Research on Big Data Platform Security Based on Cloud Computing

Xiaxia Niu(✉) and Yan Zhao

School of Information, Beijing Wuzi University, Beijing, China
15136212624@163.com, 605671232@qq.com

Abstract. Emerging services such as cloud computing, the Internet of Things, and social networking are driving the growth of human society's data types and scales at an unprecedented rate. The age of big data has officially arrived. The use of cloud computing technology to bring great convenience to big data processing, solve various deficiencies in traditional processing technology, make big data more application value and service value, but at the same time, it also brings new security problems. By analyzing the security threats faced by cloud computing-based big data platforms, a cloud computing-based big data platform security system framework is proposed, and a security deployment strategy is given.

Keywords: Cloud computing security · Cloud computing · Big data

1 Introduction

In 2010, the global data volume entered the era of ZB. According to IDC prediction, by 2020, the world will have the amount of 35 ZB data. Massive data will affect our work and life in real time. Even the national economic, social development and big data era has arrived.

The arrival of the era of big data has put forward higher requirements for the real-time and effectiveness of data processing. Traditional IT technology has been unable to meet the needs. Cloud computing technology has made the data analysis, data mining and data processing of mass data become a reality. The large data platform based on cloud computing provides a better service to users, and it also brings a series of security problems. On the one hand, big data means massive amounts of data, and it means more complex and sensitive data. This data will attract more potential attackers. Once a successful attack, hackers can get more data, which in turn reduces hackers. The offensive cost increases the "yield rate." On the other hand, the ownership and use rights of some sensitive data are not clearly defined. Many big data services or big data applications based on data analysis do not take into account the possibility of user privacy issues. In addition, under the cloud computing deployment architecture, computing, storage, and network resources are loosely coupled, resources are allocated on demand, and network boundaries are blurred, which also brings security challenges [1].

In this context, the purpose of this paper is to propose a cloud computing-based big data platform security system framework and provide a big data platform security strategy.

J. Li et al. (Eds.): SPNCE 2019, LNICST 284, pp. 38–45, 2019.
https://doi.org/10.1007/978-3-030-21373-2_4

2 Overview

2.1 Cloud Computing and Big Data

2.1.1 Cloud Computing and Its Architecture

Cloud computing is a developing concept. There is a variety of explanations for what exactly is cloud computing. There is no universally accepted definition. The definition given by Wikipedia [2] is that cloud computing is a computing model that provides dynamically scalable virtualized resources to users through the Internet. Users do not need to know how to manage the infrastructure supporting cloud computing. The definition of NIST provided by the National Institute of Standards and Technology [3] is: Cloud is a parallel distributed system composed of a group of interconnected virtualized computers. It is based on the dynamics of service contracts between service providers and consumers. Computing resources.

Cloud computing can provide elastic resources on demand. It is a collection of services. Combining current cloud computing applications and research, its architecture can be divided into three layers: core services, service management, and user access interfaces. The core service layer abstracts the hardware infrastructure, software operating environment, and application programs into services. These services have the characteristics of strong reliability, high availability, and retractable scale, which meet diverse application requirements. Service management provides support for core services to further ensure the reliability, availability, and security of core services. Users access the interface layer to achieve end-to-cloud access.

2.1.2 Big Data and Its Main Features

Generally speaking, big data means large and complex data sets which are difficult to be processed by existing database management tools or traditional data processing software. People use the word “big data” to describe and define the mass data produced in the era of information explosion, and to name the technology development and innovation related to it.

Large data has the characteristics of “4V”, that is, which is quantified, diversified, fast and valuable.

(1) Volume: refers to a very large amount of data, that is, a large amount of data storage, a large amount of calculations, and the data has jumped from the TB level to the PB level.
(2) Variety: It means that big data includes not only structured data tables and semi-structured texts, videos, images, and other information, but also the interaction between data is very frequent and extensive, including unstructured data. The proportion has increased year by year. Many types of data impose higher requirements on data processing capabilities.
(3) Value density is low. Big data has a relatively low value density. For example, with the wide application of the Internet of Things, information perception is omnipresent and information is massive, but the value density is low, and there is

a large amount of irrelevant information. Therefore, it is necessary to make predictable analysis of future trends and patterns, and use machine learning, artificial intelligence, etc. to perform in-depth and complex analysis. How to more quickly complete the value extraction of data through powerful machine algorithms is a difficult problem to be solved in the era of big data.

(4) Velocity: It refers to the continuous updating of data and the rapid growth. At the same time, the processing speed of data storage and transmission is also very fast. This is the most significant feature of Big Data that distinguishes it from traditional data mining.

2.2 Big Data Platform Security System Framework Based on Cloud Computing

Based on cloud computing, the security protection of big data platform is based on data, from the aspects of data access, use, destruction, modification, loss, and leakage. It mainly includes the following aspects [4].

(1) Network security: refers to the design, construction, and use of the platform network itself, as well as various security-related technologies and methods based on the network, such as firewalls, IPS, and security auditing.
(2) Server security: including server virus protection and server security configuration and reinforcement.
(3) Storage Security: Data preservation and backup and recovery design.
(4) Virtualization platform security: It includes the isolation, configuration, and reinforcement of virtual machines, malicious virtual machine protection, and monitoring.
(5) Platform software security: System security including software such as operating systems, databases, data processing software, and platform components, as well as system security analysis and hardening using security assessment management tools to improve the security of these systems.
(6) Application security: Including application access security and application data security.
(7) Security management: complete user identity authentication and security log audit trails, as well as unified analysis and recording of security logs and events.
(8) Interface security: including authentication of internal and external interfaces, transmission security, and interface call control. The cloud-based big data platform security system framework is shown in Fig. 1.

2.3 Data Security Threats Faced by Cloud Computing

Due to the large scale of cloud computing systems, the application and privacy data of many users are concentrated. At the same time, cloud computing has unprecedented openness and complexity, and its security faces more severe challenges than traditional information systems. Cloud Security Alliance CSA and Hewlett-Packard jointly listed seven aspects of cloud computing security issues.

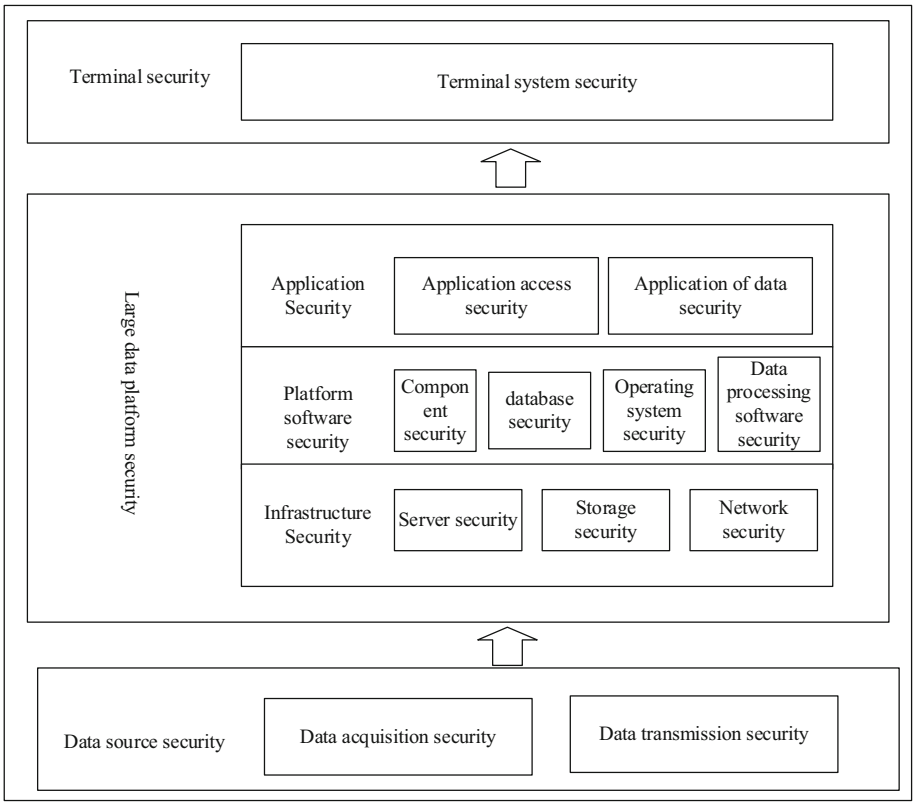

Fig. 1. The cloud-based big data platform security system framework

(1) Data loss and leakage. The security control of data in cloud computing is not high. The lack of security mechanisms and management deficiencies may cause data leakage. Whether it is private data or important national data, if it is lost or leaked, it will have bad consequences.
(2) Sharing technology vulnerabilities. Cloud computing is a large shared data center. The greater the degree of sharing, the more vulnerabilities and the more attacks.
(3) Supplier reliability is not easy to assess. To avoid the theft of sensitive information, a reliable service provider is needed, but how to conduct a credible assessment of the service provider remains to be studied.
(4) The identity authentication mechanism is weak. Since a large amount of data, applications, and resources are concentrated in the cloud, if the authentication mechanism of the cloud computing is weak, the intruder can easily obtain the user account and log in to the user's virtual machine to perform various illegal operations.

(5) Unsafe application interface. The cloud computing application system is very complicated, and the security is more difficult to ensure. Some interfaces may be used by the attack program, which is prone to security problems.
(6) The cloud computing is not running correctly. In terms of technology use, hackers may progress faster than technicians, use legal identity to fish in troubled waters, and illegally run cloud computing.
(7) Unknown risks. The transparency of the service allows the user to use only the web front-end interactive interface. It does not know which platform the vendor uses or which security mechanism to provide. In other words, the user cannot know whether the cloud service provider has fulfilled the service agreement as promised.

3 Security Strategy of Large Data Platform Based on Cloud Computing

3.1 Infrastructure Security Strategy

3.1.1 Internal Network Security

(1) Security domain division. According to the nature of system resources and different security levels, data resources can be divided into different security domains, such as:

(a) Application area: According to the application system's importance and safety level requirements, it can be further subdivided.
(b) DMZ Zone: Internal servers that need to provide external services are deployed in the region to ensure the security of internal business systems.
(c) Management and maintenance area: The platform management server area mainly includes public and platform servers such as network servers, host management servers, database management servers, and security management servers.

(2) Security domain isolation and protection technology deployment. Through the deployment of authentication, role-based access control, ACL, VLAN, MPLS, VSAN and other technologies, the isolation of different security domains can be achieved. Deploy an intranet security system such as a firewall, IPS/IDS, and baseline analysis technology on the core switch to implement intelligent security monitoring and control. At the same time, through the in-depth analysis of application layer protocols such as FTP and HTTP, the content transmitted by these protocols is accurately restored, and the data sent from the internal network device is discovered to prevent the leakage of sensitive information. Different resources are deployed in different service areas. The service areas are isolated by VLAN and virtual firewalls [5].

3.1.2 Storage Security

In the process of uploading data, when uploading to the cloud, data may leak due to server failure. After the cloud platform suffers from illegal access, data may be forged, tampered with, and stolen. Therefore, after the data is stored in the cloud, it is encrypted and needs to be operated by the corresponding encryption technology. The big data in the cloud is divided into two types: static data and dynamic data. The corresponding

data encryption mechanism also has two kinds of static data confidentiality mechanism and dynamic data encryption mechanism [6, 7]. There are two encryption algorithms for static data encryption, namely symmetric encryption algorithm and asymmetric encryption algorithm.

After the data holder encrypts and splits, it uploads to the cloud computing platform. The user has to decrypt the data after downloading the required data. If there is a phenomenon of being stolen or lost when data is stored, transmitted, and used, it will be avoided because it is encrypted. At present, the encryption technology that adopts mainstream data for cloud computing mainly includes proxy encryption and attribute encryption. The ciphertext-based attribute encryption (CP-ABE) and key (KP-ABE) data encryption methods have their own characteristics when applied. The data encryption model of cloud computing can effectively enhance the security of data by deploying agents. Because the cloud platform is a semi-trusted agent when it is used, the architecture of the PRE can be transplanted into the cloud computing. A security scheme with a high security. If the b user wants to share the data uploaded by the user in the cloud after encryption, a needs to generate a transition key based on the user's information and b's public key. This key only has the function of mutual conversion between ciphertexts. The ciphertext of a can be changed into ciphertext for B, and after the user downloads the ciphertext accordingly, the user can perform more operations on the data shared by a.

Data security is a key part of big data cloud storage security. In the process of storing big data, the selected encryption technology must be effective and reliable, and it plays a vital role in the establishment of big data storage security system. The use of reasonable and scientific encryption technology can not only ensure the corresponding confidentiality of big data in storage, but also have important significance for cloud computing and users to achieve optimal allocation of network resources [8].

3.1.3 Server Security

(1) Security configuration and hardening

The operating system should follow the principle of minimum installation and install only the required components and applications. Configure operating system and database system access control measures, identify and authenticate logged-in users, and set security policies such as password policies, account policies, audit policies, and user rights. Security assessment and optimization are performed for devices such as compute nodes, storage nodes, and management nodes, and security hardening is implemented when necessary. Deploy a unified policy to upgrade server operating system software and application software.

(2) Server security protection

Monitoring of important servers, including monitoring of the server's CPU, hard disk, memory, network and other resources. Deploy security measures for the server, such as host firewall, host IDS, and so on.

(3) Server virus protection
Deploy a unified management of anti-virus and anti-malicious code products to achieve centralized management of virus and malicious code protection to ensure that the server is free from virus threats. Ensure that antivirus products can be upgraded in a timely manner.

3.2 Platform Software Security Policy

3.2.1 System Hardening

System hardening mainly reinforces the operating system, database, etc., and fixes the loopholes in the software itself. The following hardening procedures can be adopted: (a) Minimizing cutting: Abolishing unnecessary components and services. (b) Code Security: Optimize the code according to the specification. (c) Security configuration: Configure according to the configuration hardening specifications in the industry. (d) Security Test: Tests are conducted using the industry's leading scanning tools. (e) Integrity protection: review by means of an integrity check tool.

3.2.2 Database Security Audit System Deployment

The database audit system is deployed in the intranet security system, collects, analyzes, and recognizes the data stream that accesses the database through the bypass interception method, and monitors the running status of the database in real time, records multiple access behaviors, and monitors abnormal access operations, etc. The distributed deployment mode can be used to centrally manage the domain database auditing system in the high-efficiency management area. The branch nodes are no longer independent, and all under the unified supervision, effectively improve the efficiency of regional security auditing.

3.3 Application Security Policy

3.3.1 Application Access Security Policy

The application system shall provide a dedicated login control module to perform identity identification and authentication of the logged-in user, and important information (such as user accounts and passwords) is encrypted and stored; identity authentication, uniqueness of user identity identification, complexity check of user identity authentication information, and login are provided. Fails to process functions, and configures related parameters according to security policies. Provides security audit function that covers each user. Audits important application security events. The contents of the audit records should include the date, time, originator information, and type of the event, descriptions and results, etc.

3.3.2 Application Data Security Policy

Deploy a web application attack defense system to detect and prevent as many attempts as possible to tamper with static web page files, dynamic web page script files, or dynamic web page data. At the same time, web page tampering events caused by other unknown types of attacks can be promptly restored. Blurring sensitive data involved in Web presentation data.

4 Conclusion

The emergence of cloud computing and big data has changed the traditional data storage model and data computing model. It has brought tremendous changes to human society. People are enjoying the benefits of this kind of change. It is therefore increasingly recognized the importance of information technology. The combination of cloud computing and big data provides users with better services but also creates new security threats. This paper proposes a cloud computing-based big data platform security architecture framework, and gives security strategies for big data platforms. Only by providing users with reliable, reliable, and cost-effective cloud computing services can enterprises succeed in the field of cloud computing.

Acknowledgments. The study is supported by the National Nature Science Foundation of China "Research on the warehouse picking system blocking influence factors and combined control strategy" (No. 71501015), Intelligent Logistics Collaboration Center and Tongzhou Canal Project Leaders.

References

1. Ban, R., Tu, L., Liu, H.: Cloud computing platform for big data security strategy research based on J. Post Design Technol. **10**, 74–78 (2017)
2. Cloud Computing - [EB/OL] (2013). http://zh.wikipedia.org/wild/wikimeco
3. Mell, P., Grance, T.: The NIST definition of cloud computing [R/OL], 11 February 2010. http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc
4. Yaofeng, Z.: Analysis of cloud data security in the background of big data. Cyber Secur. Technol. Appl. **11**, 102–103 (2014)
5. Li, Y.: Cloud security data center network security deployment. Netw. Inf. (6), 38–39 (2012). Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010)
6. Bellare, M., Rogaway, P.: Introduction to Modern Cryptography. UCSD CSE,207,2005,207
7. Riyest, R.L., Shamir, A., Adleman, L.: A method for obtaining digital signatures and public-key cryptosystems. Commun. ACM **21**(2), 120–126 (1978)
8. Sun, L.: Discussion on the security of big data storage based on cloud computing. Netw. Secur. Technol. Appl. (2018)