How Big Data Can Improve Cyber Security

Haydar Teymourlouei¹, Lethia Jackson²
¹Department of Computer Science, Lincoln University, PA, USA
²Department of Computer Science, Bowie State University, Bowie, MD, USA

Abstract - The research presented in this paper offers how Big Data can improve cybersecurity. Cybersecurity concerns affect organizations across virtually all industries, including retail, financial, communications and transportation. Big Data and analytics are some of the most effective defenses against cyber intrusions. Hackers continue to develop powerful offensives to prevent what used to be considered highly effective cyber defenses. With the resources available to hackers today, they can move around a defense-in-depth strategy to breach data systems. With tools and techniques that now exist to handle the volume and complexity of today’s cyber-attacks, enabling enterprises to stay ahead of evolving threats. Big Data along with automated analysis brings network activity into clear focus to detect and stop threats, as well as shorten the time to remedy when attacks occur.

Keywords: Big Data, cyber threats, logs, data reduction method, network, security

1 Introduction

The ability to accumulate large amounts of data provides the opportunity to examine, observe, and notice irregularities to detect network issues. Better actionable security information reduces the critical time from detection to remediation, enabling cyber specialists to predict and prevent the attack without any delays. Data is analyzed using algorithms which give critical insight to organizations in order to provide assistance in improving their services. Big Data is continuing to be used on bigger platforms including financial services, health services, weather, politics, sports, science and research, automobiles, real estate, and now cybersecurity. An important way to monitor your network is to set up a Big Data analysis program. A common response to evolving attacks is to either add more security tools or increase the sensitivity of the security tools already in place. Big Data analysis is the process of viewing large data sets to reveal hidden patterns, unknown correlations, market trends, customer preferences and other important information. To ingest all the data, filter and aggregate the data first, but it is tricky and difficult to decide what to filter out and what to keep. With the help of log monitoring tools and advanced data reduction method we can defend against cybersecurity. To help prevent cyber-attacks it is necessary to monitor our network.

2 What is a Big Data?

Data is a collection of facts, such as values or measurements. Information is often the result of combining, comparing, and performing calculations on data. Big Data is high-volume, high-velocity and high-variety information assets [9]. Billions of bytes of data are collected through various mediums. Big Data demands cost-effective and innovative forms of information processing for enhanced insight and decision making. There is always an issue with storage and processing large amounts of data. Privacy and security can be compromised while storing, managing and analyzing the large quantities of data [2]. When dealing with Big Data [5], it is necessary to maintain a well balanced approach towards regulations and analytics. Data comes from multiple intrusion detection systems, as well as sensors from high noise environments. Using analytical tools, data management and examination techniques help us detect attacks. Big Data research and development is needed in the academic, industrial and government research labs to develop solutions that defend and protect large data sets.

3 Data Management Techniques

Changing the organization’s operation from the normal system usage of data to handling and maneuvering Big Data brings the difficulties of tricky advanced persistent threats. Data privacy, integrity and trust policies should be examined inside the context of Big Data security. The cybersecurity of Big Data operates in high volumes. Data can come from multiple intrusion detection systems, sensors as well as sensors from high noise environments. As indicated by break-out group participants, denial of informational attacks and the demand to deal with malicious opposition are threats to data privacy. Using analytical tools, data management and examination techniques that integrate data from hosts, networks, social networks, bug reports, mobile devices, and the internet of things provides sensors to detect attacks. Hashing techniques and data technologies also play a major role in strengthening security systems.

Data management and examination techniques such as biometric authentication [12] defend against cyber-attacks by providing solutions to security issues of protecting massive amounts of data. Many data communities are developing large data and solutions for efficiently managing and analyzing large sets of data. Continuous Big Data research and development is needed in the academic, industrial and government research labs to develop solutions that defend and protect large data sets including cloud data management. Cloud data management includes malware detection, insider threat detection, intrusion detection, and spam filtering [8].

4 Managing the Cyber Threats
Government and industry authorities are now enforcing regulations on how companies combat cyber-attacks. With the help of latest technology they are able to defend against cyber threats. Network Behavior Analysis (NBA) has been an emerging technology that serves as a security direction tool to improve the current network status. NBA proctors traffic coming in and traffic going out of the network to ensure that nothing is getting into the host package and application. By 2011, approximately twenty-five (25) percent of large security systems will use NBA. A major disadvantage of NBA is that it does not catch the security breach before it becomes a problem. There are many challenges when it comes to enforcing access and security control policies in Big Data storage environments [3]. Enforcing access and security control policies creates challenges, because in some instances each user would have fined grain access control based on their specific job or responsibility [1]. Companies ensure that their employees are within compliance, and it keeps logs of user activity. Keeping track of user activity in real time helps control threats as the data can be analyzed for unusual or suspicious activity and then treated accordingly. In addition, an event management technology uses Big Data technologies and methods to combat attacks. While the development of the data has encouraged an increase of cyber threats it concurrently manages and reduces their occurrence.

5 Different Methods to Prevent of Cyber Threats

With the help of Big Data log analytics we can prevent cyber threats by monitoring data. When Big Data log analytics is combined with JIT (Just in Time) analysis it collects information on the machines that have an open connection with locations outside the local network. It also predicts future attacks and gives you information about previous attacks that might have occurred on your system. An IBM report shows us that forty-six (46) percent of businesses are experiencing security breaches; which shows that the need to protect our information is very high. IBM developed a solution using Big Data that protects data from threats and fraud. IBM’s solution detects risk and intrusion while analyzing structured and unstructured data. QRadar performs real-time correlation, anomaly detection and reporting for immediate threat detection, and also sends enriched security data to IBM Big Data products, such as IBM InfoSphere BigInsights [8].

Moreover, Hadoop is a Java-based programming framework that helps to store, organize, process, and analyze data. The analytical ability of Hadoop allows CounterTack, a form of APT, to analyze system level information collected from any kind of data to detect intrusion or malicious behavior. CounterTack uses Cloudera, which uses Hadoop algorithm, to protect against security threats. Hadoop allows us to manipulate through data affording us complete access to the information. Also, Apache Storm is a free open-source real-time streaming analytics. It is similar to Hadoop, however, it was introduced for real-time analytics. It is fast and scalable, supporting not only analytics in real-time but machine learning too. The Apache Storm algorithm reduces the number of false positives found during security monitoring. Apache Storm is commonly available in cloud solutions that support antivirus programs which includes Big Data used to analyzed as well as uncover threats.

6 How Log Monitoring Can Prevent Cyber Threats

Monitoring data validates access to the systems, tracks logged-in activities, prevents a breach as well as manages passwords efficiently. Surveillance is a major function of securing any network and the first deterrent for hacking or infiltrating anything in general. Continuously monitoring surveillance deters the hacker from successfully infiltrating the system. Other prevention methods include regular scans, updated virus protection software and firewalls. Free software is available like Barracuda’s Firewall, McAfee’s Advanced Protection or Patrol that monitors and protect against security breaches.

All operating systems come equipped with standard log in features and are often overlooked. Software that may be used to monitor the system for security has a built-in event logger found in Windows based operating systems. This tool gives administrators (audit on Windows Server) the ability to view events that vary from login actions against files to either access a file, delete a file or create a file, and much more. These log files keep track of events that are produced from hardware and software operations. They can range in notification from informational events to warnings and subsequently to critical errors. However, not all events are collected by default, important audit settings must be turned on in domain group policies and on the folder or device that contains valuable data in order to receive the events. Log files can alert an administrator that a significant file has been modified, deleted or an unsuccessful attempt to change the file raises a red flag.

Log monitoring usually consists of three tiers; log generation, log analysis, and log monitoring. Log generation involves the hosts making their logs accessible to the servers. Log analysis receives log data that is converted into the necessary formats so it can become a single standard readable format. Log monitoring includes console or software that handles monitoring and reviewing of log files. With log monitoring we can monitor system traffic to detect and prevent cyber-attacks. Additional log monitoring benefits include tracking suspicious behavior, meeting regulatory compliance, detecting/preventing unauthorized access, as well as monitoring user activity. Also, there is tool called System Information and Event Managements which handles the generation of reports, alerts when an abnormality is present, manages incidents, and correlates and analyzes events.

6.1 How to Prevent Cyber Attacks using Event Viewer

When a windows server or machine has problems or issues, the administrator would want to know what is the activity that
caused the problem. The event viewer allows the administrator to view the activity. Each activity, referred to as an event, is assigned a number. There are certain events that require more caution than others and warrant more attention to prevent an intrusion. For example, Event 4724 represents an innocent password reset or is it someone trying to wreak havoc on your system. Another example, Event 1102 which represents someone clearing the security log or is it someone trying to erase their tracks after their attempt to intrude in unauthorized territory. Windows 2008 and later versions contain built-in alerting of certain events. To prepare Windows to give an alert of certain events, perform the following steps. Step 1: Open the task manager and create a basic task. Step 2: Give the task a name, for example monitorEventxx, and a brief description of the event. Step 3: Set the trigger to identify when a specific event is logged. Step 4: Specify which log to look at and the event Id number. Step 5: Select the action whether email, message, or start program when the event is logged. Verify the message is correct when completing this step. In the case that an event is logged, Windows will perform the alert. From this point the administrator can validate if the event requires more attention or not. Using this method is highly preventative of cyber threats and can provide real time speed to thwart a concurring intrusion.

6.2 How Data Compression and Reduction Method Could Help Cyber Threats

In many instances, investigators require only a portion of Big Data; however, data is delivered in voluminous amounts. This leaves researchers to combat the cost it takes for Big Data to be transferred. This goal to compress Big Data is achieved using a reduced data size.

The size of large datasets must be reduced in order to successfully identify anomalies. The algorithm offers an efficient data reduction method that effectively condenses the amount of data. This algorithm enables users to store the desired amount of data in a file and decrease the time in which observations are retrieved for processing by using a reduced standard deviation range that significantly minimizes the original data. View the diagram below for an outlook on how large Big Data is presented.

![Figure 1: Actual Plotted Big Data](image1.png)

It is clear that when attempting to plot large sets of Big Data, researchers’ ability to attain desirable results is hindered. Often times only a small portion of the data sets is of interest to the researcher.

![Figure 2: Condensed Version Plotted from Big Data](image2.png)

The following diagram reveals the detection of a real cyber threat. Data reduction method was able to successfully locate the threat and provide a range of severity of the threat from low to high after reducing a large data set to a reasonable size. The use of a reduction method effectively condensed the amount of
The range of severity for cyber threat detected in this example is high. The algorithm enables users to compress Big Data by a percentage value. The program instructs the user to input the following:

- maximum and minimum threats
- desired percentage that such Big Data should be reduced

The data in Table 2 includes many negative numbers. However, the program ignores the numbers that indicate a severity range of low such as negative numbers (-1). Therefore, in this particular illustration, there are only a few data values that are of significance based on the percentage that the program was instructed to perform. This critical reduction makes data easily retrievable and accessible. In addition, it provides researchers with the amount of data that examiners request.

7 Conclusions

Organizations need to be current with the latest vulnerabilities to prevent known attacks. Big Data will quickly solve the problems of cybersecurity. The reality is that Big Data and analytics will allow companies to identify anomalies and advanced attack vectors. Cybersecurity requires risk management and actionable intelligence that is common from Big Data analysis. While it is great to have tools that can analyze data, the key is to automate tasks so that the data is available more quickly and the analysis is sent to the right people on time. This will allow analysts to classify and categorize cyber threats without the long delays that could make the data irrelevant to the attack at hand. Big Data will also help analysts to visualize cyber attacks by taking the complexity from various data sources and simplifying the patterns into visualizations.

Nearly one million malware threats are released daily and the estimated cost of cyber-crimes average billions. The importance of continuous vulnerability management should not be overlooked because exploits and viruses are constantly evolving. Although cyber criminals will always pursue weaknesses in computer systems, the countermeasures put in place today will help deter attacks in the future. Big Data analytics has already produced positive results in its efforts to reduce cyber threats. 84 percent of Big Data users say their agency has successfully used Big Data analytics to thwart cybersecurity attacks and 90 percent have seen a decline in security breaches [13].

8 References


