ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

Data Confidentiality and Verifiable User Access Policy for Big data Storage in Cloud

Naveen Ghorpade[1], Dr.P. Vijaykarthik[2]

Assistant Professor, Department of Computer Science & Engineering, Sri Krishna Institute of Technology,

Chikkabanavara, Bangalore, India

Professor & HOD, Department of Information Science Engineering, Bangalore, India

ASTRACT: Cloud Computing has been envisioned as the next-generation architecture of IT Enterprise. It moves the application software and databases to the centralized large data centers, where the management of the data and services may not be fully trustworthy. This unique paradigm brings about many new security challenges, which have not been well understood. This work studies the problem of ensuring the integrity of data storage in Cloud Computing. Dataconfidentiality of access policy are completely ignored in the existing approach and it must important to update the access policy for improving data security and other dynamic activity cause by the user. In the proposed scheme we improve the existing NTRU cryptosystem to overcome the decryption failure of original NTRU. Our schemes facilitate to provide a key to each user in order to make verification of the right user by the data owner who can access the data from the cloud server to avoid the collusion attack and cheating. We improve the NTRU algorithm to increase the data security which is the main issue for the existing NTRU algorithm. Extensive security and performance analysis show that the proposed scheme is highly efficient and provably secure.

I. INTRODUCTION

Big data is an important service of cloud computing, which allows data owners to store a large set of data. More and more owners start to store the data in the cloud .which makes it more voluminous and complex which cannot be managed by the traditional data processing application software which introduces new security challenges .BIG data is a high volume, and/or high velocity, high variety information asset, which requires new forms of processing to enable enhanced decision making, insight discovery, and process optimization. Most existing approaches for securing the outsourced big data in clouds are based on either attributed-based encryption (ABE) or secret sharing. ABE based approaches provide the flexibility for a data owner to predefine the set of users who are eligible for accessing the data but they suffer from the high complexity of efficiently updating the access policy and ciphertext. Secret sharing mechanisms allow a secret to be shared and reconstructed by certain number of cooperative users but they typically employ asymmetric public key cryptography such as RSA for users' legitimacy verification, which incur high computational overhead. Moreover, it is also a challenging issue to dynamically and efficiently update the access policies according to the new requirements of the data owners in secret sharing approaches.

As Big Data expands with the help of public clouds, traditional security solutions tailored to private computing infrastructures, confined to a well-defined security perimeter, such as firewalls and demilitarized zones (DMZs) are no more effective. Using Big Data, security functions are required to work over the heterogeneous composition of diverse hardware, operating systems, and network domains. In this puzzle-type computing environment, the abstraction capability of Software-Defined Networking (SDN) seems a very important characteristic that can enable the efficient deployment of Big Data secure services on-top of the heterogeneous infrastructure. SDN introduces abstraction because it separates the control (higher) plane from the underlying system infrastructure being supervised and controlled. Separating a network's control logic from the underlying physical routers and switches that forward traffic allows system administrators to write high-level control programs that specify the behavior of an entire network, in contrast to conventional networks, whereby administrators (if allowed to do it by the device manufacturers) must codify functionality in terms of low-level device configuration.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

The scheme motivates us to develop a verifiable access control scheme for securing the big data stored in clouds, tackling the challenges of the following security services:

• **Security.** The proposed scheme should be able to defend against various attacks such as the collusion attack. Meanwhile, access policy update should not break the security of the data storage, disclose sensitive information about the data owner, and cause any new security problem.

• **Verification.** When a user needs to decrypt a stored ciphertext, its access legitimacy should be verified by other participating users and the secret shares obtained from other users must be validated for correct recovery.

• **Authorization.** To reduce the risk of information leakage, a user should obtain authorization from the data owner for accessing the encrypted data

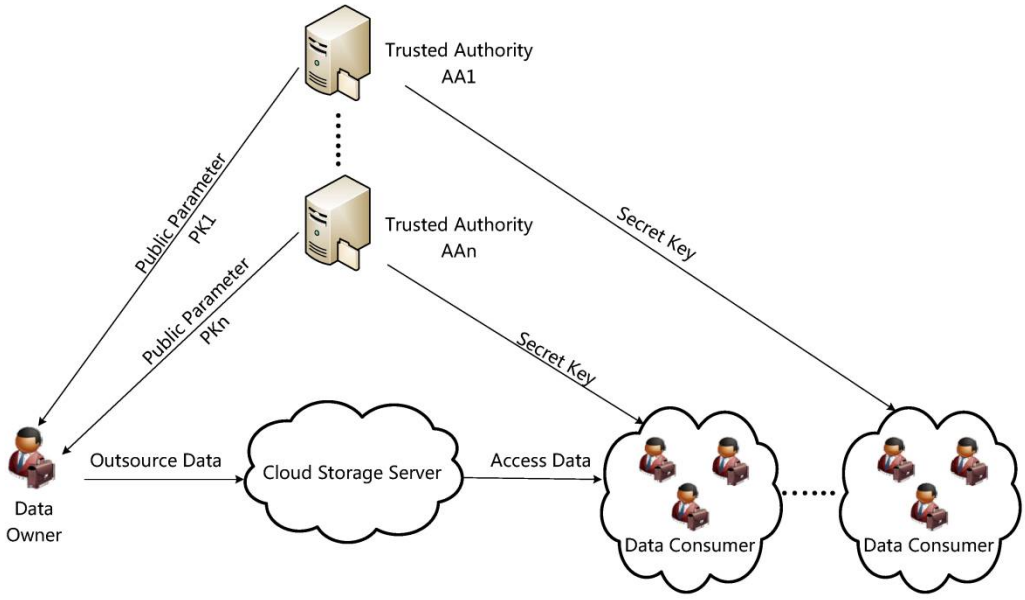
II. RELATED WORK

Because big data containsvoluminous personal identifiable information, it is biggest challenges to provide access control policy secuerly, we mainly summarize the state of-the-art of securing the big data stored in clouds. retrieving to clouds is one of the most effective approaches to securing the big data storage, in which the data get encrypted by owners based on cryptographic primitives and store the encrypted data to the clouds. To contract out a a secure mechanism should be created between a data owner and a cloud. In order for the cloud to perform operations over the encrypted data, "Fully Homomorphic Encryption" (FHE) was usually adopted, which allows direct addition and multiplication operations over the ciphertexts while preserving decryptability. Homomorphic encryption was also applied to guarantee the security of data storage. Nevertheless, it is an immature cryptosystem, and is extremely inefficient in practice, which renders it hardly applicable in real world applications. Securely outsourcing big data computations to the clouds was also extensively studied but this topic is out of the scope of the paper.

Adequate access control is key to protect the stored data. Access control has traditionally been provided by operating systems or applications restricting access to the information, which typically exposes all the information if the system or application is hacked. A better approach is to protect the information using encryption that only allows decryption by authorized entities. Attribute-Based Encryption (ABE) is one of the most powerful techniques for access control in cloud storage systems. In the past years, quite a few attribute-based access control schemes were proposed, in which the data owners define the access policies based on the attributes required by the data and encrypt the data based on the access policies. By this way the data owners are able to ensure that only the users meeting the access policies can decrypt the ciphertexts. However, it is difficult to update the policies when these ABE based schemes are applied because the data owners do not store the data in their local systems once they outsource the data into the cloud servers. It is also difficult to verify the legitimacy of the downloaded data as the clouds housing the data are not trustworthy. Besides, the operations of encryption and decryption in ABE have a high computational overhead and incur a large energy consumption.

IV.MODEL AND PRELIMINARIES

SYSTEM MODEL



ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

Cloud server:-A cloud server provides spaces for data owners to store their outsourced ciphertext data that can be retrieved by the users. It is also responsible for updating the ciphertexts when the data owner changes its access policy.
Owners:-A data owner designates the access policy or its data, encrypts the database don the access policy before outsourcing the data to the cloud server, and requests the cloud server to update the encrypted data when a new access policy is adopted. It can also check whether the ciphertext at the cloud server is correctly updated.
Users:-Each user is assigned with a sub-key for an encrypted data that the user is eligible to access. In order to decrypt the ciphertext, the user's eligibility must be verified by at least t−1 other users that are also eligible to access the data. The information provided by the t − 1 verifiers must be validated by the user for correct message decryption based on the (t,n)-threshold secret sharing.

IV. PRELIMINARIES

We introduce two cryptographic primitive: **NTRUcrypotosystem** and **secret sharing**

THE NTRU CRYPTOSYSTEM

The NTRU cryptosystem is based on the shortest vector problem (SVP) in a lattice that makes it lightning fast and resistant to quantum computing attacks. It has been proved to be faster than RSA.
The arithmetic of NTRU depends on three integer parameters (N, p, q). Let Zq = Z/qZ denote the ring of integers modulo q. The operations of NTRU took place in the ring of truncated polynomials P = Zq[X]/ XN − 1 . In this ring, a polynomial f is defined by its coefficients in the base 1, X, X2 , . . . , XN−1 as
f = (f0, f1, . . . , fN−1) = f0 + f1X + . . . + fN−1XN−1
The addition of two polynomials f and g is defined as pairwise addition of the coefficients of the same degree
f + g = (f0 + g0, f1 + g1, . . . , fN−1 + gN−1),
and multiplication, noted " ∗ " is defined as a convolution multiplication
f ∗ g = h = (h0, h1, . . . , hN−1), with hk = X i+j≡k (mod N) figj .
The Euclidean norm or the length of a polynomial f = (f0, f1, · · · , fN−1) is defined as
kfk = vuut N X−1 i=0 f 2 i
Let B(d)) be the binary set of polynomials defined for a positive integers d as the set of polynomials of R with d coefficients equal to 1 and all the other coefficients equal to 0. The set B(d)) can be written as
B(d) = (f(X) = N X−1 i=0 fiXi ∈ P | fi ∈ {0, 1}, N X−1 i=0 fi = d)
Different descriptions of NTRUEncrypt, and different proposed parameter sets, have been in circulation since 1996. The 2005 instantiation of NTRU is set up by six public integers N, p, q, df , dg, dr and four public spaces Lf , Lg, Lm, Lr.

- N is prime and sufficiently large to prevent lattice attacks.
- p and q are relatively prime numbers.
- q is much larger than p.
- Lf = B(df) is a set of small polynomials from which the private keys are selected.
- Lg = B(dg) is a similar set of small polynomials from which other private keys are selected.
- Lm = Zp[X]/ XN − 1 is the plaintext space. It is a set of polynomials m ∈ Zp[X]/(XN − 1) that represent encryptable messages.
- Lr = B(dr) is a set of polynomials from which the blinding value used during encryption is selected. The key generation, encryption and decryption primitives are as follows:

1. Key generation

- Randomly choose a polynomial f ∈ Lf such that f is invertible in P modulo p and modulo q.
- Compute fp ≡ f −1 (mod p) and fq ≡ f −1 (mod q).
- Randomly choose a polynomial g ∈ Lg.
- Compute h ≡ g ∗ fq (mod q).
- Publish the public key (N, h) and the set of parameters p, q, Lf , Lg, Lr and Lm.
- Keep the private key (f, fp).

2. Encryption

- Represent the message as a polynomial m ∈ Lm.
- Randomly choose a polynomial r ∈ Lr.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

• Encrypt m with the public key (N, h) using the rule $e \equiv p * r * h + m \pmod q$.

3. Decryption

• The receiver computes $a \equiv f * e \pmod q$.

• Using a centering procedure, transform a to a polynomial with coefficients in the interval $-\frac{q}{2}, \frac{q}{2}$.

• Compute $m \equiv fp * a \pmod p$

The decryption process is correct if the polynomial $p * r * g + f * m \pmod q$ is actually equal to $p * r * g + f * m \in Z[X]/\langle X^N - 1 \rangle$, that is without using modulo q. We have

$a \equiv f * e \pmod q$

$\equiv f * (p * r * h + m) \pmod q$

$\equiv f * r * (p * g * fq) + f * m \pmod q$

$\equiv p * r * g * f * fq + f * m \pmod q$

$\equiv p * r * g + f * m \pmod q$.

Hence, if $a = p * r * g + f * m$ in $Z[X]/\langle X^N - 1 \rangle$, then

$a * fp \equiv (p * r * g + f * m) * fp \equiv m \pmod{}$.

AN IMPROVED NTRU CRYPTOSYSTEM

There exist two types of decryption failures in NTRU: the Wrap failure and the Gap failure. In other words, the decryption process of NTRU cannot always output a correct decrypted message.To overcome this problem,we propose an improved NTRU crytosystem in this section. To proceed, we first analyze the reasons of the decryption failures in the original NTRU. When a ciphertext $e = p\varphi * h + m \pmod q$ is received, it is decrypted according to the following two steps:

1) Compute $a = e * f \pmod q$;

2) Calculate $m0 = a * fp \pmod p$.

As analyzed , the NTRU decryption could correctly recover the plaintext m if $a = (a0, a1, a2, \cdots, aN-1)$ remains unchanged after the module q operation in step 1. This indicates that as long as $ai \in (-\frac{q}{2}, \frac{q}{2})$, $i = 0,1,\cdots,N-1$,before the mod q operation, m can be correctly recovered. If there is at least one ai with $|ai| \geq \frac{q}{2}$, the Wrap failure could happen, resulting in $m0 \neq m$ with a high probability; if $\max_{0 \leq i \leq N-1}\{ai\} - \min_{0 \leq i \leq N-1}\{ai\} > q$, the Gap failure could happen, leading to $m0 \neq m$ with a high probability. From the above analysis, one can see that the Gap failure leads to the Wrap failure and that over coming the Wrap failure can also getting rid of the Gap failure.

In order to overcome the Wrap failure, we propose the following improved NTRU decryption algorithm. For $a = e * f = p\varphi * h * f + m * f$, we first figure out $\Gamma = \max\{|\max_{0 \leq i \leq N-1}\{ai\}|, |\min_{0 \leq i \leq N-1}\{ai\}|\}$ in a, and then compute $\tau = b\Gamma q/2c$. If $\tau = 0$, which implies that each $ai \in (-\frac{q}{2}, \frac{q}{2})$ in a, we can decrypt e to get m via the traditional approach; otherwise, we adjust a to obtain $a0 = a - c(1) - c(2) - \cdots - c(\tau)$, where $c(1), c(2), \cdots, c(\tau)$ are the adjusting vectors and $c(j) = (c(j)\,0, c(j)\,1, \cdots, c(j)\,N-1)$ for $j = 1,2,\cdots,\tau$.

V. CONCLUSION AND FUTURE WORKS

In this paper, we first propose an improved NTRU cryptosystem to overcome the decryption failures of the original NTRU and then present a secure and verifiable access control scheme based on the improved NTRU to protect the outsourced big data stored in a cloud. Our scheme allows the data owner to dynamically update the data access policy and the cloud server to successfully update the corresponding outsourced ciphertext to enable efficient access control over the big data in the cloud. It also provides a verification process for a user to validate its legitimacy of accessing the data to both the data owner and $t-1$ other legitimate users and the correctness of the information provided by the $t-1$ other users for plaintext recovery. The security of our proposed scheme is guaranteed by those of the NTRU cryptosystem and the (t,n)-threshold secret sharing. We have rigorously analyzed the correctness, security strength, and computational complexity of our proposed scheme.

Designing a secure, privacy preserving, and practical scheme for big data storage in a cloud is an extremely challenging problem. In our future research, we will further improve our scheme by combining the (t,n)-threshold secret sharing with attribute based access control, which involves an access structure that can place various requirements for a user to decrypt an outsourced ciphertext data in the cloud. Meanwhile, we will investigate the security problems when a data

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

owner outsources its data to multi cloud servers and consider an attribute-based access structure that can be dynamically updated, which is more applicable for practical scenarios in big data storage.

VI.ACKNOWLEDGMENT

This research was partially supported by the National Science Foundation of the US under grants CNS-1318872, CCF-1442642 and IIS-1343976, and the National Natural Science Foundation of China under grants 61672119, 61373027, 61472044, 61272475, 61571049, 61472403, and 61672321, and the Fundamental Research Funds for the Central Universities (No.2014KJJCB32).

REFERENCES

[1] M. A. Beyer and D. Laney, "The importance of big data: a definition," Stamford, CT: Gartner, 2012.
[2] V. Marx, "Biology: The big challenges of big data," Nature, vol. 498, no. 7453, pp. 255–260, 2013.
[3] G. P. Consortium et al., "A map of human genome variation from population-scalesequencing,"Nature,vol.467,no.7319,pp.1061–1073, 2010.
[4] A. Sahai and B. Waters, "Fuzzy identity-based encryption," Advances in Cryptology–EUROCRYPT 2005, pp. 457–473, 2005.
[5] C. Hu, F. Zhang, X. Cheng, X. Liao, and D. Chen, "Securing communications between external users and wireless body area networks," in Proceedingsofthe2ndACMworkshoponHottopicsonwirelessnetwork security and privacy. ACM, 2013, pp. 31–36.
[6] C. Hu, H. Li, Y. Huo, T. Xiang, and X. Liao, "Secure and efficient data communication protocol for wireless body area networks," IEEE Transactions on Multi-Scale Computing Systems, vol. 2, no. 2, pp. 94– 107, 2016.
[7] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in Proceedings of the 13th ACM conference on Computer and communications security. ACM, 2006, pp. 89–98.
[8] B. Waters, "Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization," Public Key Cryptography–PKC 2011, pp. 53–70, 2011.
[9] C. Hu, N. Zhang, H. Li, X. Cheng, and X. Liao, "Body area network security: a fuzzy attribute-based signcryption scheme," IEEE journal on selected areas in communications, vol. 31, no. 9, pp. 37–46, 2013.
[10] A. Lewko and B. Waters, "Decentralizing attribute-based encryption," Advances in Cryptology–EUROCRYPT 2011, pp. 568–588, 2011.