Security mechanism building for big data based on Trusted Computing

Lv Hai-yan, Zhou Li-Jun, Zhao Yuan
Naval Aviation University, Shandong Yantai, 264000, China
xwc112329@126.com

Abstract: In view of the security risks faced by large data in the process of collection, storage, transmission and use, the security threats faced by large data in these processes are analyzed. The characteristics of trusted computing and its advantages in information security protection are analyzed. On this basis, a secure storage, transmission and authentication mechanism of large data based on trusted computing is proposed to achieve active network defense and ensure that large data can be controlled, trusted and manageable in the process of storage, transmission and authentication.

1. Introduction
In the era of large data, the original means of information resource processing have not adapted to the rapidly increasing data magnitude. The original computing environment also has changed with the development of data mining, association analysis and other large data technologies [1]. Centered by offering variable service of development and application of big data information technology needs credibility as the foundation and development premise, and it needs more credible measurement, computation and control. Trusted computing can achieve trusted immunity, active protection, and ensure that big data is credible, controllable and manageable.

2. Problems of big data security
Rapid development of Internet technology has indeed brought many conveniences to people's lives, but the development of everything has its two sides. From the "prism gate" incident to Crip's "safety door" incident, the problems of information security in the era of big data are noticeable [2].

2.1. Security threats of data storage
The collection of massive heterogeneous data makes the big data analysis platform must have the characteristics of large capacity, distributed, and cluster management, and can support the processing of large-scale structured, semi-structured and unstructured data. Because of the complexity and diversity of big data storage, data storage management and security protection measures are difficult to avoid, which can easily cause data leakage and tampering. The centralized storage of data has also attracted the attention of hackers, and it is necessary to enhance the security level to prevent hackers from attacking.

2.2. Data processing problems
If data processing is erroneous, it will lead the masses to the wrong direction and make wrong decisions. However, due to the large scale of the data, it is impossible to screen out the wrong data well, and the lack of effective supervision and inspection in the process of data transmission, resulting
in serious deviation and large direction error, so that the fundamental purpose can not be reached. Therefore, the correctness and accuracy of data processing is very important.

2.3. Privacy threats
Because the big data has been applied to all walks of life, and the information contained in the big data is varied, there is no lack of personal privacy related information, such as geographical location, mobile phone number, and so on. Because of improper data processing, sensitive information is not effectively protected, so that personal privacy information is published in public [3-7]. Some criminals can extend their customers’ information to a larger scope by collecting and analyzing the basic information of customers. At present, our country still lacks the relevant regulations of the user information management in the era of big data, and there is no better supervision system, and more users usually lack of self-protection consciousness, all the above problems have caused a lot of loss brought by information disclosure.

3. Trusted computing and information security protection
Shen Chang-xiang, the academician of the Chinese Academy of engineering, proposed that the trusted computing refers to the provide security protection at the time of calculation operation, so that the calculation results are always the same as expected, and the whole process can be measured and controlled without disturbed[4]. Credibility includes many aspects, such as correctness, reliability, security, availability, efficiency and so on. It can be popularly understood as: trustworthiness = reliability + security. The overall goal of trusted computing is to improve the security of the computer system, and the main goal of this stage is to ensure the integrity, the secure storage and the remote proof of the credibility of the platform.

The key three components of trusted computing are trusted platform module, trusted storage and trusted network connection.

1) Trusted platform module
TPM is the core of the trusted computing platform. It is a small SoC chip system including cryptographic computing components and storage components, and it consists of the central processor, memory, cipher operation processor, random number generator and I/O components. The TPM mainly complete the storage of the trusted metric, the report of the trusted measure, the key generation, and the encryption. And signature, data security storage and other functions, as well as system startup, user authentication, system monitoring, encryption and signature, and other security and trust functions.

2) Trusted storage
TCG's trusted storage provides a manageable, enterprise level method for full disk encryption. The drive used is a self-encrypted drive which simplifies the enterprise’s encryption process for handling sensitive data, because all data, applications and drivers are encrypted inside the drive. TPM implements hardware based encryption, trusted storage implements encryption and authentication functions by self encrypting drive.

3) Trusted authentication
Trust Network Connect (TNC) is an industrial standard architecture for providing network security and network security access. According to TNC standards, administrators can control network access according to user’s identity and device status, monitor the network operation, and immediately respond to trusted computing group once a problem occurs [8]. IFMAP architecture has been developed for the multi tenant characteristics of cloud computing, which is a platform for information interaction and sharing between traditional network security devices (such as firewalls, intrusion detection, traffic control, etc.) and trusted network connection components.

4. Large data security solution based on Trusted Computing

4.1. Trusted security storage
In view of the 3 main problems, such as storage environment, storage mode and storage protection of large data security storage, a secure storage scheme for large data is designed by using trusted computing technology, as shown in Figure 1. The trusted platform module TPM is used as the root of trust, and the security features provided by trusted computing can improve the storage security of big data. Among them, the security and reliability of storage environment is the premise of large data storage. The TPM built in the computer system is a trusted trust root, and the integrity measurement mechanism is used to judge whether the large data storage environment is attacked and tampered. Integrity measurement is the collection of the current state of large data system platform, and its measurement process is the establishment process of trust chain. The large data system platform starts from the trust root, first measures the system BIOS, stores the measurement results in the corresponding system platform state register PCR (Platform State Register) by extending the measurement results, and then continues to measure the operating system startup module, the system kernel, the application on the system, and establish them for them. As a trust chain, every measure is done, and the measure is stored in the PCR. Once the platform state is measured and stored, the attacker can't forge the state of the platform to steal the data. When users access the large data platform, the measurement values of the platform integrity report are compared with the stored measurement standard reference values, and the integrity of the stored values is identified by the trusted platform. If the agreement is consistent, the storage environment is fully trusted; otherwise the data has been changed during the system start-up process.

![Diagram](image)

Figure 1 Trusted security storage based on trusted computing

4.2. Trusted security transmission

During the transmission process, data must be authenticated by password or digital certificate to ensure that data comes from a trusted party. The sender encrypts a signature with a private key such as name, certificate number, and the receiver uses the public key to decrypt it. If it is successful, the data can be trusted; otherwise the data is not trusted. Assuming that the server sends the file to the requestor, it takes at least three steps: (1) the server encrypts the file with its private key; (2) the server sends the encrypted file to the requestor; (3) the requestor declassified the received file with the public key of the service party. In order to ensure that the data information is not leaked and tampered in the transmission process, the trusted authentication service is introduced into the large data platform to increase the authentication of the identity and platform legitimacy, which ensures the safe transmission of the data, thus ensuring the reliability and security of the large data in the transmission. Since the key of TPM is managed by the tree structure, the server encrypts the file from the sub key, the parent key to the root key SRK with its own private key before transmission to ensure the confidentiality of the data in the transmission process. When the requestor receives the file sent by the server, it must first from the root key SRK to that. All the parent keys of the key are decrypted layer by layer until the key corresponding data is completely decrypted so that the plaintext of the file can be seen [9].

In order to ensure that data is protected from accidental or intentional modification during transmission, it can be verified by verifying the integrity of the transmitted data. Before data transmission, the corresponding hash value is obtained by SHA - 1. After the receiver receives the data, the hash value of the data is compared with the hash value of the transmitted data to judge whether the data is tampered. If the comparison does not indicate that it has been damaged or tampered with, it refuses to receive further data. If the comparison is consistent, the transmitter and data are trusted. Because the security requirements of large data are different in the transmission process, different data
encryption requirements are different, so in order to ensure the security and reliability of the data in the transmission process, different data transmission schemes should be provided in accordance with different data security requirements under trusted computing.

4.3. Trusted security authentication

User authentication is the main foundation of access control. Authentication and access control in the service of big data are more important. Based on trusted computing platform, the security authentication in big data service can be realized, and it can provide more authentication than user name and password.

![Figure 2 Authentication Based on Trusted Computing](image)

As shown in Figure 2, the public key process (PP) is started by the process management module (PM) and submits the public key to the operating system process management module, and then the process that carries the public key enters the waiting state. After PM receives the public key, the public key is submitted to the authentication module (CA) to activate the CA operation. After CA gets the public key, it extracts the private key from the specified location, and uses the specified algorithm to authenticate, then returns the authentication result to PM. If authentication is passed, PM allows the public key process (PP) to continue to run. If authentication fails, PM prohibits the operation of PP with public key. In this mechanism, CA is not visible to PP and is only scheduled by PM. PM submits the public key to CA in pipes or message queues. The private key generator of the hardware system provides the private key, the authentication algorithm is a RSA asymmetric encryption algorithm that supports the 2048 Byte key, and the secret key length can be selected according to the actual operation condition.

5. CONCLUSION

Huge amount, high speed production, variety and distribution synergy of large data make it face lots of security risks in the process of its collection, storage, transmission and use. At present, the means of information security protection can no longer meet the needs of information security in the era of big data, and can not adapt to the continuous development of big data technology. In this paper, under the trusted computing, it takes the data access control as the core to realize the active defense, and ensures that the large data can be controlled, trusted and manageable in the process of storage, transmission and authentication, so that the large data is more secure and effective.

REFERENCES

[1] Viktor Mayer-Schonberger, Kenneth Cukier Big Data: A Revolution That Will Transform How We Live, Work and Think[M].Hodder Export,2013.
[2] Sidhu JSS.Security in Computing and Communications[M].Springer Berlin Heidelberg,2014.
[3] Goh W, Yeo C K. Teaching an Old TPM New Tricks: Repurposing for Identity-based Signatures Security & Privacy [J], 2013, 11(5): 28-35.

[4] Oliveira DP R, Andreia FML, Delisandra FV, et al. Energy Consumption Analysis of the Cryptographic KeyGeneration Process of RSA and ECC Algorithms in Embedded Systems [J]. Latin America Transactions, 2014, 12(6): 1141-1148.

[5] SHEN Zhi-dong, LI Li, YAN Fei, et al. Cloud computing system based on trusted computing platform [C]. Proceedings of the International Conference on Intelligent Computation Technology and Automation, 2010: 942-945.

[6] AHAMED B Bazeer, MOHAMED S Syed Sabir. Implementation of trusted computing technologies in cloud computing [J]. International Journal of Research and Reviews in Information Sciences, 2011(1): 7-9.

[7] Trusted Computing Group. TCG Guidance for Securing IoT [EB/OL]. http://www.trustedcomputinggroup.org/wp-content/uploads/TCG_Guidance_for_Securing_IoT_1_0r14-public-review.pdf, 2016-03-10.

[8] O’driscoll A, Daugelaite J, Sleator R D. Big data, hadoop and cloud computing in genomics [J]. International Journal of Information Management, 2013, 46(5): 774-781.

[9] Wei L, Zhu H, Cao Z, et al. Security and privacy for storage and computation in cloud computing [J]. Information Sciences, 2014, 258: 371-386.