On Data and Virtualization Security Risks and Solutions of Cloud Computing

Xiangyang Luo*
1. China Institute of Electronic Equipment System Engineering, Beijing 100141, China;
2. Zhengzhou Information Science and Technology Institute, Zhengzhou 450002, China;
3. State Key Laboratory of Information Security (Institute of Information Engineering, Chinese Academy of Sciences), Beijing 100093, China.
Corresponding author, Email: xiangyangluo@126.com

Lin Yang and Dai Hao
China Institute of Electronic Equipment System Engineering, Beijing 100141, China
Email: yanglin61s@yahoo.com.cn

Fenlin Liu
Zhengzhou Information Science and Technology Institute, Zhengzhou 450002, China
Email: liufenlin@vip.sina.com

Daoshun Wang
Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China
Email: wangdaoshun@gmail.com

***Abstract*—Data security and virtualization security issues are two key bottlenecks restricting the application of cloud computing promoting and applications, especially for the Cloud-based media computing system. In this paper, states of the art of the techniques on cloud computing data security issues, such as data encryption, access control, integrity authentication and other issues is surveyed, on this basis, the key technical issues of the cloud computing data security should concern about and focus on are indicated, and some corresponding countermeasures and suggestions are presented. For the virtualization security problem introduced by private cloud computing, the security risks induced by virtualization are analyzed and classified, and then based on the divide-conquer idea, for each kind of security risk, some corresponding solutions are presented.**

***Index Terms*—Cloud Computing; Data Security; Virtualization Security; Countermeasure Suggestions**

I. Introduction

Service-based cloud computing is an important form of the information infrastructure in the Internet era, which adopts new business model to provide high-performance, low-cost computing and data services, supporting all kinds of Informationization application. Along with the rapid development and application of this new network technology, new security problems appear constantly, and become an important factor in restricting industrial development. A series of non-traditional security threats brought by the rapid development of cloud computing need completely new and higher demands on information security. We need to intensify the research efforts of the security technology synchronously and urgently, and provide security support of the technology, products and infrastructure for cloud computing technology development and application. Following the development and application of cloud computing industry, although a number of IT companies in the world have launched their cloud computing products, security problems are not resolved, related security events occurred one after another. Coupled with the popularization of cloud computing, and the deepening understanding of cloud computing, the security issue has become the biggest concern in the using cloud computing and the migration to cloud computing. If the bottleneck problem of cloud computing security cannot be resolved, cloud computing technology is difficult to carry out the industrial upgrading and application promotion.

Following cloud computing security research in the academic community, part of literatures have begun to pay attention to cloud computing security issues, but the vast majority of literature still remain in the research of cloud computing deployments, services, applications and other related issues, and depth research on cloud computing security issues has not yet commenced, and the key security issues like data privacy protection and virtual security related to cloud computing are still lack of support of basic theory and effective technology. So carrying out study of cloud computing security issues has important practical and theoretical value [1].

Currently, cloud computing security issues have obtained increasing attention. The International Conference RSA on information security listed cloud

doi:10.4304/jnw.9.3.571-581

computing security as focal issues. Many research institutions, enterprises and standardization organizations have launched related research. Security vendors are concerned about and developing various types of security cloud computing products. This paper will focus on data security and virtualization security issues of cloud computing. Berkeley Cloud Computing White Paper [2] sets out 10 issues and opportunities facing Cloud Computing, in which, related security issues include data loss, the security and auditability of data, and the virtualization security. This shows important status of the data security and the virtualization in a series of security issues of cloud computing.

In terms of data loss, the software sets have already been improved a lot in the cross-platform, but essentially, API (Application Programming Interface)of cloud computing is still private, or a uniform standard is not established currently. So, it's difficult for users to migrate their data and programs from one station to another, and this is also the reasons that a lot of users do not want to use cloud computing. In terms of the security and auditability of data, current cloud offers the public network essentially, and would be suffered more attacks. So it's difficult for users to put sensitive data into the cloud.

In terms of research in the data security of cloud computing, the article [3] also believes that data security is an important security issues of cloud computing. Handing the data originally stored in the local control to an external service center of cloud computing is not easy. With the development of network technology, the bandwidth will not be the main obstacle, but the security is still the most important concerns. However, as money has long been accustomed saving in the bank, in the future, data Bank will certainly appear, sooner or later. Technique may not be the main stumbling block, and institutions, laws, Integrities, habits and ideas, these non-technical factors will directly influence the popularity of cloud computing. the article [4] investigates the risks of cloud computing, and puts forward a system design scheme can be used to capture information flow in the cloud for whether users of cloud computing could use their own information at any time, and how to prevent their information to be illegally obtained. Providing safe and effective access in large-scale off-site data is an important component of cloud computing, and the article [5] gives a different key to encrypt the data block to provide a mechanism for the access control based on the elastic encryption, against the security issues in this mode of data owner "Write" - user "read". The article [6] discussed the cloud storage, raised cloud storage architecture assumptions and involved issues, including storage security, but did not give a corresponding solution. The article [7] and [8] analyzed privacy, security and the issue of credibility of cloud computing, and discussed a number of methods which can enhance the credibility and security.

In terms of research in the virtualization security of cloud computing, the paper [9] in order to enhance the cloud of virtual machine (VM) security, proposed a measure framework of virtual machine operation. In this framework, there is a module in virtual machine to transmit the operation of virtual machine to a trusted virtual machine, by comparing with a good reference table prepared in advance to determine whether the virtual machine in a normal state of operation.

From the above literatures, as the applications and services model of cloud computing is different from the traditional end-to-end implementation of the encrypted communication to ensure data security, an untrusted third party will participate the process of virtualization storage and processing of massive data, this brings new data security and virtualization security issues. Certainly, the encryption technology is still a powerful measure to ensure data security of cloud computing. But how to implement the highly efficient encryption, how to quickly search on the data encrypted, how to carry on disaster recovery and fast recovery, how to proceed the access control of data and so on, all of above are a series of key and difficult problems that data security of cloud computing must resolve. Consider the data security and virtualization security of cloud computing are two key issues which are relative and need to be solved, this paper focuses on these two issues. Firstly, the research progress of data security issues of cloud computing such as data encryption, access control, and integrity authentication, is reviewed; Secondly, the key technologies and key issues that the data security issues of cloud computing should be concerned about are pointed out; Thirdly, some corresponding countermeasures and suggestions on data security are proposed; Fourthly, based on the idea of partition, the cloud computing virtualization risks are classified and analyzed respectively; Lastly, the corresponding response measures and safety recommendations of virtualization risks are given. The proposed solutions can effectively avoid the virtual cloud computing security threats, and improve the security of cloud computing systems, especially the private cloud computing systems.

II. Research Progress of Related Technologies on Data Security Issues of Cloud Computing

In cloud computing environments, the usual data transmission and storage mode is shown in Figure 1:

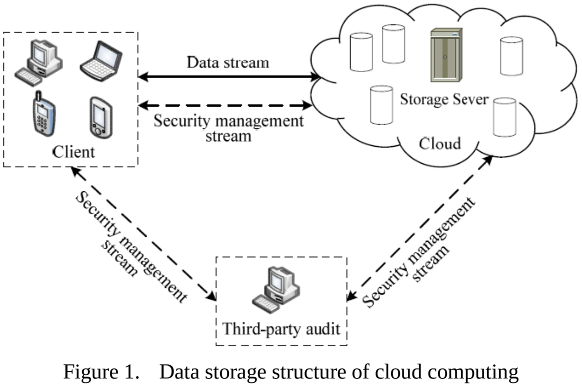

Data stream
Security management stream
Client
Storage Sever
Cloud
Security management stream
Security management stream
Third-party audit

Figure 1. Data storage structure of cloud computing

The effective protection of user privacy data is the primary problem of cloud security. From the above figure, we can see that cloud computing stores a large number of data files on the distant servers, and users can reduce the burden of storage and computation, then enjoy the flexible and efficient service brought by cloud computing. But the characteristics of cloud storage make users' data faced with many security risks, includes: (1) the traditional security region partition is invalid. Because of the cloud storage service must be scalable, security boundaries and protection equipment cannot be clearly defined, which increases some difficulty for the implementation of specific protection measures; (2) the cloud storage transmits data through the network. The service interruptions, data destruction, information stolen and tampered caused by the malicious attacks in the network pose a severe challenge to the security of data communications, access authentication and confidentiality; (3) from the user's view, the cloud storage of data makes cloud computing service provider obtains the data access control, and the user's data is faced with privacy security threats. People worry about that the sensitive personal data will be disclosure, misuse or missing by putting the data in cloud environment. To solve the above issues, in recent years, researchers made a lot of research work in the data security access control mechanisms, data integrity, authentication, ciphertext to retrieve and data encryption technique of cloud computing environment.

A. Data Security Access Control Mechanisms of Cloud Computing Environment

As the cloud service requires secure cross-domain collaboration, and needs to provide protection for being mistaken of personnel and equipment identity, users need to have the convenient and comprehensive ability to control their own data. If not to access all resources like an interoperable stand-alone system, it will affect the usability of cloud. Therefore, cloud service providers should provide appropriate mechanisms to support users with regulation, authorization and access control on their own data, and let the user know whether there are other users to access or copy data stored in the cloud.

Secure identity management and control are essential for any network environment, but it will become more complex in the cloud computing environment. Cloud Security Alliance believe that the management of authentication and access control for company is still the biggest challenge facing IT department, however, the existing cloud service providers themselves didn't comprehensively place authentication service in their cloud computing platform. In the long term, spreading authentication services into the cloud services is necessary. The implementation of sound data classification mechanism can determine data sets involved in the things, and determine the control mechanism applied to the data set in a particular case. Taking these initiatives can help companies make correct decisions. Related research progress can be divided into two respects.

1) Access Control Based on Virtualization Technology

As virtualization technology is becoming the core technology of cloud computing, security control for cloud computing is concentrated on virtual machine-based access control. Currently it mainly uses strong access control mechanisms to achieve isolation in the communication of virtual machines. sHype implemented strong access control module in the Xen Virtual Machine Manager, and using this module can control communications between multiple virtual machines on a single physical node. IBM proposed Trusted Virtual Domain (TVD) [10], and the security control for TVD is the control for the inter-domain communication. Payne et al. proposed a hierarchical access control model [11], and for the complex problem of the virtual machine-based access control strategy, they proposed a hierarchical classification framework to simplify it.

2) Cross-domain Access Control

When the users' cross-domain access resources they need to set up certification services in the domain boundaries, and make a unified identity management for accessing to shared resources. Since each trusted domain has its own access policy, so it needs to support the synthesis of strategies. It is first proposed by Mclean in mandatory access control framework, and synthesizes two security grids to a new grid structure. The synthesis of strategies while also guarantees the security of the new strategy, and the new synthetic strategies must not be contrary to the original access control policies in various domains. In this aspect, Bonatti [12] proposed an algorithm for synthesis of access control policies, it use synthesis operator to synthesize the security policy based on set theory. Wijesekera et al. [13] proposed a synthesis algebraic framework of a strategy based on the license status changes.

B. Data Integrity Certification of Cloud Computing Environment

In cloud computing environment, users do not have to reserve the data locally, but they must be convinced that their own data in the cloud can be well preserved and maintained. So users need a safe way to regularly verify the correctness of their own stored data. Users have not enough time, ability or resources to manage their data, thus they put this task entrusted to a trust TPA. This brought a lot of new security challenges unresolved, in which the most concerned is the validation issues of the integrity of data stored on untrusted servers.

The existing research achievements include: Provable Data Possession (PDP) [14] model and Proof of Retrievability (POR) [15] model.

Atenises et al. define the Provable Data Possession (PDP）model to confirm the existence of the file on the server untrusted. Their programs audit data with RSA-based homomorphic tags. So they can provide public verifiability. But Atenises did not consider the dynamic data storage, and the using may reveal the users' data information. Since then, Atenises et al. also proposed a dynamic version of the PDP [16]. But this system has confined of inquired number, and cannot support dynamic data manipulation, for example, it does not support insertion of data blocks.

Juels et al. define the Proof of Retrievability (PoR) model, using spot-checking and error-correcting codes to ensure persistence and recoverability of data. Specifically, some special data blocks called sentinels will be embedded into the data file F for the purpose of detection, and then file F will be encrypted to hide the position of these special blocks. However, this method also only supports a limited inquiry, and special precomputation hinder the dynamic data updating. In addition, this program does not support public verifiability. Shacham [18] designed an improvement scheme of PoR, and gives a complete proof of security in the security model defined in paper [17], but still only considered the static data files. Gellman [19] gives an exploratory construct for dynamic provable data possession, and they improved PDP model mentioned in paper [14] to make it supports provable data possession. Particular, in order to support the update of the data, they did not use the method of calculation of the index information in the original model, but before certification process, use the authenticated skip list data structure to authenticate the tag information to updating data blocks.

In addition to the above methods, some literatures also raised other data validation methods. For example, Schwarz et al. raised method based on algebraic signature [20], Yun et al. raised the program based on tree structure "MAC tree" [21], Wang et al. raised method based on homomorphic signature and RS Error-correcting codes [22]. In addition, NEC proposed the PDI (provable data integrity) method [23], and this method improve and increase the processing speed of the POR method and the size of verification objects, and be able to support open authentication.

Known from the existing data validation, the problem of supporting open authentication and data dynamically update has not been completely solved. How to design safe and efficient programs to solve these two problems at the same time is still a very challenging task.

C. Ciphertext Retrieval Technology

In order to solve the problem of data protection, a common method is that users encrypt data, and then put ciphertext into the server. When the encrypted data stored in the cloud formed the scale, retrieval of encrypted data become an urgent problem which needs to be solved. Under normal circumstances, the ciphertext is not available for retrieval semantic and statistical properties. So retrieval of ciphertext is a more difficult problem. The existing ciphertext retrieval methods mainly include Linear searching method, Public key based on keyword searching method, Security index searching method, Order Preserving encrypted searching method and so on. The following is a brief description.

1) Linear Searching Method

Song et al. first proposed the search method for ciphertext data [24], the method used was a linear search algorithm. In the linear search algorithm, they first use the symmetric encryption algorithm to encrypt the plaintext messages. For any ciphertext information corresponding with each keyword, it will generate a bunch of pseudo-random sequence, and generates a check sequence determined by the pseudo-random sequence and the ciphertext. Sum of the Pseudo-random sequence length and the test sequence length is equal to the length of the ciphertext information. Pseudo-random sequence and test sequence encrypted the ciphertext again. When searching, the user submits the plaintext information to be searched, and then searching method calculates the corresponding ciphertext sequence. On the server side, the ciphertext message sequence is linear model 2 plus with each sequence in the searching range. If the obtained results meet the parity relations, it shows that the ciphertext information sequence matches successfully, otherwise, it continues searching. Linear searching method is a one-time pad, extremely resistant to statistical analysis. However, such methodologies have an obvious drawback. It needs to match ciphertext information successively, and the time complexity degree is very high, then it is difficult to apply to large data sets searching in the case of cloud computing environments with mass data.

2) Public Key Based on Keyword Searching Method

Boneh et al. proposed the public key based on keyword searching method [25]. This method can access the remote database to obtain data with the lack of client storage and computing resources. The encrypted data has a number of different sources, and it needs to do a searching for such encrypted data. A viable idea is that data is encrypted with public key encrypt. It first generates a public key and a private key, and then encrypts plaintext keywords which will be stored with the public key to generate the ciphertext can be used to search. During the searching, it encrypts the plaintext sequence provided by the user intended to search with the public key, and then carries out ciphertext keyword matching.

3) Security Index Searching Method

Boneh et al. proposed the security index searching method, to solve the disadvantage that simple index is vulnerable to statistical attacks. The mechanism is that the key used for per encryption is a set of pre-generated inverse Hash sequence, and encrypted index is being put in Bloom filter. When searching, it first uses the inverse hash sequence key to generate a trapdoor, and then does a Bloom detection. The returned ciphertext is the document required. The shortcoming of this method is the need to generate a large number of key sequences. With the increase of the number of searches, the computing complexity degree increases linearly. In the above searching methods for encrypted information, all searching models are the Boolean models, and thus it cannot do a sorting operation based on the relevance of the retrieved documents. In the actual situation, especially in the cloud storage applications with larger data size, there may be many documents containing a same query keyword. How to identify the most relevant one or several documents in a number of possible documents need to be addressed.

4) Order Preserving Encryption Searching Method

Swaminathan et al. proposed the order preserving encrypted searching method. In this method, the term frequency of each keyword in the document is encrypted by order preserving encryption algorithm. After the encrypted documents that users need to query are submitted to the server, the server first searches encrypted document containing keywords ciphertext, and then makes a sort treatment for corresponding ciphertext of frequency encrypted with order preserving algorithm. Finally, the encrypted document with a high evaluation will be returned to users. This method can sort encrypted document in the case of being given many relational documents, and thus returns the most relational document to users. However, this method does not apply to the query that contains a number of query terms, because sorting method does not know which query terms according to. Moreover, this method only uses the word frequency information in the document, and cannot use the inverse document frequency of the word, and thus the vector space model cannot be directly applied.

Known from the above ciphertext retrieval technology, although the existing methods in the state of the ciphertext can quickly retrieve the required information, but computational cost is very high, all of them do not apply to large-scale data retrieving. In the cloud computing environment, the relevant data for retrieving are often very large, and existing sorting methods are almost unrealistic, therefore, these methods cannot solve these issues. Search methods can sort the documents accordance with the relevance taking advantage of the word frequency information in the document, and improve the retrieval accuracy and the return rate, although the frequency of some commonly used words in the document is very high, it distorts the actual relevance. Therefore, the practical application is also faced with further word frequency analysis and processing work.

D. Study of Data Encryption Technology in Cloud Computing Environment

Since invention of the public key encryption appears, how to deep and unlimited analyze encrypted information without impacting its confidentiality has plagued cryptographers for several decades. Once this issue is resolved, it will completely solve the difficulties faced by the various types of ciphertext retrieval method mentioned in the previous section.

A characteristic of the traditional encryption means is storing data in the box and not letting the public use or analyze data, unless using the decryption key to open the box. Fully homomorphic encryption scheme allows you to analyze and calculate data in the case of data encryption. If an encryption algorithm for addition and multiplication can all find the corresponding operation, then we called it fully homomorphic encryption. In brief, the nature of fully homomorphic encryption technology is arbitrary complex expressly operation can construct the corresponding ciphertext operation in the addition and multiplication. This is different from the partial homomorphic encryption which applies only to single operation (addition or multiplication).

Plaintext data encrypted by fully homomorphic encryption algorithm can be retrieved without restoring the plaintext message. That is, the most relevant documents are returned to the user, and it realizes data analysis and calculation in case of data encryption. It not only protects the user's data security, but also improves the retrieval performance.

In 2009, IBM Fellow C. Gentry, at ACM International Symposium on Theory of Computing (STOC), published a papers entitled "Fully homomorphic encryption based on the ideal lattice" [28], solved the fully homomorphic encryption problem proposed by well-known cryptographer R. the Rivest and L. Adlemanand 30 years ago. The paper's publishing not only generates a great sensation in academia, but also has a major impact on industry. Fully homomorphic encryption technology was known as the title of "the holy grail of cryptography". After Gentry's paper published, fully homomorphic encryption once again has become the hot issue in the field of cryptography nearly two years. In the top three cryptology annual meetings (CRYPTO, EUROCRYPT, and ASIACRYPT) in the field of cryptology, the cryptology scientists have researched fully homomorphic encryption on the basis of the work done by Gentry and made a lot of new results [29-39]. Paper [40] published by Smart and Vercauteren in the PKC 2010 Conference improved Gentry's program, achieved further optimization of efficiency, and had designed integer-based fully homomorphic encryption scheme. Now large companies such as IBM and Google are pushing fully homomorphic encryption technology to the practical application, and apply it to respective system.

The basic principle of fully homomorphic encryption algorithm is as follows, note the encryption operation as E, plaintext as m, and get e after Encryption. That is, $e = E(m)$, $m = E'(e)$. Known the plaintext operation f, we can construct F for E, satisfying $F(e) = E(f\ (m))$. So E is a homomorphic encryption algorithm for f. It assumes that f is a very complex operation. With homomorphic encryption, the sender can send the e encrypted to a third party. The third party does an operation F. The sender gets back the $F(e)$, get $f(m)$ after decryption. The third party completes the work on behalf of the sender, and still knows nothing about the m. This is a surprising result! However, looking for such an E is not easy. Purely from the view of mathematic, $E(x) = x$, is homomorphic. But unfortunately there is no cryptographic effectiveness. The RSA algorithm is homomorphic for the multiplication operation. The corresponding operation F is also a multiplication. Others such as the addition will not be able to construct the corresponding F. But the Paillier algorithm is homomorphic for addition.

If an encryption algorithm that multiplications and additions all can find the corresponding operation, this encryption algorithm will be fully homomorphic. Fully homomorphic encryption technology gets an output with the processing for encrypted data. The result of decrypting the output is same to the output using the same approach for the original, un-encrypted data. In other words, fully homomorphic encryption technology can construct the corresponding cryptographic operations for arbitrarily complex expressly operation. Special significance of the homomorphic encryption is that it can construct the corresponding F for any f. In this way, you can get some incredible applications. I can solve your problem, even though I do not know your problems-this is a less proper metaphor.

We can completely believe that if we really achieve a mature fully homomorphic encryption, and the loss of encryption efficiency compared with classical encryption algorithms is not too much, it will be perfectly used in cloud computing environments, and the range of applications will be very extensive.

III. Problems and Possible Countermeasures Should be Concerned about

A. Cloud Computing Data Security System Research

For the data security problems cloud computing brought, how to implement total life cycle management to cloud computing data security is an important research direction. Figure 2 shows several key aspects related to cloud computing data security. How to give out a more complete cloud computing data security technology system is worthy study, and it is also a key step to clear the problem needed to be resolved in cloud computing data security.

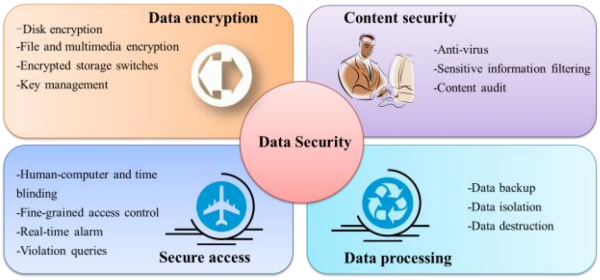


Figure 2. Security system diagram of data life cycle

In addition to study data security system of data life cycle, there is an important research idea at this stage, which is to weaken or transfer the outstanding existing security risks. For example, the most worried thing for the existing cloud computing users is whether system administrators could spy on users' data. Whether it can be considered that before data comes into the user data library, we carry out a conversion operation (similar to encryption) in order to converse user data to massy codes that system administrator cannot understand. Although this method extends the processing time of the cloud, but it can effectively prevent the system administrator peeping user data, to meet the needs of some applications. Such weakening or transfer of outstanding security risks may be a good expedient before cloud computing data security issues are fully resolved.

B. Authentication and Access Control

Authentication and access control is a common network security measures. Compared with traditional network service model, cloud computing is more flexible, and it has higher requirements on authentication and access control. Authentication and access control list provided by Amazon S3 security is a reference available of authentication and access control issues. S3 uses HMAC-SHA1 digital signature to determine the user's identity. HMAC-SHA1 is a message authentication protocol based on cryptographic Hash functions and shared key, which can effectively prevent the data during transmission be intercepted and tampered with so as to maintain data integrity, reliability and security. The core of the HMAC-SHA1 message authentication mechanism is an encrypted hash function, a random key encryption and a secure key exchange environment. When new user registers, Amazon assigns an Access Key ID and a Secret Access Key for him. Access Key ID is used to determine the sender of service request, and Secret Access Key to participate in the process of digital signature to prove that the user is the lawful owner of the account sending service requests. Access control list is a list provided by S3 for user-defined access control policy, and users may set the appropriate access control lists in terms of his need, in order to share his own files with others and achieve the purpose of preventing unauthorized users' access at the same time.

Identity recognition and access management IAM (Identity and Access Management) is one of the key security measures to the ensure safety of cloud computing operation. Many traditional IAM management tools can be transplanted into the cloud, such as authentication, access control, single sign, the segregation of duties, privileged user management, user account management, user self-service and compliance report, etc. In particular, cloud service providers should focus on the following security technologies in authentication and access control:

1) SaaS (Software as a Service) Account Management

2) SaaS Collection and analysis of safety records

3) PaaS (Platform as a Service) application access policy

4) Control technology of privileged user in the IaaS (Infrastructure as a Service)

5) Monitoring and auditing techniques

Cloud-based services need to secure cross-domain collaboration, and ensure that the personnel identity and equipment will not be abused. Therefore, for any identity and access control systems, especially systems involving substantial assets, we should use the authentication framework using in-person proofing or a similar strong authentication processes, as well as strong encryption certificates to authenticate. These certificates will help to establish a "claims-based access control. For transaction and connection specified in the cloud services, after they provide the necessary information, the system should allow transmission and verification statement, without the need to provide more information.

C. Data Integrity and Availability Certification

Huge communication cost caused by the large-scale data makes it impossible for the user to verify the integrity and availability after the data is completely downloaded from the cloud. Thus, cloud users must judge with high confidence probability whether the cloud data is complete through some kind of knowledge proof agreement or probability analysis means under the case of getting few data.

In terms of cloud storage protocol design, the main considerations currently are data integrity authentication, data error positioning, dynamic updating of data, etc. But the design is not perfect, for example, the dynamic data operation of the protocol is only at data block level. Therefore, we hope to find a way to extend the existing agreement operation to the data level. The main difficulty lies in that there is no clear correspondence between the data and tag, and may lead to unnecessary computing and communications consumption.

In terms of the variety demands of protocol design, we can consider to use erasure-coded data, combining distributed authentication of homomorphic token and challenge-response agreement to ensure data integrity and fault location, in order to enable it to resist the Byzantine failure, malicious data changes, server conspiracy and other attacks.

D. Practical Homomorphic Encryption Technology

The homomorphic encryption technology is of revolutionary significance for solving the problem of data security of cloud computing. Once the practical full-homomorphic encryption technology appears, data security issues plagued cloud computing applications will be fundamentally resolved. Therefore, it is foreseeable that the practical full-homomorphic encryption technology has been one of the focus issues of cloud computing security research, until it be completely resolved.

Homomorphic encryption technology research can commence in two ways. One is to optimize the design of the existing homomorphic encryption scheme, the other is to construct new full-homomorphic encryption algorithm.

1) Optimal Design of the Full-Homomorphic Encryption Scheme

Existing full-homomorphic encryption technology, whether based on distance vector lattice obscure, the integer group Approximate Maximum Common Divisor obscure, or the fusion part of the same state and multiply the same state, or the use of the ideal lattice containing the wrong selection (LWE) issues, is without exception calculated in the large algebraic structure, which makes the efficiency of encryption and decryption differ differently to symmetric cryptosystem, and it is difficult to meet the actual demand. Optimize the design of the existing homomorphic encryption scheme, such as through refining the analysis the difficult problems of the existing full-homomorphic encryption algorithm to put forward constraint conditions strategy, and design more efficient probability decryption algorithm to achieve the desired lower number of Gravitas, or the reduction of die remaining elements in integer operation mode in surplus ring, may be able to reduce the length of the encryption and decryption keys necessary to improve the operation efficiency and reduce the expansion of the ciphertext data to meet some of the practical application of needs.

2) Construction of New Full-Homomorphic Encryption Algorithm

Look at the process of development from the source of the full-homomorphism technology to the emerging new approaches nowadays, after a certain deformation or transformation, many math problems can be used to construct public key cryptography, in which should be no lack of good performers. The current program starts from 1-bit data decryption, and extends on this basis. Its computational efficiency is a far cry from the actual needs, and in most cases ciphertext division is very difficult to solve. Therefore, exploring construction of the new practical full-homomorphic encryption scheme, designing practical and efficient homomorphic encryption algorithm is of great significance to solve the problem of cloud computing, data security.

IV. Briefly Description of Virtualization

At present, there still did not have the unification cloud computation definition, NIST defined the cloud computation essential characteristics, the service models and the deployment models, has certain representation, specific as shown in Figure 3 [41].

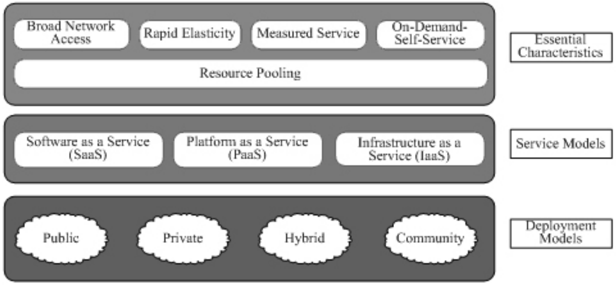
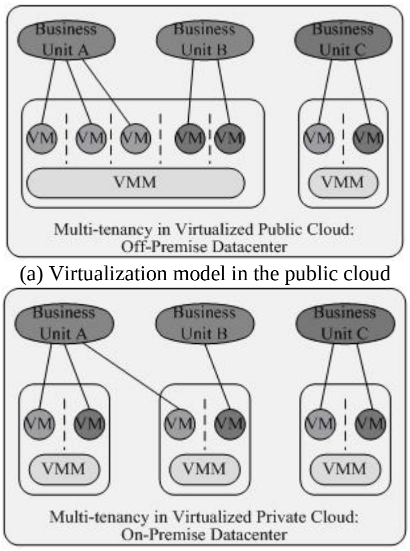


Figure 3. Cloud computation essential characteristics, service models and deployment models defined by NIST

The cloud computation unfolds five essential characteristics, have represented its difference from the tradition computational models. In these five characteristics, has contained the virtualized resource pool, and has used the virtualization technology. Cloud computing provider’s resources, including the memory, processor, memory, network bandwidth as well as the virtual machine and so on, collect in the resource pool, then virtualization uses the multi-renter model, according to the user need, assigns or redistributes the different physics and the hypothesized resources dynamic for many consumers. Although there is the position independency in a certain extent, namely the user beyond control or is unable to know that uses the resources’ accurate physical location, but virtualization may assign the position in the high abstract stratification plane (e.g. country, state, province, or data central).

Refers to the different existing literatures, there are large differences about the virtualization definition. In the virtualization vigorous development’s stage, there did not have the strict standard and the definition defines the virtualization. It is usually thought that the virtualization is refers to the computer parts operate in the hypothesized foundation, but not in the real foundation, and the virtualization is one solution to simplify the management and optimized resources. Virtualization technology assigns flexibly workload different physical machines to achieve resource sharing, is a method that running multiple virtual machine operating systems independent in a physical machine [42], shown in Figure 4, where VM is the virtual machine, and VMM means the virtual machine monitor. The main purpose of virtualization is to simplify the IT infrastructure, and it can simplify access to resources and resource management. Users access the resource through the standard interfaces supported by the virtual resources. By using the standard interfaces, when the changes happen on the underlying physical resources, the case does not affect the user's use. At the same time the overall management of IT infrastructure can be simplified, virtualization reduces the coupling between the user and resource, so the user does not depend on specific physical resources. By using this loose coupled relationship, managers can make the management on IT infrastructure in the base of influence the user minimum. Today's virtualization technology includes microprocessor virtualization, file virtualization and storage virtualization.


Business Unit A
Business Unit B
Business Unit C
VM
VMM
Multi-tenancy in Virtualized Public Cloud: Off-Premise Datacenter
Multi-tenancy in Virtualized Private Cloud: On-Premise Datacenter

(a) Virtualization model in the public cloud

(b) Virtualization model in the private cloud

Figure 4. Virtualization models in cloud computing

V. Risks Analysis of Virtualization Security

Virtualization technology itself is not new, but after being applied to cloud computing, it brings some new security risks and vulnerabilities may be exploited maliciously and may cause major security incidents. Virtualization security has two aspects, one is own safety of virtual technology, another is the introduction of new virtualization security issues. Specifically, we think the following risks of the private cloud virtualization security are worth concerned especially.

A. Risks of Resource Access Control

First, when resources are unified into the same logical platform to storage and use, there is no fixed security border and isolation, the user can only see the logical storage location, do not know the specific storage location of information, this information may be in the off-premise storage infrastructure, which there is the possibility of leakage of secret information, including the infrastructure manager and control the infrastructure by using the vulnerability of virtualization platform might extract the secret information. Second, when the virtualization platform is attacked, the administrator's privileges may be stolen and used maliciously, for example for some users to upgrade or delegated authority, these rights run away will lead to secret information out of control. Zhang et al. [43] proposed a solution based on the nested virtualization, CloudVisor, which separates the virtual machine monitor resource management functions and the security protection functions. The security tool further is implanted into the lower layer of the virtual machine monitor. CloudVisor primary design goal is that, even if the virtual machine monitor and the virtual machine are invasion of malicious behavior, but also can ensure the privacy of user data and the consistency of the virtual machine. This work tries to close to the underlying hardware, which is not a complete sense of the monitoring system, but with the ability of perception and control to illegal behavior.

B. Risks of DOS Attack

Virtualization services have a risk to slow or even stop when a large number of applications and even unlimited use virtualization platform for processing. This is similar to a traditional network, when a server or a group of emergency visits, the server will reduce the speed of response, processing business is slow, in particular, a large flow of cases and the application layer DDoS (Distributed Denial of Service) attacks, cause server downtime and service is stopped. DDoS has a greater risk and use potential at virtualization platform. Due to the virtual platform is designed to provide each user with the necessary service, when an excessive number of users or malicious applications were a lot of services, it will take up too many resources and make the virtualization platform cannot run effectively, thus all users will be affected.

For the former, that the excessive numbers of users are normal using virtualization platform, you can expand the virtualization platform and increasing the resources to resolve. But for the latter, i.e. a malicious application that was a lot of services to take up a lot of damage for the purpose of unlimited resources, which for any kind of high-performance virtualization platform is not affordable, can be very destructive. In addition, this behavior is similar as Botnet in Internet, the malicious users apply a large number of services to attack other virtualization platform, and as a result all virtualization platforms will denial of service.

C. Risks of Virtualization Platform In Building Network

The network connect client with sever is based on the software hubs usually. If two clients are the members of same network, which share the same virtual interface, then the two clients can see the server and all traffic of the client end because the traffic of two networks connected by a virtual switch machine is through the same physical network card. For all client ends of a server, all other client ends and servers share the same software stack. Network stack sharing is a major problem of virtualization security, if all the client ends and the server share the same network software stack, the attacker can access to the entire stack by only attack a computer of clients.

D. Risks of Virtualization Platform'S Security Management

The virtualization platform built in private cloud is physical isolation with Internet, thus the library of viruses and Trojans for the virtualization platform cannot update rapidly, and the vulnerabilities of virtualization platform cannot be repaired in time. Internet environment will always produce new viruses and Trojans, Internet-based business can update security software through accessing Internet servers in real time and ensure obtaining the latest security services. For the private cloud, in the LAN (Local Area Network) environment, there is a "time poor" in updating virtualization platforms virus and Trojans library and fixing the vulnerabilities, this bring a risk for the security of the virtualization platform.

There is a qualitative change between the management of virtualization platform and that of traditional network, this bring a new kind of risk for virtualization platform, namely the management risk. In the virtual world, all the concepts is change from hardware to software, the system administrator of managing the virtual switching network cannot any longer use the simple tools for monitoring and troubleshooting. The administrators cannot approach the virtual machines, plug into a laptop computer, add a network splitter, make a port mapping, or view the statistics of a virtual device. All of these skills and knowledge about virtualization beyond the capabilities of the general network administrator, and the virtualization technique conceal the software controller, GUI (Graphical User Interfaces) of management, dedicated kernel module and the binary systems, so the general network administrator cannot see them. Only the designers, developers and senior administrator of virtualization system know how to implement effective manage, thus increasing the cost and difficulty of management.

VI. Solutions of Virtualization Security Risks

For virtualization security, the appropriate technical measures can be used include: encryption and integrity checking of virtual memory image files, isolation and reinforcement of virtual machines, access control of virtual machines, vulnerability checking of virtualization, monitoring of virtual machine, security migration of virtual machine, and so on. In the face of a wide range of infrastructure, a wide range of services, and large user groups in virtualization environment, the overall strategy of virtualization security solution is to divide and conquer in this paper, namely, according to the user and manipulate objects of different categories take appropriate safety measures.

(1) For the first kind of risks, for from the LAN, WAN (Wide Area Network) and Internet, different users, its access to different positions and different content, we need to use different security policies. This paper considers that control the number of users is not conducive to the maintenance, since the virtualization platform has all the data and applications under the same standard, it will have different access to the unified management of data and applications are classified under the same security zone, and then take security measures to the security zone which can avoid the object is encryption protection for each objection. Directory services can be used to manage identities and provide the capability of access control. When the user needs to access the resource of cloud, one-time grant permission for the client, specify the scope of access, and it can only be displayed to access the secure area, other areas is hidden to prevent the user to access, thus can avoid the user which lack of rights to know the path of secret information.

(2) For the second kind of risks, in the special case when the number of users and applications services requests increased dramatically, we need to adopt the workload equilibrium and migration strategy to move to another work area. At the same time, take audit mechanism for users apply services, review each application to prevent a malicious user to apply a lot of resources on the virtual platform. Taking in emergency situations, the virtual platform is outage and loose data by attack, virtualization platform requires a rapid recovery and return to normal working mechanism, where the strategy is to establish the backup of their own and restore mechanism for each security zone, parallel reconstruct without disturbing each other, to restore service. For the data protection and disaster recovery, we can refer to Symantec's data protection and disaster recovery solutions [44].

(3) For the third kind of risks, in each virtualization area we can setup a separate administrator (can be two processes), a process running in a secure area within the real-time, involved in virus and Trojan killing, as well as bug fixes, etc., while another process in an isolated area, communicate with the outside world, updating the virus and Trojan library and obtain the latest vulnerability information, and download those patches stored in a separate isolated "box", the first process to obtain the information of the "box", and the security is maintained throughout the area.

(4) For the fourth kind of risks, they belong in the virtual platform design problems, added to the virtualization platform security solution does not solve these two problems sometimes. To solve these two problems fundamentally, we need to add some new mechanisms into the virtualization platform design process. Current security policy is mainly audit. In a cloud environment, the physical server is integrated into multiple virtual machines instances on the virtual server. The firewalls, intrusion detection and prevention, integrity monitoring and log checking all can be deployed as a virtual machine as software to increase the protection of virtual machines. Among them, it needs to stress that the log file must be tamper-proof to ensure its integrity and authenticity. We can adopt the digital signature technique, the digital watermarking and other techniques to ensure the integrity and authenticity of the log file. For the virtual machine which must use an embedded hypervisor API, we should customize a security mechanism to monitor the flow of VM backboard, where

the flow is not visible for the traditional network security monitor equipment.

VII. Conclusions

Cloud computing is a rapidly developing emerging industry currently, and it has broad prospects for development, but meanwhile the challenges of security technology they face are unprecedented. The security issue has become a bottleneck restricting cloud computing industrial applications. This article focuses on the important issue of cloud computing, data security. First, we summarize the research progress of issues of data encryption, access control, and integrity authentication and so on with respect to cloud computing data security. On this basis, we point out the key technologies and key issues the cloud computing data security issues should be concerned about, and then give the corresponding countermeasures and suggestions. In addition, for the virtualization security problem of private cloud computing, the security risks induced by virtualization are analyzed and classified, and then based on the results of risk analysis, for each kind of security risk, some corresponding solutions are presented. We believe that based on the proposed solutions in this paper, the security risks of private cloud computing virtualization can be reduced reliably, and the security level of whole private cloud computing can be enhanced. Overall, the research of cloud computing security is at the initial stage of development. In terms of cloud computing virtualization security and data security, there are still a large number of key issues to be studied in depth.

Acknowledgment

This work was supported by National Natural Science Foundation of China (Grant No. 61379151, 61373020 and 61272489), China Postdoctoral Science Foundation (Grant No. 201104911838, 2012T50842).

References

[1] D. Feng, M. Zhang, Y. Zhang, Z. Xu. Cloud Computing Security Research. *Chinese Journal of Software*, 2011, 22(1) pp. 71-83.

[2] Above the Clouds: A Berkeley View of Cloud Computing. http://www.eecs.berkeley.edu/Pubs/ TechRpts/2009/EECS-2009-28. pdf

[3] J. Yao. The Future Needs Cloud Computing. *Development and Application of High Performance Computing*, 2009, 26(1) pp. 7-9.

[4] L. Sumter. Cloud Computing: Security Risk. *Proceedings of ACM Southeast Conference*, 2010, 1-4.

[5] W. Wang, Z. Li, R. Owens, B. Bhargava. Secure and Efficient Access to Outsourced Data. *Proceedings of ACM Cloud Computing Security Workshop*, 2009, 55-65.

[6] W. Zeng, Y. Zhao, K. Ou, Wei Song. Research on Cloud Storage Architecture and Key Technologies. *Proceedings of ACM International Conference on Information System*, 2009, 1044-1048.

[7] H. Takabi, B. D. James, A. Gail-Joon. Security and Privacy Challenges in Cloud Computing Environments. *IEEE Security & Privacy*, 2010, 10(6) pp. 24-31.

[8] S. Pearson, A. Benameur. Privacy, Security and Trust Issues Arising from Cloud Computing. *Proceedings of IEEE International Conference on Cloud Computing Technology and Science*, 2010, 693-702.

[9] Q. Liu, C. Weng, M. Li, Y. Luo. An In-VM Measuring Framework for Increasing Virtual Machine Security in Clouds. *IEEE Security & Privacy*, 2010, 10(6) pp. 56-62.

[10] IBM Research. http://domino.research.ibm.com/comm/ research_projects.nsf/pages/ssd_tvd. index.html.

[11] B. D. Payne. Improving Host-Based Computer Security Using Secure Active Monitoring and Memory Analysis. *A Thesis Presented to The Academic Faculty*, Georgia Institute of Technology, 2010.

[12] P. Bonatti, S. Vimercati, P. Samarati. An Algebra for Composing Access Control Policies. *ACM Transactions on Information and System Security*, 2002, 5(1) pp. 1−35.

[13] D. Wijesekera, S. Jajodia. A Propositional Policy Algebra for Access Control. *ACM Transactions on Information and System Security*, 2003, 6(2) pp. 286−325.

[14] G. Ateniese, R. Burns, R. Curtmola, et al. Provable Data Possession at Untrusted Stores. *Proceedings of 14th ACM Conference on Computer and Communications Security*, 2007, 598–609.

[15] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik. Scalable and efficient Provable Data Possession. *Proceedings of the Conference on Security and Privacy in Communication Networks*, 2008. Doi: 10.1145/1460877. 1460889

[16] A. Juels, B. S. Kaliski. PORs: Proofs of Retrievability for Large Files. *Proceedings of 14th ACM Conference on Computer and Communications Security*, 2007, 584–597.

[17] H. Shacham, B. Waters. Compact Proofs of Retrievability. *Proceedings of ASIACRYPT'08*, 2008, 90-107.

[18] R. Gellman. Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing. Report, February 23, 2009.

[19] T. Schwarz, S. Ethan, L. Miller. Store, Forget, and Check: Using Algebraic Signatures to Check Remotely Administered Storage. *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems*, 2006, 12−21.

[20] A. Yun, C. Shi, Y. Kim. On Protecting Integrity and Confidentiality of Cryptographic File System for Outsourced Storage. *Proceedings of the 2009 ACM Workshop on Cloud Computing Security*, 2009, 67−76.

[21] Q. Wang, C. Wang, J. Li, et al. Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing. *Lecture Notes in Computer Science*, 2009, vol. 5789, 355−370.

[22] K. Zeng. Publicly Verifiable Remote Data Integrity. *Lecture Notes in Computer Science*, 2008, vol. 5308, 419−434.

[23] D. Song, D. Wagner, A. Perrig. Practical Techniques for Searches on Encrypted Data. *Proceedings of the IEEE Symposium on Security and Privacy*, 2000, 44-55.

[24] D. Boneh, G. Crescenzo, R. Ostrovsky, et al. Public Key Encryption with Keyword Search. *Proceedings of the 23rd Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT'04), Lecture Notes in Computer Science*, 2004, vol. 3027, 506-522.

[25] D. Park, K. Kim, P. Lee. Public Key Encryption with Conjunctive Field Keyword Search. *Proceedings of the Workshop on Information Security Applications, Lecture Notes in Computer Science*, 2004, vol. 3325, 73-86.

[26] A. Swaminathan, Y. Mao, G. Su, et al. Confidentiality-Preserving Rank-Ordered Search. *Proceedings of the ACM Workshop on Storage Security and Survivability*, 2007, 7-12.

[27] C. Gentry. Fully Homomorphic Encryption Using Ideal Lattices. *Proceedings of the ACM International Symposium on Theory of Computing*. 2009. 169−178.
[28] D. Stehle, R. Steinfeld. Faster Fully Homomorphic Encryption. *Proceedings of ASIACRYPT 2010, Lecture Notes in Computer Science*, 2010, vol. 6477, 377-394.
[29] M. Dijk, C. Gentry, S. Halevi, V. Vaikuntanathan. Fully Homomorphic Encryption over the Integers. *Proceedings of EUROCRYPT 2010, Lecture Notes in Computer Science*, 2010, vol. 6110, 24-43.
[30] C. Gentry. Toward Basing Fully Homomorphic Encryption on Worst-Case Hardness. *Proceedings of CRYPTO 2010, Lecture Notes in Computer Science*, 2010, vol. 6223, 116-137.
[31] C. Aguilar, P. Gaborit, J. Herranz. Additively Homomorphic Encryption with d-Operand Multiplications. *Proceedings of CRYPTO 2010, Lecture Notes in Computer Science*, 2010, vol. 6223, 138-154.
[32] C. Gentry, S. Halevi, V. Vaikuntanathan. i-Hop Homomorphic Encryption and Rerandomizable Yao Circuits. *Proceedings of CRYPTO 2010, Lecture Notes in Computer Science*, 2010, vol. 6223, 155-172.
[33] K. Chung, Y. Kalai, S. Vadhan. Improved Delegation of Computation Using Fully Homomorphic Encryption. *Proceedings of CRYPTO 2010, Lecture Notes in Computer Science*, 2010, vol. 6223, 483-501.
[34] C. Centry, S. Halevi. Implementing Gentry's Fully-Homomorphic Encryption Scheme. *Proceedings of EUROCRYPT 2011, Lecture Notes in Computer Science*, 2011, vol. 6632, 129-148.
[35] D. Boneh, D. Freeman. Homomorphic Signatures for Polynomial Functions. *Proceedings of EUROCRYPT 2011, Lecture Notes in Computer Science*, 2011, vol. 6632, 149-168.
[36] R. Bendlin, I. Damgard, C. Orlandi, S. Zakarias. Semi-Homomorphic Encryption and Multiparty Computation. *Proceedings of EUROCRYPT 2011, Lecture Notes in Computer Science*, 2011, vol. 6632, 169-188.
[37] Z. Brakerski, V. Vaikuntanathan. Fully Homomorphic Encryption from Ring- LWE and Security for Key Dependent Messages. *Proceedings of CRYPTO 2011, Lecture Notes in Computer Science*, 2011, vol. 6841, 505-524.
[38] J. Coron, A. Mandal, D. Naccache, M. Tibouchi. Fully Homomorphic Encryption over the Integers with Shorter Public-Keys. *Proceedings of CRYPTO 2011, Lecture Notes in Computer Science*, 2011, vol. 6841, 487-504.
[39] N. Smart, F. Vercauteren. Fully Homomorphic Encryption with Relatively Small Key and Ciphertext Sizes. *Proceedings of PKC 2010, Lecture Notes in Computer Science*, 2010, vol. 6056, 420-443.
[40] NIST. http://www.nist.gov/groups/SNS/cloud-computing/index.html
[41] J. W. Rittinghouse, J. F. Ransome. *Cloud Computing: Implementation, Management, and Security*. Published by CRC Press, 2011.
[42] F. Zhang, J. Chen, H. Chen and B. Zang. CloudVisor: retrofitting protection of virtual machines in multi-tenant cloud with nested virtualization. *Proceedings of 23rd ACM Symposium on Operating Systems Principles (SOSP)*, 2011, 203-216.
[43] P. Ferrie. Attacks on Virtual Machine Emulators. White Paper, Symantec Corporation, January 2007. http://www.symantec.com/avcenter/reference/Virtual_Machine_Threats.pdf

Xiangyang Luo received the B.S. degree, the M.S. degree and the Ph.D. degree from Zhengzhou Information Science and Technology Institute, Zhengzhou, China, in 2001, 2004 and 2010, respectively. He has been with Zhengzhou Information Science and Technology Institute since July 2004.

From 2006 to 2007, he was a visiting scholar of the Department of Computer Science and Technology of Tsinghua University. From 2011, he is a postdoctoral of Institute of China Electronic System Equipment Engineering Co., Ltd. He is the author or co-author of more than 70 refereed international journal and conference papers. He is currently an associate professor of Zhengzhou Information Science and Technology Institute. His research interest includes image steganography and steganalysis. He obtained the support of the National Natural Science Foundation and the Postdoctoral Science Foundation of China.

Lin Yang received his Ph.D. from National University of Defense Technology in 1998. Now he is a Ph.D. supervisor of National University of Defense Technology. He is also a researcher of Institute of China Electronic Equipment System Engineering Co., Ltd. He is the author or co-author of more than 80 refereed international journal and conference papers. His research interests include information security and network security architecture.

Fenlin Liu received his B.S. from Zhengzhou Institute of Information Science and Technology in 1986, M.S. from Harbin Institute of Technology in 1992, and Ph.D. from the Eastnorth University in 1998. Now, he is a professor of Zhengzhou Institute of Information Science and Technology. His research interests include network and information security.

Daoshun Wang received his B.S. from Department of Mathematics at LanZhou University, China, in 1987, and a Ph.D. degree from the Department of Mathematics, Sichuan University, China, in 2001. He is currently an associate professor of the Department of Computer Science and Technology, Tsinghua University. His research interests include Information security, visual cryptography and digital watermarking.

Hao Dai received his M.S. degree from Tsinghua University in 1982. He is Chinese Academy of Engineering and a researcher of Institute of China Electronic Equipment System Engineering Co., Ltd. he is also a Ph.D. supervisor of Xi'an University of Electronic Science and Technology. He won the first prize in the National computer Applications, national Science and Technology progress Award in engineering practice. He is the author or co-author of more than 50 refereed international journal and conference papers. His research interests include automated network and network security.