International Conference on Recent Trends in Engineering Science and Technology (ICRTEST 2017) ISSN: 2321-8169
Volume: 5 Issue: 1(*Special Issue 21-22 January 2017*) 222 - 225

Survey on Big Data Information Security: The Current Security challenges by Big Data and Current Information Security Protection Method of Big Data.

Rahul Shahane
Asst. Professor WCEM, Nagpur
rahul.m.shahane@gmail.com

Abstract **-** With the fabulous development of information technology, big data application prompts the development of storage, network and computer field. It also brings new security problems. This security challenge caused by big data has attracted the attention of information security and industrial community domain. This paper summarizes the characteristics of big data information security, and focuses on conclusion of security problems under the big data field and the inspirations to the development of information security technology. Finally, this paper outlooks the future and trend of big data information security.

*****

1. Introduction

The development of the current big data is still faced with many problems especially security and privacy protection [1]. On the Internet People's behavior are known by Internet merchants [2], such as Amazon, DangDang know our reading habits, and Google, Baidu knows our search habits. A number of actual cases show that personal privacy will be exposed even after harmless data being collected [1]. In fact, the meaning of big data information security is much extensive. The threat person facing with is not only personal privacy leak, but also the protection of big data itself and knowledge acquired from it.

Currently many organizations realize the big data security issues and actively take actions on big data information security problems. In 2011, CSA formed a working group on big data to find solutions for data security and privacy issues. In this paper, based on the status of big data research, we analyzed the current security challenges by big data, and elaborated the current information security protection method of big datThe improvement of the current big data is still confronted with numerous issues particularly security and security assurance [1]. On the Internet People's conduct are known by Internet traders [2], for example, Amazon, DangDang know our perusing propensities, and Google, Baidu knows our hunt propensities. Various genuine cases demonstrate that individual protection will be uncovered even after safe data being gathered [1]. Indeed, the significance of big data security is much broad. The risk individual confronting with is close to home security spill, as well as the insurance of big data itself and information procured from it.

As of now numerous associations understand the big data security issues and effectively take activities on big data security issues. In 2011, CSA shaped a working gathering on big data to discover answers for data security and protection issues. In this paper, in light of the status of big data explore, we investigated the present security challenges by big data, and expounded the present data security assurance strategy for big data. a.

2. Threats of Big Data Security

Similarly as Gartner stated: "big data security is a vital fight"[3]. Today, big data has infiltrated into different enterprises, and has turned into a sort of creation variable which assumes a vital part. Later on it would be the most astounding purpose of the opposition. With the improvement of fast preparing and investigation innovation, the potential data it contained can rapidly catch the profitable data keeping in mind the end goal to give reference to basic leadership. Be that as it may, as big data setting off an influx of profitability and shopper excess, the test of data security is coming either.

2.1 Data Acquisition

The wellspring of big data is differing qualities. In this way, the initial step to prepare big data is to gather data from source and preprocess, with a specific end goal to give uniform top notch data set to the ensuing procedure. Subsequently, because of the immersion of data securing, huge data turn out to will probably be "found" as a delicate target, and be increasingly consideration. On one hand, big data implies the tremendous measures of data, as well as means more unpredictable and more delicate data. These data would draw in more potential aggressors, and turn into a more alluring target. Then again, with data gathered, the programmer could get more data in one fruitful assault, and lessen programmer's assault costs.

The classification of data alludes that as indicated by a predetermined prerequisites, data can't be revealed to unapproved people, elements or forms, or gave the attributes of its utilization. A lot of data accumulation incorporates an extensive number of undertakings working data, client data, individual security and a wide range of conduct records. The centralized stockpiling of these data builds the danger of data spillage, and not manhandled of these data additionally turns into a part of the individual security. There is no

certain definition to the proprietorship and appropriate to utilization of delicate data. What's more, numerous examination in view of vast data did not consider the individual security issues included either.

The uprightness of data alludes to every one of the assets which must be adjusted by approved individuals or with the type of approval. The reason for existing is to keep data from being altered with unapproved clients. Because of the openness of big data, during the time spent system transmission, data would be harmed, for example, programmers blocked, intrusion, altering and imitation. Encryption innovation has understood the data classification prerequisites and additionally ensuring data respectability. Be that as it may, encryption can't tackle the greater part of the security issues.

2.2 Storage of Data

The development of system society makes the stage and channel of asset sharing and data trade for the big data in the field of different ventures. Organize society in view of cloud calculation gives an open situation to big data. Organize get to and data stream gives the premise of fast versatility push of the assets and the customized benefit. As of late, from the chain response of client record data being stolen on the Internet, it can be seen that big data will probably draw in programmers, and once being assaulted, the volume of stolen data is tremendous.

Before big data, data stockpiling is isolated into social database and document server. What's more, in current big data, differing qualities of data sort makes us ill-equipped. For over 80% of the unstructured data, NoSQL has the benefits of scalability and accessibility and gives a preparatory answer for big data stockpiling. However, NoSQL still exist the accompanying issues: one is that with respect to the strict get to control and security administration of SQL innovation; besides, in spite of the fact that NoSQL programming pick up involvement from the conventional data stockpiling, NoSQL still exist a wide range of hole.

2.3 Data Mining

With the improvement of PC system innovation and computerized reasoning, arrange hardware and data mining application framework is increasingly generally used,to give advantageous to big data programmed effective gathering and clever element investigation. From one perspective, big data itself exits spill. Big data itself can be a bearer of reasonable assault. Infections and noxious programming code covered up in expansive data is elusive. Then again, the strategy of assault progresses. In the meantime of the big data innovation, for example, data mining and data examination picking up esteem data, the assailant utilizing these big data innovation either, similarly as the two after aspects.A extensive number of actualities demonstrate that inability to appropriately handle big data will bring about incredible infringement to clients' protection. As per the diverse substance should be ensured, security insurance can be further partitioned into area security assurance, mysterious identifier security, unknown associations et cetera. The danger People confronted with is close to home security spillage, as well as expectation and conduct of the general population in light of big data. Indeed, mysterious insurance can't secure protection exceptionally well. Investigate on informal organization additionally demonstrates that client properties can be found from the gathering features[4].Currently accumulation, stockpiling, administration and utilization of client data is shy of particular, and regulation[5][6]. Clients can't decide their security data use. In business situation, client ought to have the privilege to choose how their data be utilized, and understand clients' controllable security assurance. A general view about big data Is: data itself can tell everything, the data itself is a reality [7]. Truth be told, if not precisely screened, the data can swindle individuals, similarly as individuals can once in a while be misdirected by their eyes.one of the danger of big data believability is fake or purposely producing data, and the wrong data frequently prompt to wrong conclusions. On the off chance that data application situations is plainly, somebody could intentionally producing data, and make a "false fragrance", to initiated experts reach the conclusion that was on their side. As a result of false data regularly covered up in a great deal of data, it make difficult to distinguish validness of data, in order to make wrong judgment. Because of the creation and proliferation of false data in system group is turning out to be increasingly simple, its belongings ought not be thought little of and essentially utilizing data security innovation to recognize the realness of all sources is outlandish.

3. Reason Analysis

With the development and progress of information technology, the security of sensitive data is facing with unprecedented challenges. This is a serious impediment to the spread of new applications. Safety problems mainly displays in the following respects.

3.1 Lack of world recognized laws and regulations for data security and privacy protection

Privacy is not a new problem, but with the development of network technology, privacy has also been gradually amplifier, especially e-commerce (Electronic Commerce, EC) privacy issues, which has become one of the most important issues in the network economy. However, for existing privacy regulations and policies, there are still somewhere to improve [8].

First of all, because of the different of specifications and law cultural of different countries, privacy law only applies to certain territorial limits which impact limited on the global network. Secondly, many countries are not willing to

weaken the economic rise of the Internet brought by the economic boom, so they try to avoid joint intervention with other countries. Moreover, because of the long-term and stability of the law, legal measures can't meet the needs of the rapid development of the Internet.

3.2 The cloud infrastructure has not a uniform and reliable authentication, which cannot prove it's credible

With the rapid development of cloud storage, more and more users choose to use the cloud storage to store information. The key characteristic of cloud storage is stored as a service. Users can upload their data to the public API in the cloud. But due to the loss of the users' absolute control of data, some hidden danger of data security arises: (1) Rely on customer management of the certificate too much. (2) The granularity of data storage protection is not enough.(3) Do not consider the perfect data sharing requirements.(4) The lack of an effective regulatory pathway to ensure that the storage of data would not be lost, leak, or abuse.

Nowadays, many cloud storage service provider provides cloud storage services with a very low price or even free. Because of the loss of control of data caused by cloud storage, user is difficult to check the data integrity and confidentiality in cloud storage environment. In the worst case data is stored in the unknown "corner" of service pool, which lead to the poor cloud storage environment disaster resistance [10].

3.3 Lacking of Creditable Authentication in Cloud Computing Service

While bringing convenience, there are problems in cloud computing, among which security issues are the most critical ones and the main factors enterprises users worry about. CSA(Cloud Security Alliance) puts forward the risks cloud computing faces, including data center security, event responding security, application security, key management security, authentication and access control security, virtualization layer security, backup for disaster recovery and business alignment. At the same time, people have realized there are differences between cloud computing security and traditional security. In traditional IT systems, the owner and the user of the fundamental facility are identical. When it comes to cloud computing, CSP(Cloud Service Provider) owns the fundamental facility which offers computing service, while users have the access to it. This makes adversarial relationship between CSP and users. Cloud computing is a trusted model in its nature, CSPs prove the creditability of its service and users build up confidences in it through CSPs' proof[12].

4. Data Security Protection Technique

Key technologies in Security protection fields are in great demands to face the security challenges. In this section, we introduce important relevant field.

4.1 Individual User

As with individual users' information in big data environment, the core and basic techniques to provide privacy protection are still in developing period. Take typical K-anonymity scheme as an example, its early version [13] and optimized version divide quasi-identifiers into groups through tuple generalization [14] and restraining method. When an equivalence class has identical value on some sensitive attribute, attackers are able to confirm its value. In response to this issue, researchers proposed 1-diversity [15] anonymity.

Current edge anonymity schemes are mainly based on adding and deleting of the edges. Edge anonymity can be effectively achieved by adding, deleting and exchanging edges randomly [16]. There are problems in such methods that noises randomly added are exiguity, and protections to anonymous edges are insufficient. An important method is to perform division and aggregation operations to super nodes such as node aggregation based anonymous method, genetic arithmetic based method and simulated annealing method based method.

4.2 Internet Enterprise

Information security is critical important for Internet enterprises. System security adopts techniques such as redundancy, network separation, access control, authentication and encryption [18]. Security issues are caused by openness, boundless, freedom of the networks, the key to solve such issues are making network free from them and turning network into controllable, manageable inner system. As network system is the foundation of application system, network security becomes principal issue. Ways to solve network security issues are network redundancy, system separation and access control

4.3 Cloud Service Provider

CSPs provide following measures to prevent security issues in cloud environment. In order to prevent CSPs from peeping users' data and program, separating power and hierarchical management are needed to control access to data in cloud. Provide different authority in accessing data to service provider and enterprise to ensure data security. Enterprise should have total authority and limit authority to CSP.

In cloud computing environment data separation mechanism prevents illegal access to data, however, we should take care of data leakage from CSPs. Mature techniques as symmetrical encryption, public key encryption are available to encrypt data and then upload data to cloud environment. In cloud environment data division is often used with data encryption i.e. encrypted data are scattered in user end and

spread in several different clouds. In the way, any CSP is not able to gain complete data.

5. Conclusion and Prospect

Information security in big data environment is a promising fields in information security. This paper introduces impact to information security from two aspects of big data and cloud computing. In general, improving system efficiency and provide general cloud storage functions on premise to ensure user data and access authority are the research direction of future safe cloud computing. At present, more things need to be done in cryptograph searching and reduplicate data removing.

After all, there is an urgent need of improved solutions concerning the users to control the use of their data and more research should be done in this field and there is also a need for more robust approaches in key management limitation, which could extend traditional approaches to Cloud computing.

REFERENCES

[1] Viktor Mayer-Schonberger, Kenneth Cukier. Big Data: A Revolution That Will Transform How We Live, Work and Think. Boston: Houghton Mifflin Harcourt , 2013
[2] Meng Xiao-Feng, Ci Xiang. Big Data Management: Concepts, Techniques and Challenges. Journal of Computer Research and Development, 2013, 50(1): 146-169 (in Chinese)
[3] Chen Mingqi, Jiang He. USA Information Network Security New Strategy Analysis in Big Data [J]. Information Network Security. 2012(8):32—35
[4] Narayanan A, Shmatikov V. How to break anonymity of the Netflix prize dataset. ArXiv Computer Science e-prints, 2006, arXiv:cs/0610105: 1-10
[5] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. Exploiting geographical influence for collaborative point-of-interest recommendation.//Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval(SIGIR'11), Beijing, China, 2011: 325-334
[6] Goel S., Hofman J.M., Lahaie S., Pennock D.M. and Watts D.J.. Predicting consumer behavior with Web search. National Academy of Sciences, 2010, 7 (41):17486– 17490
[7] http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
[8] Study Finds Web Sites Prying Less: Shift May Reflect Consumer Concerns[EB/OL]. http://www CNN.com, 2002-03-18
[9] A survey of data disclosing in 2010 by Verizon[EB/OL].[2012-05-10].
[10] 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
[11] Sweeney L..k-anonymity: a model for protecting privacy. InternationalJournal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10 (5): 557-570
[12] Sweeney L..k-Anonymity: Achieving k-Anonymity Privacy Protection using Generalization and Suppression.
[13] AshwinMachanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1):1-52
[14] Ying X. and Wu X.. Randomizing social networks: a spectrum preserving approach. //Proceedings of the SIAM International Conference on Data Mining (SDM'08), Georgia, USA, 2008: 739-750
[15] Lei Zou, Lei Chen and M. Tamer zsu. k-automorphism: a general framework for privacy preserving network publication. // Proceedings of the 35th International Conference on Very Large Data Bases (VLDB'2009), Lyon, France, 2009: 946-957.