SECURITY ISSUES IN THE BIG DATA: A BIG PROBLEM

Madan Mohan [#1]

#1 Ph.D. Scholar, Noida international university, Greater Noida, India. Email: mmphdcse@gmail.com

I. ABSTRACT

Big data is a great way to be followed especially for large companies that have a lot of information about their system.

This paper delves into the new challenges associated with big data. It points out safety challenges on Big Data as the main issues that organizations seek to address on a day-to-day basis. These challenges include securing the trusted environments, enough access management, performing due diligence, combating AVI vulnerabilities, and security automation. They can be solved by maintaining strict access strategies that only allow their esteemed and responsible employees to log in and also set the systems in such a way that they can detect abnormalities and allow for investigations while there is still time. The paper has addressed big data challenges as well as their solutions which are always be considered in the case of an organization as they have long-term consequences if not put into consideration.

Keywords: *Big Data, Systems, AVI Vulnerabilities, Security Automation, Solutions, Trusted Environments*

BIG DATA

Big data is a word used to describe large-scale structured and unresolved data very much, it is very difficult to process data using traditional database and software technology.

Volume: The quantity of produce and stored data, many factors contribute to an increase in volume, live streaming data and Storing transaction data, live streaming data and data collections from sensors

Variety: There are data in all types of formats

Variability: Together with speed, the data flow may not be consistent with the periodic top of something height.

Complexity: When data comes from multiple sources, you also need to consider the complexity of the data. Before actual processing, the data must be linked, matched, cleaned and converted to the desired format.

II. BIG DATA SECURITY

The 21st Century business milieu regards data as its most crucial asset. Earlier on it was only the technological field that believed in the value of data but now all industries ranging from manufacturing, health, education, media and all others are already in the clique. Due to this growth, the volumes of data to be handled in organizations has increased and facilitated towards the formation of Big Data for the sake of saving all the information in the most original and quality form. Big Data has however been associated with issues of security and privacy that have been in existence for a long period [1]. Hacking activities motivate people to access information with no authorization and make it available to the wrong parties who might take advantage of the data obtained.

Privacy and security threats have been a challenge when dealing with Big Data for the longest time possible but there are new issues to watch for in 2018. The protecting of trusted environments, sufficient access management, performing due attentiveness, combating API weaknesses and security automation are the most recent challenges in Big Data security which should be solved to ensure for trustworthy operations of the systems [2].

Most organizations consider using the cloud when it comes to big data although it is not restricted. However, cloud computing and Big Data go hand in hand towards ensuring that the information of a company is maintained in the most effective way possible. However, since many hackers have realized that Big Data holds all the valuable details about the running and development of the firm, the system becomes a weakness (Lystra's & Papadopoulos, 2018). The following are the recent problems in Big Data security and how to solve them for a better future in protecting organization's information.

The main causes of security issues on big data can be shown through the chart below:

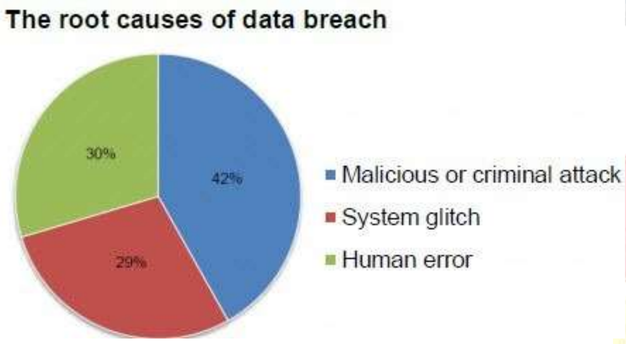

Figure 1: Illustration for the causes of big data security issues

Source: http://www.billchamberlin.com/cost-of-a-security-data-breach-rises-according-to-9th-annual-ponemon-institute-study/

III. SECURING TRUSTED ENVIRONMENT

As mentioned earlier, Big Data does not have to entirely depend on the cloud and many large companies use internal environments to run NoSQL or Hadoop databases. Such environments are crucial in preventing the systems against any external security attacks but suffer from insider threats. Not all employees in an organization are happy about their position in the organization and this state might lead them into accessing valuable information about the company and using it for their own self-interests (Mishra & Kumar, 2018). Rivals and enemies most of the times use an inside person to acquire finer details of a company and use it to damage its reputation which might affect its position in the market or its ability to make significant growth.

An organization using other trusted environments should, therefore, do its best to secure them from misuse both by internal and external parties. A technique referred to as "anomaly detection" is the best strategy to apply in this case as it establishes the baseline of employee interactions with the systems [3]. If the unusual behavior is detected, then the Big Data systems will automatically send alerts to management which calls for the investigation of the possibility of any intrusion or suspicious activity. However, it can sometimes be difficult to detect intrusions from the authorized staff which in many cases are responsible for the biggest downfalls of reputable firms. Another way to secure the data is by ensuring that the servers contain sensitive and valuable information have two-factor verification during access so as to reduce the number of people who have the right to use these servers.

IV. SUFFICIENT ACCESS MANAGEMENT

Most of the severe security breaches on cloud usually occur because of inaccurate access administration. Poorly managed access systems increase the probability of employee errors while login or trying to acquire particular information during their day to day operations. The weak access patterns facilitate for future successful attacks and data breaches.

The best way to handle this situation is by enforcing the principles of least rights for workers to minimize any possible damages that come with a compromised account. The company should also allow IT managers to control the users within all the systems of the cloud and allow for a more visible and transparent procedure [4]. If all employees are aware of the details regarding the access of information in the systems or alternatively allow only a few people to exclusively login.

V. PERFORMING DUE DILIGENCE

Businesses are trying hard to compete and hence ending up copying each other's styles of operation. Some businesses are therefore taking on the Big Data as well as cloud computing as a way of storing and accessing data. The trend makes other organizations do it just because everyone is doing it and end up failing to execute thoroughness on their selected cloud suppliers. It is, therefore, essential for all businesses to first evaluate and examine the multiple solutions and vendors of Big Data and cloud putting into consideration their security features. The thoroughness in the selection activity enables them to work only with the best and reduce the risks of security threats and data breaches.

VI. COMBATING API VULNERABILITIES

This factor focuses on cloud vendors as opposed to the users in the previous challenges. The vendors work directly with the providers and they expose the API to the consumer ensuring the relations with service providers. It is, therefore, the vendor's responsibility to make sure that the API's provided is secure enough to protect the large data stored on the servers [5]. The issue can be solved by the consistent fixing of the API and scan them of vulnerabilities to reduce any security threats and it also increases confidence in the users.

VII. SECURITY AUTOMATION

The IT staffs dealing with Big Data are faced with a great challenge of dealing with sensitive information during their day to day activities. They are required to classify, sort out and guard classified data including the Personally Identifiable Information (PII), Intellectual Property in addition to individual health data (PHI).

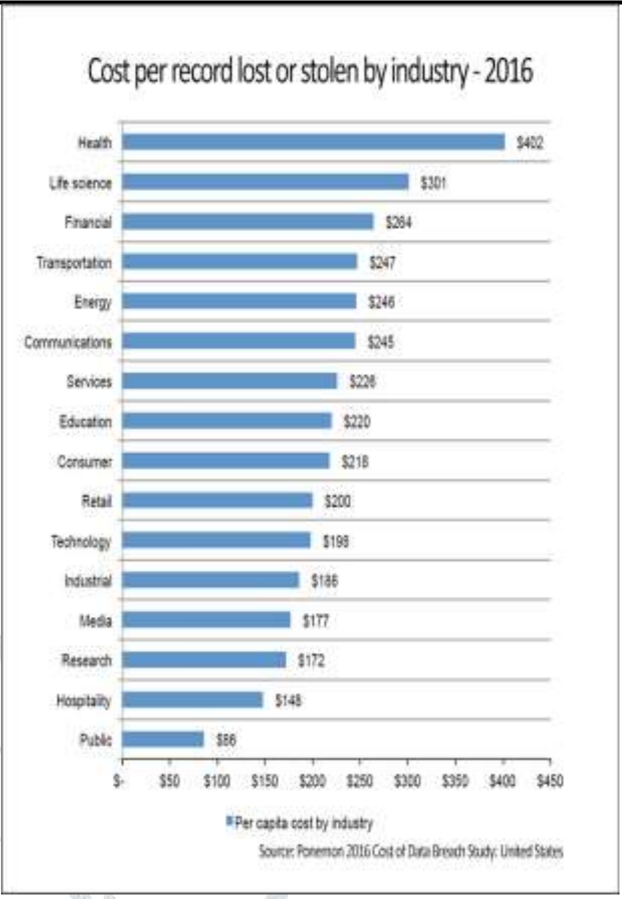

Figure 2: Illustrating the impact of data breach for different industries in the US-2016

Source: https://www.calyptix.com/hipaa/healthcare-data-breaches-expensive-average/

The best-advanced solutions including machine learning can be used in cloud computing to take over the labor-intensive activities that involve sensitive data. This method is involved in determining the way data should be stored and the methods that users access the information. The advancement is appropriate for monitoring the sensitive data and helps the IT staff as they send alerts when there are abnormal patterns detected.

CONCLUSION

In conclusion, apart from the benefits of large data, the challenges are pressing equally. These challenges reduce the level of user trust while investing in modern information access and storage systems. However, users and service providers have a duty to improve the privacy of large data. The policy enforced by each organization to safeguard the data determines whether its information will be safe or not.

The recent problems in Big Data present a huge challenge and it will be a big problem for the future. So before which contributes to major failures and hence the need for more improved solutions to make sure that the future is bright for technological adoption in organizations.

REFERENCES

[1] L., Daphne and M. A. Salem Durai. (eds.). *HCI challenges and privacy preservation in big data security*. New York, IGI Global, 2017.

[2] M. Brojo Kishore and K. Raghvendra. *Big data management and the internet of things for improved health systems*. New York, IGI Global, 2018.

[3] W. Yichuan, K. LeeAnn and B. Terry Anthony. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, vol. *126*, pp. 3-13, 2018.

[4] Q. Meikang., G. Keke, T, Bhavani., T, Lixin. and Z. Hui. Proactive user-centric secure data scheme using attribute-based semantic access controls for mobile clouds in financial industry. *Future Generation Computer Systems*, vol. *80*, no. C, pp. 421-429, 2018.

[5] L. Miltiadis D., & P. Paraskevi. *Applying big data analytics in bioinformatics and medicine*. New York, IGI Global, 2018.