ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 6, July 2017

Use of Data Analytics to Recognize Threats in the Security Systems of Banks

Prashasthavardhana Prahlad[1], Sachin K. Rai[1], Dr. D.V.Ashoka[2]

UG student, Department of Information Science and Engineering, JSSATE, Bengaluru, India[1]

Professor, Department of Information Science and Engineering, JSSATE, Bengaluru, India[2]

ABSTRACT: Big Data refers to a large volume of data, both structured and unstructured. Big Data Analytics refers to collecting and analysing sets of data by the application of algorithms to find hidden patterns. Big Data Analytics tools spanning around SIEM (Security Information and Event Management) can be of great help. Combining Big Data Analytics with Predictive Analytics can be a very crucial tool in removing security bottlenecks. Big Data Analytics can help identify the risk prone technologies and finding their alternatives, while the Predictive Analytics can be used to extract useful information from existing data sets by doing advanced statistical analysis of which can give reliable idea about the future trends. This will immensely help the developers to be future ready and build software and security tools for tomorrow to guarantee best guard against the security and data breaches. This paper gives the overview of big data analytics to recognize threats in the security frameworks of banking systems.

KEYWORDS: Big Data, Big Data Analytics, SIEM, Predictive Analytics.

I. INTRODUCTION

I. What is Big Data?

Big Data is a generalized term that describes large amount of data. This data can be both structured, as well as unstructured. The data that can be grouped as big-data is the huge volume of information that is generated by an organization on a transactional basis. Although the data generated is very large, the more important point is what these organizations do with the data. Big-data can be analysed in a very detailed manner to make changes in the existing strategies, or bring about new plans and make better decisions, which ultimately help the organization grow.
According to McKinsey, "Enormous Data alludes to datasets whose size are past the capacity of regular database programming apparatuses to catch, store, oversee and break down"[1].

II. History of Big Data

About 90% of the available data has been created in the last two years. The term Big Data has been around since 2005, when it was launched by O'Reilly Media in 2005. However, the usage of Big Data and the need to understand all available data has been around much longer.

III. Big Data Characteristics

Big data was initially given 3 defining characteristics, which are called as the "3Vs of Big Data". These characteristics were "Volume, Variety and Velocity". But with the improvements in the analytics field, it was found out that Data Scientists needed to take care of few more "Vs", which are "Veracity, Volatility, Validity", in addition to the original 3Vs[2].
1. Volume: This deals with the size of the data, and as the name suggests, it is very big.
2. Variety: This refers to the various sources from which different types of data, both structured and unstructured is derived.
3. Velocity: This deals with the rate at which the date flows in organizations, business machines or human interaction with websites, mobiles and so on.
4. Veracity: This refers to the biases, noise and abnormality in the data.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 6, July 2017

5. Validity: This is mainly concerned with the correctness of the data and whether the obtained data is accurate, not out-dated and is used for the intended purpose.
6. Volatility: This tells us how long the data is valid. It also tells for how long a particular data set can be stored.

IV. What is the importance of Big Data?

For organizations it is important to make maximum use of their data. Big data analytics allows this and use it to identify new strategies to reduce cost and optimize the business or organizational process. It also helps organizations to develop better products. Further, it helps them to identify potential risks and take preventive actions. By doing so, it helps the organizations to derive efficient operations, greater profits and happier customers[3].

II. SECURITY INFORMATION AND EVENT MANAGEMENT

I. Introduction

Security Information and Event Management (SIEM) services can be called as the combination of SIM (Security Information Management) and SEM (Security Event Management) services. SIEM is primarily used for the dynamic analysis of the security aspects of network. The solutions of SIEM are available as software and appliances which generate a detailed report of the security system and also helps to discover potential threats[4].

II. Key Focus of SIEM

The key focus of SIEM is to monitor and help manage user and service privileges, directory services and other system-configuration changes; as well as providing log auditing,it's review and incident response.

III. SIEM Capabilities

1. Data Aggregation: Log management aggregates data from many sources, including network, security, servers, databases, applications, providing the ability to consolidate monitored data to help avoid missing crucial events.
2. Correlation: Looks for common attributes, and links events together into meaningful bundles, in order to turn data into useful information.
3. Alerting: The automated analysis of correlated events and production of alerts, to notify recipients of immediate issues.
4. Dashboards: Tools can take event data and turn it into informational charts to assist in seeing patterns, or identifying activity that is not forming a standard pattern.
5. Compliance: Applications can be employed to automate the gathering of compliance data, producing reports that adapt to existing security, governance and auditing processes.
6. Retention: as it is unlikely that discovery of a network breach will be at the time of the breach occurring, employing long-term storage of historical data to facilitate correlation of data over time, and to provide the retention necessary for compliance requirements is of high importance in forensic investigations.
7. Forensic analysis: The ability to search across logs on different nodes and time periods based on specific criteria[4].

III. PREDICTIVE ANALYTICS

I. What is Predictive Analytics?

Predictive Analytics is a branch of “Advanced Data Analytics”, that utilizes many concepts from data mining, machine learning and predictive modeling, to analyze the current data to identify certain trends and make predictions about unknown future events. Predictive Analytics involves the analysis of transactional data to identify relationships between data sets, and based on other relevant factors, make predictions about the future based on the observed trends. This becomes very important for organizations as the assumptions or predictions are based on data and not a hunch. This allows the organization to identify potential risks, and gives them more time to employ strategies to counter these risks[10]. It may also go further and suggest certain decisions to be taken based on the predictions and its implications. Prediction process can be divided into four steps: (1) collect and pre-process raw data; (2) transform pre-processed data into a form that can be easily handled by the (selected) machine learning method; (3) create the learning model (training) using the transformed data; (4) report predictions to the user using the previously created learning model[5].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 6, July 2017

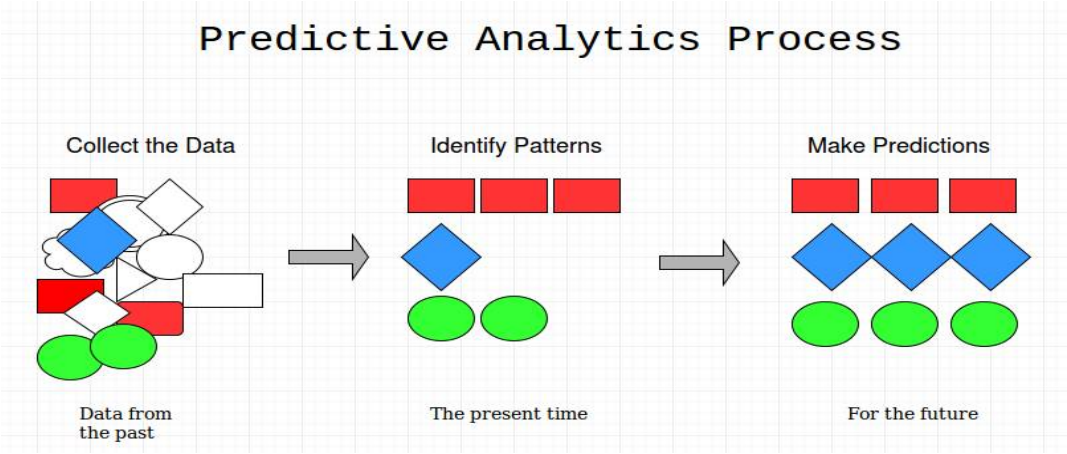

Figure: Predictive Analytics Process.

II. General Predictive Analytics Applications

1. Fraud Detection: Fraudulent and inaccurate credit transactions, either offline or online, can be detected to prevent theft and also invalid insurance claims.
2. Direct Marketing: A combination of the best product versions, medium of communication and marketing material can be decided to entice more number of customers.
3. Underwriting: Predictive Analytics can be used to identify future customer risk behaviour based on the application level data. It can also be used to detect illness, bankruptcy et al.
4. Health Care: It can also be used to identify the diseases that a patient might develop in the future such as diabetes, asthma et al.
5. Risk Management: It is used to identify potential risks which may hamper the organization from getting maximum profit from a certain product[5].

III. Predictive Analytic Tools

Some of the popular "Open-Source Predictive Analytic" tools include:
1. R
2. Apache Mahout
3. OpenNN

Some of the popular "Commercial Predictive Analytic" tools include:
1. MATLAB
2. IBM SPSS Statistics and IBM SPSS Modeller
3. Oracle Advanced Analytics

IV. DATA BREACH

Data breach particularly refers to gaining access to important and confidential data using unauthorised means. This can result in grave losses to an entity whose data has been breached, especially when it comes to financial firms such as banks. Hackers finding their way into the security by stealing passwords using sophisticated techniques such as dictionary attacks, brute force attacks and hybrid attacks to directly injecting malicious viruses or malwares into the security system[6].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 6, July 2017

I. HISTORY OF DATA BREACHES

Whenever there is large amount of data either being produced, consumed or stored, it is always likely that this data can fall in to the wrong hands. Before computation brought the era of digitisation, the number of data breach cases weren't that frequent, but ever since 1980s these cases has been on the rise. With their frequent occurrence, people became aware of the potential threat that these breaches can directly cause to them. Most of the major data breaches occurred after the year 2005. The primary reason for this was huge amount of data that was produced ever since. According to a 2012 report by *CSC*, it was estimated that by 2020 data production is going to be 44 times more than what it has been in 2009, also there is going to be 4300% increase in annual data generation by 2020[6].

Year	No. data breach cases	No. of records exposed(in millions)
2005	157	66.9
2006	321	19.1
2007	446	127.7
2008	656	35.7
2009	498	222.5
2010	662	16.2
2011	419	22.9
2012	447	17.3
2013	614	91.98
2014	85.61	783

Table: Data breach trend in the past decade ***Source***: *Statista.com[8]*

One can see from the above table that in recent times, since 2011 number of data breach cases in USA has been on the rise. Also the number of data breaches has increased almost 500 % in 2014 from 2005. One of the sectors worst hit by data breaches is "*Banks*". The 2016 Annual Verizon Data Breach Report revealed that 89% of the total breaches had financial or espionage motive, making banks the most vulnerable entity[7].

In October 2016, just a month before the major disruption in the Indian Economy terrain, in the form of '*Demonetisation'* when high value currency notes ceased to be legal tender overnight, several Indian banks reported a data breach, biggest of its kind in India. Findings of the SISA report revealed that the news regarding the breach surfaced only after couple of months from the date 21st May, when the breach actually occurred . It was estimated that almost 3.2 million debit card data was compromised. Major Indian lenders including SBI, ICICI, Yes Bank, HDFC and Axis Bank were among the worst hit. Various reports surfaced about unauthorised use of cards by the hackers in China and USA. This resulted in one of the India's biggest card replacement drive in the banking history, with SBI blocking and replacing almost 6,00,000 debit cards. SISA report further revealed that a malware was injected into the payment gateway network of Hitachi Payment Services Pvt. Ltd. Which many banks use for their ATM transaction processing. It was also revealed that the malware had captured the debit card number and PIN of those customers who used their cards at the affected ATM, thus the accounts of these customers became vulnerable putting their financial assets in the hands of the hackers. The demonetisation drive is going to increase the pressure on the Indian financial firms to strengthen their payment security features cause of the added impetus to digital payments now.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 6, July 2017

II. Major Causes

A 2015 study conducted by *Ponemon Institute* for IBM revealed that malicious attachments, shellshock and DoS(denial of service) constitute the top three cyber security that affects the financial sector[7]. It stated that 47% of the data breaches that take place to malicious criminal attacks such as phishing, hacking and DDoS attacks. Another very famous 2015 Annual *Verizon Data Breach Report* states that POS(Point of Sale) intrusions accounted for 28.5% of confirmed data breaches and 0.7% of total incidents, followed by Crimeware attacks attributed to malwares and accounting for 18.8% of confirmed breaches and 25.1% of total incidents, which was followed by Insider Misuse, attributing for 10.6% of confirmed breaches and 20.6% of total incidents. Hackers can be attributed to be the biggest culprit to this grim situation of rising data breach cases. About 76% of the total attacks of hacking involved exploitation of weak credentials such as passwords. Authentication based which primarily comprises of getting the hold of users account passwords by sophisticated guessing techniques or using other sites on the target system comprised of about 4 of every five hacking incidents that occurred in 2012, says Verizon. Verizon also found that stolen passwords played a key role in data breaches, almost 48% of all the hacking involved stolen passwords which could have been found by the repository of previous stolen password lists of previous data breaches[7][9]. On a fair analysis we can see that data breaches caused to malware attacks and hacking has been on the rise recently.

III. Recommendations and Solutions

One of the critical issues attributed to data breaches is the delay that occurs in knowing that a data breach has actually occurred. Like in the most recent data breach event that took place in India, security experts came to know about data breach occurrence only after couple of months of the initial attack. This delay exacerbates the impact of the data breach and it exponentially increases the losses. It also results in more delay in finding the solution to tackle the security bottlenecks and make the security system robust with new enhanced security techniques. This problem can be tackled using Big Data Analytics SIEM(Security Information and Event Management) tool. This tool can be of a great help if used efficiently. Log management available can help in aggregating data related to previous breach events such as their cause, the tools used for breach, new sophisticated techniques. This helps in finding every key aspect without missing any critical event that caused the attack. Correlation technique can be then applied to this aggregated data to find the links and patterns between the way the attack on the data occurred, their cause and amount of data breached. This can be supplemented with automated analysis of correlated events and production of alerts which is a reasonably fast way to notify recipients of the immediate issues. Long-term storage of previous data expedites the correlation of data overtime. Applications can be employed to automate the gathering of data that can expedite the whole process. SIEM tools can be further strengthened by supplementing them with Predictive Analytics. Predictive analytics can be efficiently used to achieve customer relationship management by using analytics to gain an insight about customer's life cycle, their activities trend etc. It can further help in expediting Fraud detection and Risk Management by predicting accurately and reliably any possible future attacks on the security system. Since the main culprits of various data breaches can be attributed to phishing, hacking, malware injections, there is a need to find novel and better security solutions to tackle the ever changing new ways of breaching the data. This may include the organization to strengthen their network firewall or by developing applications that detect any foreign code. By doing so, the organization is reducing the time between the data breach detection and the actual occurrence. This may save potentially up to billions of customer's money in banks.

V. CONCLUSION

The security systems of most organiztions around the world are becoming robust day-by-day. However, it cannot be ignored that even the frequency at which large scale data breaches are occuring is also increasing. While studying the topic, it became quite clear that there was a huge time difference between the time of the data breach and the time of it's discovery. This is where Predictive Analytics becomes a very strong method in detecting a data breach. Any abnormal transactions are quickly brought to the notice of the organization due to the fact that the past events have been thoroughly analysed, and reliable predcitions based on modelling can be made. If this is solved quickly, the organization can take steps to prevent any more data from being breached, thereby saving a lot of important data, and in the case of banks, the customer's valuable money. Thus, it can be said that Predictive Analytics will play a prominent role in the strengthening of the current security system of banks.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Special Issue 6, July 2017

REFERENCES

[1] Sunny Sharma, Prithvipal Singh, "A Review toward Powers of Big Data" in International Research Journal of Engineering and Technology(IRJET), Volume 03, Issue 04, Apr-2016.
[2] "Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity" by Kevin Normandeu, 2013/09/12, in InsideBIGDATA.
[3] Kalyani Shirudkar, Dilip Motwani, "Big-Data Security" in International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 3, March 2015.
[4] Eva Kostrecová, Helena Bínová, "Security Information and Event Management" in Paripex-Indian Journal of Research, Volume 4, Issue 2, Feb 2015.
[5] Nishchol Mishra, Dr. Sanjay Silakari, "Predictive Analytics: A Survey, Trends, Applications, Oppurtunities & Challenges" in (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3(3), 2012.
[6] "The History of Data Breaches" by Nate Lord on www.digitalguardian.com
[7] http://www.crn.com/slide-shows/security/300076528/verizon-report-top-9-causes-of-data-breaches.htm/pgno/0/7
[8] https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/
[9] http://www.darkreading.com/attacks-breaches/the-eight-most-common-causes-of-data-breaches/d/d- id/1139795
[10] http://www.predictiveanalyticstoday.com/what-is-predictive-analytics/