Cyber Security and Big Data Analytics

Priya Bhardwaj*
Nidhi**

Abstract

This research report examines the differences between the traditional methods of cyber security with the Big Data analytics. Since Internet is growing at a fast pace, the need for cyber security measure is essential in almost every aspect of our lives. This growing network is becoming more of a threat these days than a blessing, because today many people are using it as a safe and secure passage for committing high-profile crimes. This paper tells us the importance of Big Data in the field of cyber security.

Keywords: Cyber Security, Big Data Analytics, MapR, Hadoop

I. Introduction

Cybercrime (computer crime) can be described as a term for any felonious activity by a person whose primary means of communication is a computer. Today more research is needed to secure the cyber from these unauthorized people. Cyberspace has become today's new battlefield and cyber security continues to be a top priority for every sector.

Cyberspace can be distributed into the following resources: devices, information, networks, and people. Securing these resources is the job of cyber security. Cyber security has transformed very quickly from the technical area to the scientific notion. Globalization and Internet has given organizations, individuals and nation enormous power due to the advancing network technologies. For everyone – students, teachers, entrepreneurs, socialists, soldiers, pioneers, hackers, and even terrorists – collecting and sharing information, communication, everything has been digitized. As a result, all the sectors now have a cyber-drawback, the varsity and influence of which are hard to foretell, and the struggle facing in cyberspace are becoming more crucial than attacks taking place on the ground.

The nature of a civil security threat has not changed, but the Internet has provided a new delivery process that can escalate the speed, scale, and power of an attack. Every sector is introducing better security options to protect their systems and prevent data loss.

The year 2014 which is also known as the "Year of the Data Breach" has recorded an average cost per breach reaching $12.7 million. This has reminded us the need to use better protection and secure gateways to help fortify systems, statistics and individual. Today, the conversation regarding these crimes are less about intercepting these attacks but more about how momentarily you can expose that an attack is happening.

II. Categories of Cyber Crimes

Cybercrimes can be broadly classified into three sectors:

1. Property
2. Government
3. Individual

Each class uses a various techniques and each technique used vary from criminals to criminals.

A. Property

Just as in real word criminals can loot and mug, even in the cyber world they steal and rob. These include intellectual property crime, time theft, credit card fraud, transmission of viruses and unauthorized access to computer systems.

B. Government

The growth of Internet has broadened the horizons for criminals to use cyber space as a medium to terrorize civilians and threaten the government. Internet terrorism is a common example of this category. In this, the criminals attack government websites. The criminals here can be terrorist or unfriendly governments of other nation.

Priya Bhardwaj*
Department of IT
Institute of Information Technology and
Management, GGSIPU, Delhi, India

Nidhi**
Department of Computer Science
University of Delhi, Delhi, India

C. Individual

This type is in the form of cyber stalking, email spoofing, spamming, trafficking and cyber stalking. Today, the governments are more concerned about this category and are seeking help from other nations as well in order to arrest the perpetrators.

III. Reasons Behind the Success of Cybercrimes

The traditional method of cyber security is designed in a very sophisticated way to analyze and get rid of these security breaches. Today, IT leaders are losing control of the technology. The modern day infrastructure is growing at a rapid pace.

Traditional methods such as using firewalls and anti-viruses, is putting enterprises in jeopardy. The improvement in today's attacks are far more complicated for the traditional tools like firewalls and antivirus software. These methods are an obsolete way of securing our computer network and its high time to past them.

One of the main reasons behind these cyber attacks is the use of these software's for securing the systems. People still use passwords and pin codes to ensure the safety of their accounts in the cyber space. But these passwords and pin codes can be decrypt very easily, leaving them vulnerable against these attacks.

The average security threats are discovered approximately 6 to 9 months after the actual breach is committed. In order to avoid these attacks one should use impenetrable system that uses a unified system of software and hardware to validate any data that is send or received over the cyberspace. What we are seeing is the result that the attacks and threats are far bigger then they appear.

One such method used to identify and prevent these attacks is SIEM i.e. Security Information and Event Management systems and it uses a lot of planning to implement. This technology collects log data from different systems across the world to discover any suspicious activity or behavior of any kind. It is a very time-efficient software for the administrators.

This technology is being used in the industry since 2000 and they had a goal to help enterprises detect cyber frauds and data beaches at an early stage. SIEM programs do not reduce cyber attacks by itself, but help the authorities to detect these threats as early as possible by using functions like correlation of data from multiple devices, known anomaly patterns, etc. from many devices so that necessary steps could be taken. SIEM solutions can be very useful in events such as :

- It can assure that antivirus and Operating Systems Software, are all updated, and are capable of generating logs.
- can monitor doubtful user authentication
- can be used to monitor servers to find out if there has been any unauthorized attack
- can process which systems in the network have been affected by viruses and if the other systems are getting affected or not
- can provide all accessed files, especially the ones with private access

But despite being so efficient it does have some disadvantages. This software often ends up costing more than expected. SIEM technology requires proficiency that can sometimes be outsourced. It can be difficult to tune and it sometimes takes considerable amount to yield results.

Because SIEM technology collects structured data, there is not a very large volume of data that can be processed. This can cause collateral damage to the organization. These tools are not prepared to meet the needs of the modern architecture. We need to build software and enhance security systems that can actually add value. To have new approach and thinking.

This is where the Big Data analytics comes in play. The main aim is to give the next generation the tools and methods they need, in order to detect these attacks without the need of the Hadoop administrators or any other the scientists. The criminals have gained expertise in becoming almost invisible in the activities, it is needed to have something off their charts; hence a combination of Hadoop tools with well-programmed machine learning models.

IV. The Big Data Analytics

Big Data is large-volume, large-velocity and large-volume information assets that demand efficient, innovative forms of information processing for better decision making. Unlike traditional analysis methods, Big Data Analytics helps expose invisible structures, unknown correlations and other important organizational information. However, big data tools can analyze this data far better than the traditional methods that struggle processing big data within a specific period of time and at an unacceptable cost.

As the industry is growing, so is the data within. Today almost every organization is processing terabytes of information every day. Because of this the threat to these organizations has also increased. The term Big Data and Big Data Analytics is majorly used when we are concerned with large amount of structured as well as unstructured data with could previously not be handled. Many industries nowadays are using Big Data Analytics to overcome the security needs, for example the banking sector to detect transactions frauds.

Big Data analytics is developing effective defenses against cyber threats. Better and faster security management tools are reducing the critical time from detection to remediation, making it possible to identify frauds and breaches as soon as it occurs. It also helps us to question from observation, formulate new hypotheses, explore and discover new concepts, and make decisions. The main efforts done by big data analytic is the use of new analytics techniques on either new data or data that has been mixed in new ways.

Big Data Tools is the first step towards the cyber security, using machine learning, text mining and ontology modeling for detecting threats and attacks. One of the main components of the Big Data Analytics is the Hadoop by the Apache Network Foundation. Hadoop is an open source software mainly designed to process large amount of information. It provides a programming model called MapReduce to implement parallel processing. With this, Hadoop also has a distributed file system called HDFS (Hadoop Distributed File System) to store and process manipulate large data sets.

Equipped with these abilities, the scientific researches identify new possibilities to practice new methods and technologies. The Hadoop ecosystem has reached a level of maturity and capability such that more and more organizations can use it for more and more cases. The main purpose of Big Data Analytics is to become more effective in recognizing the pattern that represents network threats by learning more about the organizations' cyber security defense mechanism.

V. Hadoop

Hadoop is an open-sourced computational software for processing and analyzing large amounts of structured and unstructured data. It is built as a java platform and is designed to execute queries and other operations against large datasets that can be tens of terabytes and even petabytes in size.

Hadoop has multiple concepts like HDFS, Map-Reduce, HBASE, PIG, HIVE, SQOOP and ZOOKEEPER to perform the easy and fast processing of huge data. In a Hadoop cluster, data is distributed to all the nodes. A MapReduce program has two steps: the Map function analyses input data and the Reduce function collects this result to form the final result. Each cluster node has a local file system and local CPU on which the MapReduce programs runs. Data is broken into small packets, stored on different nodes, and are stored in a three different locations for security issues. There are many nodes in each cluster in the machine. Hadoop is also used for web searches, email spamming, recommending search engines and for analyzing unstructured data.

Today Hadoop is used in almost every aspect of the society. The scientific community uses Hadoop to monitor natural phenomenon. The science and intelligence society needs to analyze large amounts of data generated by servers, email, instant messaging etc. to identify potential terrorist threats. The publishing industry uses Hadoop to index and reformat huge document stores.

A Hadoop based cyber security system can unlock any mystery regarding the cyber attacks. With a variety of processing options and algorithms available, scientists and researchers can now have a real idea about the situation and what can be done. The following list gives an overview of some cyber security projects using Big Data technology (especially Hadoop and MapReduce).

A. DOFUR:DDoS forensics using MapReduce

DDoS or the Distributed Denial of Service attacks has become serious as one of the menace in the Internet society. This attack is an attempt to make the Internet services unavailable to the customers by creating large amounts of data logs. These attacks can take down a whole website in a very short period of time. A DDoS is not any random attack but well planned and coordinated. With the large data sets generated, the administrator finds it very difficult to know the exact source of the attack before its too late. The entire process is divided into 4 steps :

- Data Collection: Event data, logs are collected from firewall. This collected data is collected in big data appliances.
- Data Processing: This step checks whether the collected data satisfies certain requirements. Then it is analyzed and processed using No-SQL, Hadoop and Map reduce methods.
- Data Analysis: The data from previous step is again analyzed using prediction and classification to gain insight of the user behavior, system status or any malicious activity in the system.
- Result: If attacks are detected, it informs the administrator and terminates. Predicted information of analyzed system is reported to the authority.

B. APTs

APTs are network attacks where a criminal breaches security firewall of a network to gain access to unauthorized data. the person stays there on the network hidden until he is able to steal data or personal files to cause damage to the organization. Big Data Analytics is a suitable method for APT detection. The main problem in detecting APTs is analyzing large data-sets for any abnormality. This large amount of data makes the analyzing process look like finding for a needle in a haystack.

VI. Conclusion

The WWW seems to be a massive environment but surprisingly one of its qualities, bringing the world closer is making it a small place to live in for its customers. Cyber attacks are a danger to the organizations. Bad guys have better weapons and the organizations are becoming vulnerable. We need to develop new software's for securing networks. The problem using large volume of data for identifying a cyber-crime is the skill to quickly process and identify threats before its too late. Big data analytics are used to identify unusual behaviors, identifying threats and allowing instant action to minimize or prevent losses. It provides real-time security measures to discover threats that are unlikely to discover. By using Big Data technology such as MapR Converged Data Platform, these organizations can fortify the national security; transform teaching, and much more. Big Data Analytics tools provide enterprises a cost-efficient and stable architecture to analyze various unstructured and structured data in no time.

References

1. AFCEA Cyber Committee-*Security and Cloud Computing*
2. Journal of King Saud University – Computer and Information Sciences (2013) 25, 63–75
3. Georgia Tech Research Institute Cyber Technology and Information Security Laboratory- *Cyber Security:*
4. International Research of Scientific Research - *Cyber security a challenge to developers, is face recognition technique a solution to this problem?* Volume: 2 | Issue: 7 | July 2013 • ISSN No 2277 – 8179
5. International Research of Scientific Research - *Emerging Challenges to Cyber Security-Internet Monitoring with Specific reference to National Security*, Volume : 1 | Issue : 2 | July 2012 • ISSN No 2277 – 8179
6. Terradata *Big Data Analytics in Cyber Defense ,* Ponemon Institute, February 2013C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
7. 2015 6th International Conference on Information and Communication Systems (ICICS) - *Enhancing Security of Hadoop in a Public Cloud* (Translation Journals style)," *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9th Annu. Conf. Magnetics* Japan, 1982, p. 301].
8. Information Systems Group Saarland University - *Efficient Big Data Processing in Hadoop MapReduce*
9. Bhawna Gupta et al, / (IJCSIT) International Journal of Computer Science and Information Technologies - *Big Data Analytics with Hadoop to analyze Targeted Attacks on Enterprise Data*, Vol. 5 (3) , 2014, 3867-3870