variety of sources and this information must be verified for its authenticity and integrity before any analytics is applied on it, because sometimes certain data are inserted in malicious ways.

(2) Privacy: In order to provide high level of data security Big Data analytics can encroach upon user privacy and hence proper guidelines must be laid down to provide an ethical yet efficient security system with Big Data analytics.
(3) Securing Big Data Stores: Using Big Data with security is just one part of the coin as the bigger issue is to deal with, is security for big data, it is important to secure the massive data collection.
(4) Human Computer Interaction: Diverse sources of data are analyzed by Big Data, but still a human analyst is required to interpret the results. Though the advancements in the field of technology, for storage and computing has shown a tremendous growth, the field of human interaction with the computer still needs to be attended at. Therefore proper visualization tools must be provided to help analysts understand the expanse and patterns of data.
Once these challenges a looked into, we can be positive that these changes will bridge the gap between security intelligence and Big Data Analytics.

REFERENCES

[1] Alvaro A. Cardenas, Pratyusha K. Manadhata and Sreeranga P. Rajan, “Big Data Analytics for Security,”in IEEE Security and Privacy magazine, Feb 2015
[2] Ting-Fang Yen, Alina Oprea, Kaan Onarlioglu, Todd Leetham, William Robertson, Ari Juels and Engin Kirda,” Beehive: Large-Scale Log Analysis for Detecting Suspicious Activity in Enterprise Networks”.
[3] A case study in Big Data Analytics http://darkreading.com
[4] R Cloud Security Alliance Big data Analytics for Security Intelligence. http://cloudsecurityalliance.org/research/bigdata
[5] Sam Curry, Engin Kirda, Addie Shwartz, William H. Stwart and Amit Yoran, “Big Data Fuels Intelligence- DrivenSecurity,”Jan 2013
[6] IBM Security Intelligence with Big Data http://www-03.ibm.com/security/solution/intelligence-big-data/
[7] Security Intelligence and analytics http://www-03.ibm.com/software/products/en/category/security-intelligence
[8] The use case for big data and security analytics: An interview with Ben Wuest http://securityintelligence.com/the-use-case-for-big-data-and-security- analytics-an-interview-with-ben-wuest
[9] Four Types Of Big Data Analytics And Example Of Their Use http://www.ingrammicroadvisor.com/data-center/four-types-of-big-data-analytics-and-examples-of-their-use
[10] IBM Big Data and information management http://www-01.ibm.com/software/data/bigdata/enterprise.html
[11] IBM,” Extending Security Intelligence with Big Data solutions” Jan 2013.
[12] RSA Security Analytics https://www.emc.com/collateral/data-sheet/security-analytics-overview- ds.pdf
[13] Vijay Dheap, “What You Need to Know About Security Intelligence with Big Data,” July 2013.
[14] Types of attacks on data https://technet.microsoft.com/en-us/library/cc959354.aspx
[15] Working of RSA Security Analytics. http://www.emc.com/about/news/press/2013/20130130-01.htm

DOI 10.17148/IJARCCE.2015.411100