A Survey on Security Issues in Big Data of Ubiquitous Network

Yuehua Huo[1], Yanyu Sun[2], Weiqiang Fan[2], Xinzhou Cheng[3], Dong Li[1], and Yinlong Liu[4(✉)]

[1] Center of Modern Education Technology, China University of Mining and Technology, Beijing 100083, China
[2] School of Mechatronics and Information Engineering, China University of Mining and Technology, Beijing 100083, China
[3] China Unicom Network Technology Research Institute, Beijing 100048, China
[4] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
huoyh@cumtb.edu.cn, chinahuo007@163.com

Abstract. In the ubiquitous network, these many devices generate a huge amount of data. The sharp increase of the diverse devices has led to the various heterogeneous data types and the booming growth of the data, that is, the ubiquitous network will bring big data. Due to the heterogeneity and ubiquitous nature of ubiquitous networks, the big data of ubiquitous network faces a broad and growing range of security threats and challenges. Based on the basic knowledge of the big data of ubiquitous network, this paper analyzes the security threats in the process of data collection, storage, processing and application. System elaborated the big data of ubiquitous network encryption and integrity protection, access control, privacy protection, attack detection, security technology research status. The research direction and future development direction are also discussed in this paper.

Keywords: Ubiquitous network · Big data · Information security · Intelligent

1 Introduction

With the rapid development of network technology, both at home and abroad pay more and more attention to the study of ubiquitous network, increasing investment, making the network is widely used in social and economic sustainable development and national development strategy, extensive penetration as part of the network in all walks of life, become strategy core of each country. A large number of sensor networks, RFID [1] and other widely used, so that everything can be data, people realize that the nature of the world is the data, and the data will gradually completely change the future of the world.

Nowadays, the ubiquitous network plays a central role in the national strategic development, triggered unprecedented national investment and the further research of many scholars, and aiming at security problems exist in the process of its development, put forward the solution. The United States proposed broadband urban planning [2],

K. Long et al. (Eds.): 5GWN 2017, LNICST 211, pp. 89–98, 2018.
https://doi.org/10.1007/978-3-319-72823-0_10

the plan for the arrival of the network to pave the way at the same time, will be in the United States to set up a wide range of urban broadband network. The South Korean government also proposed the "U-Korea" strategic plan [3], ubiquitous sensor network, broadband network integration and IPv6 core network is included in the investment plan of infrastructure construction, the construction of a part of achievements in the promotion, in city management, medical treatment, military has been widely used. In China, some research on the ubiquitous network is gradually carried out, such as "perception of China", "U-Beijing" and "U-Qingdao" [4] is one of the typical representative.

With the extensive application of 4G network, 5G [5, 6] as a new wireless access technology is gradually into people's lives, it has more powerful features, will be a true sense of the integration of the network [5, 7]. In this paper, according to the characteristics of ubiquitous network data, generally reveal: safety issues faced in the ubiquitous network of different network structure and big data in different data processing stages, and the corresponding solutions and technology.

2 Big Data of Ubiquitous Network

2.1 The Concept and Characteristics of Big Data of Ubiquitous Network

Ubiquitous network is wider than the web of things, more extensive content. It is generally believed that the sensor network, the sensor node is composed of a large number of networking, personal and social network etc., which refers to at any time (Anytime) of any location (Anywhere) of any person (Anyone) and (Anything) any communication [8].

Big data [9], also known as mass data, refers to the size of the data involved to a huge scale cannot be artificial or computer, within a reasonable time to achieve interception, management, processing, and sorting into human interpretations of the form of information. Ubiquitous network is the root of big data, has a close connection, the network application can promote the research of big data. Big data is the resources and wealth, big data analysis makes the decision more scientific, more intelligent terminals, will promote the ubiquitous network more intelligent and more widely used to promote its further rapid development. Big data generated by the ubiquitous network, due to the large amount of data and fast generation and often change characteristics [10], if the encryption and decryption technology and traditional hash algorithm [11] to achieve the integrity of the identification and authentication of data, the efficiency is low and it is hard to meet the real-time, consistency and synchronization so on. The ubiquitous network is still in the stage of the Internet of things, big data mining, analysis and processing methods are not mature, at the same time, we must face many security problems of big data, in the ubiquitous network, and more complex than ever. Therefore, it is necessary for the ubiquitous network security key problems on the big data of research, for the application of ubiquitous network and big data cleared.

2.2 Security Challenges for Big Data of Ubiquitous Networks

The doubling of the amount of data challenges the ability to store data; the traditional database lacks strong scalability and better system availability, and cannot effectively store unstructured and semi-structured data such as video and audio. Cloud storage in the application of big data to bring development opportunities at the same time, also brought a huge security risk. First of all, the user data storage and security entirely by the cloud computing provider is responsible for the data is transparent to the provider. Second, in the cloud storage, the user's data is stored in the Internet server, an increase of unauthorized access to security issues.

With the increasing scale of data, the time of analysis and processing is correspondingly longer and longer, and the time requirement of information processing is getting higher and higher under big data condition. The processing of big data requires more simple and effective artificial intelligence algorithm and new problem solving method. Discover new methods, predict predictability of threats, and detect and judge unknown threats. The Prism Plan is a case of a large data technology for security analysis. The basic principle is to discover potentially dangerous situations by collecting various types of data from various countries, using safety-related data and threatening security analysis. Identify the threat before it occurs.

3 Security Problems in Big Data of Ubiquitous Network

3.1 Security Threats in the Acquisition Process

Most of the current privacy protection using location, identifier, anonymous connection and other methods. But in practice, in addition to face the threat of personal privacy, people through the vast amounts of data, analysis of people's living state and behavior, prediction, through anonymous protection in today's era of big data is not fully achieve the purpose of privacy protection. On the one hand, due to the current lack of user data collection, storage, management and use of norms, lack of supervision and effective management, mainly rely on self-discipline, cannot guarantee the user privacy information purposes; On the other hand, the data traceability and intelligent analysis technology in the big data age can analyze the sensitive information such as data sources and obtain the privacy data [12].

3.2 Security Threats to Stored Procedures

The arrival of cloud storage, big data applications with more development opportunities, but also with a huge security risk. Firstly, cloud computing providers are fully responsible for the storage and security of user data, the data provider is transparent. Secondly, in the cloud storage, the user's data is stored in the Internet server, an increase of unauthorized access to security issues. As shown in Fig. 1, it analyzed cloud storage ubiquitous network data security; research on big data security of ubiquitous network based on cloud storage includes data integrity protection and access control.

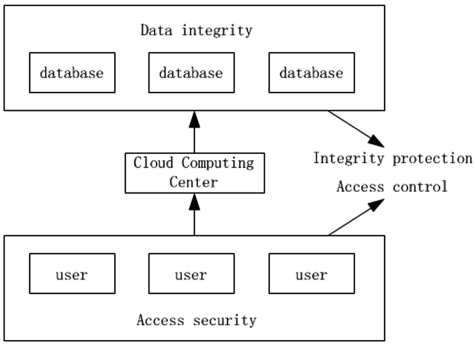

Fig. 1. Big data security analysis of cloud storage

3.3 Process Security Threats

In terms of pattern matching algorithms. In English character environment, classic DFSA application efficiency is very high, if utilized directly match characters, the construction of a Chinese character storage expansion problems when a Hash table. To this end, Shen Zhou [13], Wong [14], Gao Peng [15], respectively, put forward different algorithms. In text content analysis, the text content analysis in cloud content monitoring, the depth of the user identification of suspicious information, but also can find the current flow of hot information.

3.4 Application Process Security Threats

In the ubiquitous network environment, the content of the network is richer, and the flooding of reactionary, pornographic, violent and other undesirable contents has become a problem to be solved urgently [16]. How to ensure the legitimacy and health of data content has become a big data analysis And big data security research areas of hot issues, and cause more and more attention.

4 Security Technology in Big Data of Ubiquitous Network

4.1 Data Encryption

Research on Cloud Storage Key Management and Auditing Proxy Mechanism, Cloud Storage Service (Cloud Storage Service). In addition to the general personal data storage, but also to allow enterprises or data owners (Data Owner, DO) to facilitate the sharing of data services. But when DO wants to share specific data, especially secret information to a specific group or individual, the data access control and key management information security issues need to be considered. Zhao [17] proposed a classification proxy re-encryption technique, which enables data distributors to implement fine-grained categorization control of ciphertext delegation. Wu [18] gives a proxy-re-encryption algorithm with no certificate and an identity-based key escrow protocol. Liang [19] studied the identity-based revocable proxy re-encryption

mechanism. Fang [20] proposed an anonymous conditional proxy re-encryption scheme and a fuzzy condition proxy re-encryption scheme [21] to improve the performance of the algorithm.

4.2 Integrity Protection

Integrity protection consists mainly of two components, POR and PDP. Literature [22] builds a hierarchical architecture that provides high availability and integrity protection for data by combining two-dimensional RS coding with a challenge-response mechanism. Literature [23] gives a method to determine the integrity of the data. That is, the authentication element is generated from each file data block, in the challenge of the challenge - response mechanism, adopt the method of pseudo random, draw a small amount of data blocks, judgment data integrity can be achieved through authentication of authentication element. Literature [24] proposed method based on the pattern of challenge - response "Data type to hold Proof" (Proof of Data Possession, PDP), namely in outsourcing Data, detect is greater than a certain percentage of Data corruption. Literature [25] to solve files in multiple servers distributed storage situation, proposed the CPDP (Cooperative Provable Data Possession). The proposed scheme utilizes a homomorphic authentication response to combine responses from different cloud servers into a single response message. Currently, the big data integrity of the cloud to verify mainly rely on a third party to complete. According to whether to allow the restoration of the original data, the current data integrity verification protocol can be divided into two categories: only to verify data integrity of the PDP protocol and allow recovery of data POR protocol. Data integrity verification of the general process shown in Fig. 2.

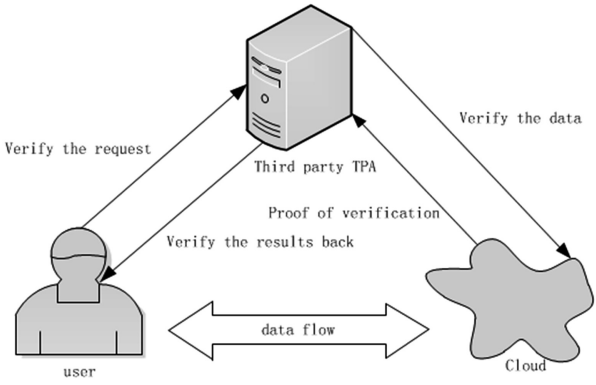

Fig. 2. Data integrity verification of the general process

4.3 Access Control

In the aspect of access control, the decryption rules are embedded in the encryption algorithm through the algorithms of KP-ABE and CP-ABE [26] (ABE), which can avoid the frequent occurrence of ciphertext access control Key distribution cost.

However, when the control strategy changes dynamically, the data owner (the Data Owner) is required to re-encrypt the data. In literature [27], Cloud service providers (cloud service provider, CSP) agent in order to reach the goal of user revocation, the cipher text from an encrypted access structure to another encrypted access structure. Undo unit has a certain limitation of the scheme, a single user cannot complete the revocation, only is a set of properties. Literature [28] cipher text access control scheme is put forward, to support fine-grained access control strategy, but when undo attribute and user privileges, DO the calculation of the price has a linear relation with the data file size. In summary, for the purpose of profit-making CSP, cannot effectively guarantee security.

4.4 Privacy Protection

Ubiquitous network game model of big data privacy protection required game basic research. Based on the previous research, we have fully understood the privacy requirements and the basic technical means at all levels. It is proposed that participants, strategy (attack and defense strategy, behavior rule set), information collection (common knowledge) Action and action sequences, utility (payment) functions and other basic elements of the game to abstract and extract. Based on the ubiquitous network heterogeneity, diversity of unstructured characteristics of big data, both offensive and defensive behavior, the environment, the diversity of information, the incomplete information dynamic game, combined with Bayesian equilibrium subgame perfect Nash equilibrium, and the Bayesian inference method, the optimal privacy policy.

4.5 Attack Detection

DDoS attack detection has been a very important research topic in network intrusion detection. In recent years, a variety of different types of detection platforms and algorithms have been proposed to solve DDoS attacks, and many technical problems have been solved and many achievements have been made. Among them, the information entropy power system is described as random degree of effective index, gradually become the research focus in the network anomaly traffic analysis and detection. We based on the theory of information entropy and chaos phase space reconstruction theory, through the in-depth study of typical DDoS attack scenario, put forward the following two methods for intrusion detection: Detection Method Based on Tsalli Entropy and Lyapunov Exponent. First, the entropy of the source IP address and the destination IP address in the network traffic packet is counted to reflect the randomness of the source IP address and the concentration of the destination IP address in the response attack. On the basis, the Lyapunov exponent in chaos theory is introduced to calculate the degree of separation between the source IP address entropy and the destination IP address entropy, and the attack traffic is distinguished from the normal traffic. Detection Method Based on Markov Chain and Kolmogorov Entropy. By analyzing the degree of protocol dependency of various network packets in attack traffic, the Markov chain model is established, and the complex Markov chain model is simplified by entropy rate sequence, and the chaotic phase space reconstruction is introduced. The data is reconstructed by using the geometric invariant feature

Kolmogorov entropy of the phase space to reflect the inherent chaos of the network traffic model to achieve the purpose of detecting the attack. The above research has basically reflected the advantages of information entropy theory in the analysis of network anomalies and network intrusion, so the above research provides a theoretical basis for big data analysis.

5 Prospect of Security Technology in Big Data of Ubiquitous Network

5.1 Technical Level

On the technical level, the subsystems are independent and closely integrated with each other. Each part belongs to an indivisible whole, it will be the perfect fusion of the subsystems together. In general can be divided into the ubiquitous network, the security of data architecture and technical analysis, related research in information industry and the practical application of hercynian information industry, and a few kinds big data pretreatment method in between data key security study four blocks of the model.

By analyzing the security requirements at different levels of the ubiquitous network and different stages of big data, the preprocessing technology of big data in the ubiquitous network, the application technology in the information industry and the specific key security models are studied. In turn, these three aspects of research are the key to promoting the development and application of large-scale data security technologies.

The big data preprocessing technology is the foundation of the research on the concrete data security model of the big data network. It can also be used in the actual information industry to solve the difficult problems in the application of the information industry and promote its further healthy development.

As shown in Fig. 3, the big data processing from front to back into data collection, data transmission, data analysis, data analysis at different stages. Then the security problems faced by different stages are analyzed. The security techniques and methods that may be adopted in each stage are determined. For example, the integrity of the big data transmission stage can be verified by digital watermarking technique.

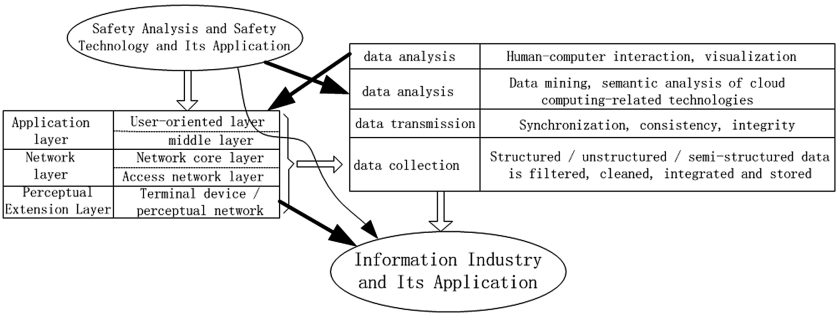

Fig. 3. Ubiquitous network and big data architecture, security requirements and technical analysis and application

5.2 Management Level

In the case of big data, a large number of users need to implement the rights management, and the user specific authority requirements unknown. In the face of unknown large amounts of data and users, it is very difficult to set the role in advance. Construction of important sensitive equipment and carrier management and monitoring system architecture and standards, and based on RFID and other Internet of things technology, from the production to the retirement of the whole life cycle management and control system to complete the relevant standards development, system development work.

First of all, the establishment of content monitoring knowledge model; then, based on "people, behavior, content," the sensitive data-aware platform, the generalized network of information data sets for preliminary screening, reduce the amount of redundant information; Finally, the establishment of dynamic awareness of the network, the suspicious, potential portals, forums and users real-time tracking analysis, so that quickly found, in time to stop. The practical application of big data in the information industry can lead to further improvement of big data analysis methods and help to improve the management level. The security model can be applied directly to the real information industry.

6 Conclusion

According to the characteristics of big data, we fully analyze the security problems and research status, and on this basis, we look at the research direction at the technical and management level. Therefore, the research on the big data and security in the network has far-reaching theoretical significance and practical application value. The research on it is not only necessary in the construction of the theoretical system, the key technology and the improvement of the security problems, Mode, laws and regulations and other aspects of exploration, but also need to carry out theoretical validation and demonstration of its application, it is better for the community to provide services. Analysis and utilization of big data in the ubiquitous network can make the future more common in the network will be more intelligent.

References

1. Van den Elzen, S.V., Wijk, J.J.: Multivariate network exploration and presentation: from detail to overview via selections and aggregations. Vis. Comput. Graph. **20**, 2310–2319 (2014)
2. Xu, X.Q., He, G.N., Zhang, S.Q., Chen, Y., Xu, S.G.: On functionality separation for green mobile networks: concept study over LTE. Commun. Mag. **51**, 82–90 (2013)
3. Kushiro, N., Higuma, T., Nakata, M., Kubota, H., Sato, K.: Practical solution for constructing ubiquitous network in building and home control system consumer electronics. IEEE Trans. Consum. Electron. **53**, 1387–1392 (2007)
4. Saito, H., Kagami, O., Umehira, M., Kado, Y.: Wide area ubiquitous network: the network operator's view of a sensor network. Commun. Mag. **46**, 112–120 (2008)

5. Zhang, H., Dong, Y., Cheng, J., Hossain, M., Leung, V.C.M.: Fronthauling for 5G LTE-U ultra dense cloud small cell networks. IEEE Wirel. Commun. **23**, 48–53 (2016)
6. Xu, L., Chen, Y., Chai, K.K.: Cooperative mobility load balancing in relay cellular networks. In: 2013 IEEE/CIC International Conference on Communications in China, Xi'an, pp. 141–146. IEEE Press (2013)
7. Xu, L., Cheng, X., Liu, Y.: Mobility load balancing aware radio resource allocation scheme for LTE-advanced cellular networks. In: IEEE International Conference on Communication Technology, Hangzhou, pp. 806–812. IEEE Press (2015)
8. Chung, Y.F., Tsai, M.Y., Wu, E.P., Chiang, D.L., Lee, C.C., Hsiao, T.C., Chen, T.S.: Bed site health care video-phone system. Appl. Mech. Mater. **284–287**, 1636–1641 (2013)
9. Hashem, I.A.T., Yaqoob, I., Anuar, N.B.: The rise of "big data" on cloud computing: review and open research issues. Inf. Syst. **47**, 98–115 (2015)
10. Guo, F., Wang, J.M., Li, D.Y.: Finger printing relational databases. In: ACM Symposium on Applied Computing, Dijon, pp. 487–492. ACM Press (2006)
11. Jiang, C.X., Sun, X.M., Yi, Y.Q., Yang, H.F.: Research on database public watermarking algorithm based on JADE algorithm. J. Syst. Simul. **18**, 1781–1785 (2006)
12. Zhang, H., Xing, H., Cheng, J., Nallanathan, A., Leung, V.: Secure resource allocation for OFDMA two-way relay wireless sensor networks without and with cooperative jamming. IEEE Trans. Ind. Inf. **12**, 1714–1725 (2016)
13. Shen, C., Wang, Y.C., Liu, G.S.: Improved algorithm of multi-pattern matching for Chinese string. Acta Agron. Sin. **21**, 27–32 (2002)
14. Zhao, J., Zheng, W., Wen, X., Zhang, H., Lu, Z., Jing, W.: Research on the resource allocation of OFDMA relay network based on secrecy ratio. J. Electron. Inf. Technol. **36**, 2816–2821 (2014)
15. Zhao, J., Lu, Z., Wen, X., Zhang, H., He, S., Jing, W.: Resource management based on security satisfaction ratio with fairness-aware in two-way relay networks. Int. J. Distrib. Sens. Netw. **2015**, 11 (2015)
16. Wang, T., An, B.Y., Peng, Z., Zhang, G.L., Zhong, B.N.: Blue card-green communications for exchanging information of mobile users. J. Comput. Inf. Syst. **10**, 8153–8160 (2014)
17. Schell, R.: Security — a big question for big data. In: 2013 IEEE International Conference on Big Data, Santa Clara, p. 5. IEEE Press (2013)
18. Bowers, K.D., Luels, A., Oprea, A.: Hail: a high-availability and integrity layer for cloud storage. In: 16th ACM Conference on Computer and Communications Security, Chicago, pp. 187–198. ACM Press (2009)
19. Zhao, J., Feng, D.G., Yang, L.: CCA-secure type-based proxyre-encryption without pairings. Acta Electron. Sin. **39**, 2513–2519 (2011)
20. Wu, X.X., Li, X.U., Zhang, X.W.: A certificateless proxyre-encryption scheme for cloud-based data sharing. In: 18th ACM Conference on Computer and Communications Security, Chicago, pp. 869–871. ACM Press (2011)
21. Liang, K., Liu, J.K., Wong, D.S., Susilo, W.: An efficient cloud-based revocable identity-based proxy re-encryption scheme for public clouds data sharing. In: Kutyłowski, M., Vaidya, J. (eds.) ESORICS 2014, Part I. LNCS, vol. 8712, pp. 257–272. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11203-9_15
22. Weng, J., Deng, R.H., Ding, X.H.: Conditional proxyre-encryption secure against chosen-ciphertext attack. In: ACM ASIACCS 2009, pp. 322–332 (2009)
23. Fang, L.M., Wang, J.D., Ge, C.P.: Fuzzy conditional proxyre-encryption. Sci. China Inf. Sci. **56**, 1–13 (2015)
24. Han, H., Wen, Y.G., Chua, T.S., Li, X.L.: Toward scalable systems for big data analytics: a technology tutorial. IEEE Access **2**, 652–687 (2014)

25. Yang, C., Liu, C., Nepal, S., Chen, J.: A time efficient approach for detecting errors in big sensor data on cloud. IEEE Trans. Parallel Distrib. Syst. **26**, 329–339 (2015)
26. Zhang, G.L., Wang, Z.N., Du, J.X., Wang, T., Jiang, Z.N.: A generalized visual aid system for teleoperation applied to satellite servicing. Int. J. Adv. Robot. Syst. **11**, 1–7 (2014)
27. Yi, X.M., Liu, F.M., Liu, J.C., Hai, J.: Building a network highway for big data: architecture and challenges. IEEE Netw. **28**, 5–13 (2014)
28. Liang, X., Cao, Z., Lin, H.: Attribute based proxy re-encryption with delegating capabilities. In: 4th International Symposium on Information, Computer and Communications Security, pp. 276–286. ACM Press, New York (2009)