International Journal of Engineering Research and General Science Volume 6, Issue 5, September-October, 2018
ISSN 2091-2730

Information Security in Big Data

Naseema Shaik[1], Mubeena Shaik[2], Nada Ahmed Mohammad[3], Fatima Ahmed Alomari[3]

[1] Lecturer, King Khalid University, Kingdom of Saudi Arabia

[2] Lecturer, Jazan University, Kingdom of Saudi Arabia.

[3] Student, King Khalid University, Kingdom of Saudi Arabia

Abstract: The Era of Big data and Information security tends to both opportunity and risk management for every business activities. In Big data, Information security standpoint big data has ushered in new possibilities in terms of analytics and security solutions to protect data and prevent future cyber attacks. In Information security point of view big data is a very large data set that is mined and analyzed to find patterns and behavioral trends. In this paper we focus on process involved in data processing and overlook the security issues of big data and describes the scope of big data in business field.

Keywords: Big data, Information security, Data Encryption, Hadoop file system.

Introduction:

Big data is becoming an immensely important part of the business plan for companies in many different areas. Analyzing huge customer datasets and other kinds of data with different tools. companies save money as well as boost revenue by targeting their marketing better, designing products to better service to their customers and make better predictions. Companies that use big data, especially if that data consists of personal information of customers, are at an elevated risk of drawing hacking attempts. Developing ways to protect that data will prove to be just as important as the data itself. Information security has experienced a profound paradigm shift from traditional perimeter protection tools towards monitoring and detecting malicious activities within corporate networks. Growing role of malicious insiders in the recent large scale security breaches clearly defines that traditional approaches to information security can no longer keep up. As the security industry's response to these challenges of security, a new generation of security analytics solutions has emerged in the past few years, which are able to collect, store and analyze large amounts of security data across the whole enterprise in real time. Big Data and advanced analytics to relate security events across multiple data sources, providing early detection of suspicious activities, rich forensic analysis tools, and highly automated remediation workflows. In this paper we are discussing the level of awareness and current approaches in information security and fraud detection in organizations around the world. Big Data security analytics initiatives, presents an overview of various opportunities, benefits and challenges relating to those initiatives, as well as outlines the range of technologies currently available to address those challenges.

I. Security issues on Big data:

Big data security is a constant concern because Big Data deployments are valuable targets to would-be intruders. A single ransomware attack might leave your big data deployment subject to ransom demands. Securing big data platforms takes a mix of traditional security tools, newly developed toolsets, and intelligent processes for monitoring security throughout the life of the platform.

Big data environments add another level of security because security tools must operate during three data stages that are not all present in the network. These are

a) data access/entry

b) Data storage

c) Data outcome

a) **Data Access / Entry:** Big data sources come from a variety of sources and data types. User-generated data alone can include CRM or ERM data, transactional and database data, and vast amounts of unstructured data such as email messages or social media posts. In addition to this, you have the whole world of machine generated data including logs and sensors. You need to secure this data in-transit from sources to the platform.

b) Data Storage: Protecting stored data takes mature security toolsets including encryption at rest, strong user authentication, and intrusion protection and planning. You will also need to run your security toolsets across a distributed cluster platform with many servers and nodes. In addition, your security tools must protect log files and analytics tools as they operate inside the platform.

c) Data Outcome: The entire reason for the complexity and expense of the big data platform is being able to run meaningful analytics across massive data volumes and different types of data. These analytics output results to applications, reports, and dashboards. This extremely valuable intelligence makes for a rich target for intrusion, and it is critical to encrypt output as well as ingress. Also, secure compliance at this stage: make certain that results going out to end-users do not contain regulated data.

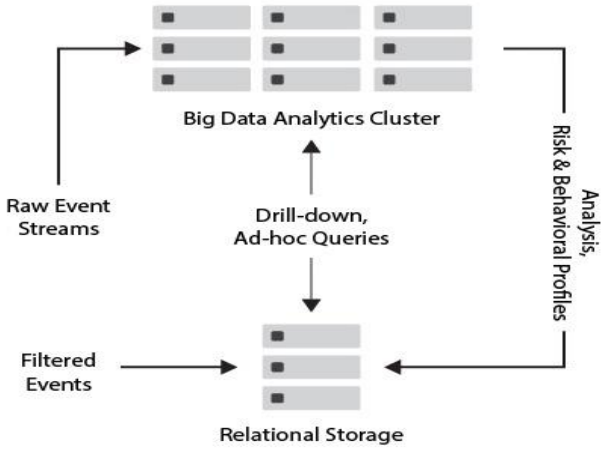

Fig.1: Data flow and Security in Bid Data

Solutions to the problem of security with privacy for big data require referring many research challenges and integrative approaches. The following are the challenges we encountered:

i) *Data mining solutions.* These are the heart of many big data environments; they find the patterns that suggest business strategies. For that very reason, it's particularly important to ensure they're secured against not just external threats, but insiders who abuse network privileges to obtain sensitive information – adding yet another layer of big data security issues.

ii) *Endpoints.* Security solutions that draw logs from endpoints will need to validate the authenticity of those endpoints, or the analysis isn't going to do much good.

iii) *Real-time security/compliance tools.* These generate a tremendous amount of information; the key is finding a way to ignore the false positives, so human talent can be focused on the true breaches.

iv) *Access controls.* Just as with enterprise IT as a whole, it's critically important to provide a system in which encrypted authentication/validation verifies that users are who they say they are, and determine who can see what.

v) *Storage.* In big data architecture, the data is usually stored on multiple tiers, depending on business needs for performance vs. cost. For instance, high-priority "hot" data will usually be stored on flash media. So locking down storage will mean creating a tier-conscious strategy.

vi) *Granular auditing* can help determine when missed attacks have occurred, what the consequences were, and what should be done to improve matters in the future. This in itself is a lot of data, and must be enabled and protected to be useful in addressing big data security issues.

vii) *Distributed frameworks.* Most big data implementations actually distribute huge processing jobs across many systems for faster analysis. Hadoop is a well-known instance of open source tech involved in this, and originally had no security of any sort. Distributed processing may mean less data processed by any one system, but it means a lot more systems where security issues can crop up.

viii) *Non-relational data stores.* Think NoSQL databases, which by themselves usually lack security (which is instead provided, sort of, via middleware).

ix) *Data provenance* primarily concerns metadata (data about data), which can be extremely helpful in determining where data came from, who accessed it, or what was done with it. Usually, this kind of data should be analyzed with exceptional speed to minimize the time in which a breach is active. Privileged users engaged in this type of activity must be thoroughly vetted and closely monitored to ensure they don't become their own big data security issues.

II. Technologies in Big Data Security:

The Big data technologies used in the security system can find the security threat earlier. For instance big data technology can detect abnormal behavior in the network, predict the attack behavior, and analyze the source of the attack.

The following are some of the techniques used in Big data security

- **Encryption:** Your encryption tools need to secure data in-transit and at-rest, and they need to do it across massive data volumes. Encryption also needs to operate on many different types of data, both user- and machine-generated. Encryption tools also need to work with different analytics toolsets and their output data, and on common big data storage formats including relational database management systems (RDBMS), non-relational databases like NoSQL, and specialized file systems such as Hadoop Distributed File System (HDFS).

- **Centralized Key Management:** Centralized key management has been a security best practice for many years. It applies just as strongly in big data environments, especially those with wide geographical distribution. Best practices include policy-driven automation, logging, on-demand key delivery, and abstracting key management from key usage.

- **User Access Control:** User access control may be the most basic network security tool, but many companies practice minimal control because the management overhead can be so high. This is dangerous enough at the network level, and can be disastrous for the big data platform. Strong user access control requires a policy-based approach that automates access based on user and role-based settings. Policy driven automation manages complex user control levels, such as multiple administrator settings that protect the big data platform against inside attack.

- **Intrusion Detection and Prevention:** Intrusion detection and prevention systems are security workhorses. This does not make them any less valuable to the big data platform. Big data's value and distributed architecture lends itself to intrusion attempts. IPS enables security admins to protect the big data platform from intrusion, and should an intrusion succeed, IDS quarantine the intrusion before it does significant damage.

- **Physical Security:** Don't ignore physical security. Build it in when you deploy your big data platform in your own data center, or carefully do due diligence around your cloud provider's data center security. Physical security systems can deny data center access to strangers or to staff members who have no business being in sensitive areas. Video surveillance and security logs will do the same.

III. How to improve Security in Big data:

The most appropriate technique to enhance Big Data security is by informing the role played by organizations who provide the platforms and systems to access the data, by providing timely blotches and security upgrades. Making security a high priority for systems puts the onus on system as well as application builders to be vigilant about security flaws, and also puts more eyeballs on the security front. With several vendors providing various solutions, you get a more precise defense against the security threats targeting Big Data applications.

Some recommendations that you can adopt to strengthen security are mentioned below:

- Do not focus all your attention on device security. Application security is of more importance.
- Keep devices and servers that contain sensitive information isolated.
- Introduce reactive and proactive protection.
- Attribute based encryption to protect sensitive information shared by third parties

  Secure open source software such as Hadoop

- Offer real-time security management Collaborating with other industry peers to create industry standards, head off government regulations, and to share best practices

- Maintain and monitor audit logs across all facets of the business

Finally, Big data presents more number of opportunities for businesses that go beyond just enhanced business intelligence. Big data offers the ability to increase cyber security itself. in order to benefit from the many opportunities big data presents, companies must enhance the responsibility and risk of protecting the data.

Conclusion:

This paper introduces impact to information security from two aspects of big data. Finally, end-users are just as responsible for protecting company data. Even though many companies use their big data platform to detect intrusion anomalies, that big data platform is just as vulnerable to malware and intrusion as any stored data. Organizations must ensure that all big data bases are immune to security threats and vulnerabilities. There is a large scope for research in the security issues in Big data.

REFERENCES:

1) Hu Kun, Liu Di, Liu Minghui. Research on Security Connotation and Response Strategies for Big Data[J]. Telecommunications Science, 2014(2):112-117,122.
2) Viktor Mayer-Schonberger, Kenneth Cukier. Big Data: A Revolution That Will Transform How We Live, Work and Think. Boston: Houghton Mifflin Harcourt , 2013
3) Mayer-Schönberger, V.; Cukier, K. Big Data: A Revolution that Will Transform How We Live, Work, and Think; Houghton Mifflin Harcourt: Boston, MA, USA, 2013.
4) Rijmenam, V. Think Bigger: Developing a Successful Big Data Strategy for Your Business; Amacom: New York, NY, USA, 2014.
5) Hrestak D, Picek S. Homomorphic Encryption in the Cloud [C] ‖ 2014 37th International Convention on Infor-mation and Communication Technology, Electronics and Micro electronics(MIPRO), 2014: 1400-1404.
6) Ulusoy, H.; Colombo, P.; Ferrari, E.; Kantarcioglu, M.; Pattuk, E. GuardMR: Fine-grained Security Policy Enforcement for MapReduce Systems. In Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security, Singapore, 14–17 April 2015; pp. 285–296
7) Cackett, D. Information Management and Big Data A Reference Architecture. Oracle: Redwood City, CA, USA, 2013.
8) Bessani A, Correia M, Quaresma B, et al. DEPSKY: Dependable and secure storage in a cloud-of clouds [C] //proc of the 6thConf on Computer System. New York: ACM, 2011:31-46
9) Dona Sarkar, Asoke Nath, "Big Data – A Pilot Study on Scope and Challenges", International Journal of Advance Research in Computer Science and Management Studies (IJARCSMS, ISSN: 2371-7782), Volume 2, Issue 12, Dec 31, Page: 9-19(2014).
10) Chen, M.; Mao, S.; Liu, Y. Big data: A survey. Mob. Netw. Appl. 2014, 19, 171–209.