Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2016, Article ID 8789397, 6 pages
http://dx.doi.org/10.1155/2016/8789397

Research Article

Enhancing the Security of Customer Data in Cloud Environments Using a Novel Digital Fingerprinting Technique

Nithya Chidambaram,[1] Pethuru Raj,[2] K. Thenmozhi,[1] and Rengarajan Amirtharajan[1]

[1]*School of Electrical & Electronics Engineering, SASTRA University, Thanjavur 613401, India*
[2]*Global Cloud Center of Excellence, IBM India, Bangalore 560045, India*

Correspondence should be addressed to Rengarajan Amirtharajan; amir@ece.sastra.edu

Received 30 October 2015; Revised 19 February 2016; Accepted 8 March 2016

Academic Editor: Yifeng He

With the rapid rise of the Internet and electronics in people's life, the data related to it has also undergone a mammoth increase in magnitude. The data which is stored in the cloud can be sensitive and at times needs a proper file storage system with a tough security algorithm. Whereas cloud is an open shareable elastic environment, it needs impenetrable and airtight security. This paper deals with furnishing a secure storage system for the above-mentioned purpose in the cloud. To become eligible to store data a user has to register with the cloud database. This prevents unauthorized access. The files stored in the cloud are encrypted with RSA algorithm and digital fingerprint for the same has been generated through MD5 message digest before storage. The RSA provides unreadability of data to anyone without the private key. MD5 makes it impossible for any changes on data to go unnoticed. After the application of RSA and MD5 before storage, the data becomes resistant to access or modifications by any third party and to intruders of cloud storage system. This application is tested in Amazon Elastic Compute Cloud Web Services.

1. Introduction

Cloud computing which is the next stage of Internet evolution ensures scalability and elasticity of the system. If the consumers and businesses are provided with Internet access, they can directly access their personal files from any corner of the world without installation. This technology enables fruitful computing by incorporating data storage, processing, and bandwidth [1]. In the cloud the data always is roaming, and in such a case the data privacy and tamper-resistance are not guaranteed. Even data can be accessed by the third party. There is a need for focusing on preventing data leakage, notification for security accident, and security incident audits. With the available conventional methods such as firewalls, security policies, and Virtual Private Networks (VPN) cloud security needs to be enhanced to get a tamperproof fertile service from it.

Security is governed by its three important aspects of confidentiality, integrity, and availability, which are the building blocks in constructing a highly secured system. These aspects ensure the security of the data, hardware, and software resources. Integrity checks data tampering. Availability is to ensure the users can use them at any time and at any place and, also, guarantee that resources for processing data and also services are available [2]. Mission critical systems availability is achieved by business continuity plans (BCPs) which are used to ensure redundancy. Confidentiality is defined as providing access only to well-authorized persons. If an unauthorized person is able to access the data, then the loss of confidentiality occurs. It can happen in both the ways either by electronic means or by the person. If the loss takes place through the communal issue, then it is called physical confidential loss. If the loss occurs while the clients and servers are not encrypting their communications, then it is known as an electronic confidential loss. RSA is chosen to achieve confidentiality [3]. Depending on the parameters called confidentiality, integrity, and availability of the consumer data it can be classified and identified as sensitive data or not.

Security and privacy are the main keys to the achievement of cloud computing and at the same time they are the most challenging issues. Recording the ownership and the data log history leads to efficient data forensics. In a cloud system,

confidentiality on the sensitive documents stored is improved by the provenance scheme called digital fingerprinting [4]. Digital fingerprint images of the users are processed to provide highly secure cloud computing. Hash value or message digest is simply a number formed from a text or string which will be smaller than the text. It is generated by a well-defined formula such that the probability of the two strings generating the same hash values is extremely low.

Hash values are used to ensure unmodified data which in turn result in a secured system. The sender generates the hash value of the message, encrypts it, and sends it along with the message. The receiver then decrypts both the message and the hash and generates another hash from the decrypted message. If both the hashes are the same, then it shows that the data is not modified during the transmission. Hash generates a 128-bit digest for the message articulated in the form of text. The MD5 approach is proposed for the same.

This paper contains 4 more sections. Section 2 is all about the literature work on the area of cloud security issues and cryptographic countermeasures. In Section 3, there is the proposed architecture for achieving the data confidentiality together with integrity verification. Section 4 encompasses the outcome of the methodology. Section 5 is the conclusion of this paper.

2. Related Work

Cloud is a shared and automated environment, which offers various services to the user. Services from the cloud are software, hardware, storage, and so forth. Based on the demand from the user cloud is scalable and chargeable too. The best part in the use of cloud is access to your concealed data from anywhere at any point in time with reliability. Because of this characteristic any organization can get into cloud particularly for the storage of data offered as storage as a service by Cloud Service Provider (CSP). In this circumstance unauthorized access needs to be blocked and also impenetrable security is needed for the data stored in the cloud. Security taxonomy for cloud and the noncloud data center is alike. Based on the domain the security frameworks are deployed.

Cryptographic techniques are used to prevent the data [5]. Nowadays medical field also transformed into e-field which means that patient records are stored in the database. Telemedicine field needs support of cloud computing where exactly mobile medicine is possible. Whenever people need to get any suggestion from a physician at any time it is possible through the cloud. But the major challenge is patient privacy; there is a need to protect data from the malicious insiders [6]. Even in business sectors also, malicious insiders who are the employees of the organization as well as the real threats to the data that belong to the organization are there. Account or service hijacking is nothing but a critical area of cloud that can be accessed [7].

Generally, in the cloud, three different elements are under threat, namely, architecture, compliance, and privacy. A data security issue comes under the privacy facet, where cryptographic algorithm always supports protecting the privacy of data available in the cloud [8]. Multilevel and factor approach is needed considering the brewing data security breaches in cloud infrastructures so hybrid encryptions can be adopted [9]. Access control of data is also one of the major issues in the cloud. If the access permission limit is set by the central authority then the trustworthy component is questionable and also not viable. The client needs an assurance that there is no collusion attack also when the data stored in the cloud is based on the identity of the data owner [10].

In addition confidentiality, integrity, and availability (CIA) factors of data are being compromised. CIA can be fortified using cryptography schemes. Symmetric, asymmetric, and hashing algorithms together contributed CIA to data. SSL (Secure Socket Layer) encryption for data confidentiality as well as MAC (Message Authentication Code) for verifying integrity of data is used to ensure CIA together with well-defined access privilege of data placed in the private or public sector of cloud [11]. Even though the owner of data is defining private or public data, the location of data in cloud is unknown and also data is in roaming state. Because of this reason integrity of data must be verified periodically [12].

Currently, data is managed and processed by a mixture of service providers so equally risk factors also increased such as unauthorized access of user sensitive data by different sectors, including service providers. So right from selection of service providers for the respective services till cryptographic techniques for protecting data from the service provider are the responsibility of the user [13, 14].

3. Proposed Methodology

This proposed model is to improve the cloud data security by incorporating various cryptographic techniques. To provide airtight confidentiality, encryption and decryption modules are added. In addition integrity of data is also verified using message digest. This is purely client side security, considering client in the cloud environment. This approach is also a reversible process. The encryption algorithm here cannot be broken easily including the popular dictionary attack. Brute force attack is also difficult to perform. The main reason for making this algorithm client side is to have the self-satisfaction and to ensure security for the clients of the cloud. Even though cloud is not trustworthy using this proposed method data can be stored in the cloud firmly. The proposed architecture is shown in Figure 1.

Data Privacy Module. Public Key Cryptography RSA is being used. In Public Key Cryptography the key for both the encryption and decryption process must be a different key. Here public key for encryption and by private key decryption is done. So this can be an appropriate algorithm for the cloud environment.

RSA: Key Generation Module

Pseudo Code. Consider the following:

(1) Select two random prime numbers which should be distinct. Assign variables to the prime numbers (a and b). Both should have similar bit length.

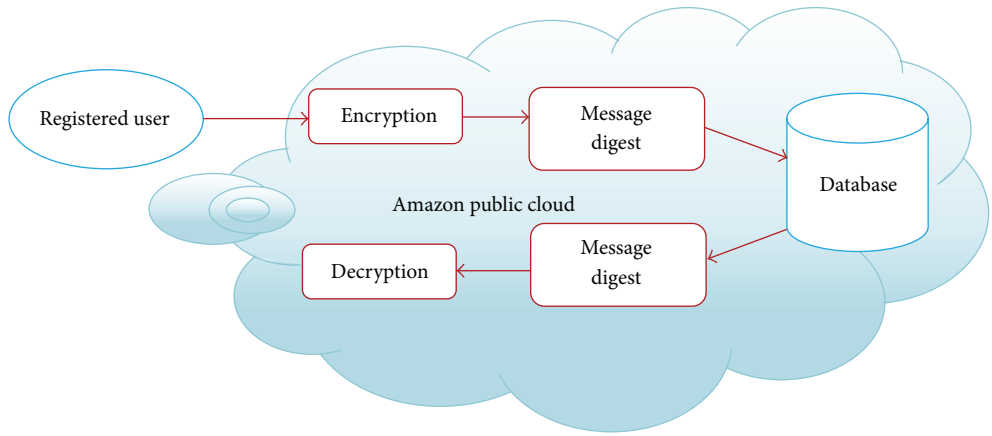

FIGURE 1: Proposed architecture.

(2) Calculate $n = ab$, where n is the number used as the modulus for private and public keys. Key length is equal to "n" length.

(3) Now calculate $\psi(n) = (a-1)(b-1)$.

(4) Select an integer K_e. It should satisfy $1 < K_e < \psi(n)$. GCD of p and $\psi(n)$ should be equal to 1; that is, K_e and $\psi(n)$ are coprime. Now K_e is the public key component.

(5) Compute K_d such that $K_d \cdot K_e \equiv 1 (\text{mod } \psi(n))$. K_d is the private key exponent. It should be kept secret.

(6) K_e and n are announced to the public for encryption.

(7) For decryption, ciphertext, K_e, and modulus n are used. By using q and $\psi(n)$ decryption key is calculated called private key K_d and must be kept secret.

Message Digest Generation. MD5 hash function algorithm has been used for the purpose of generating the message digest. The MD5 message digest algorithm generates 16-byte hash value in text format as 32-digit hexadecimal number. MD5 has been employed in a variety of areas, mainly used to check the data integrity in the cryptographic domain. Cryptographically strong digest algorithm generates a nearly unique digital fingerprint value from any source string [14]. Small change in the message leads to the predominantly different hash. MD5 even produces a hash for zero-length string.

Pseudo Code. Consider the following:

(1) The input message is divided into blocks of 512 bits. If the total number of bits is not the multiple of 512, then the padding of bits will be done.

(2) Padding is done in the following format:

First, single bit "1" is appended to the end and zeros are padded to make message length as 64 bits less than the multiple of 512.

(3) The last block represents the original message length.

(4) The MD5 algorithm utilizes 4 chaining variables of 32-bit length.

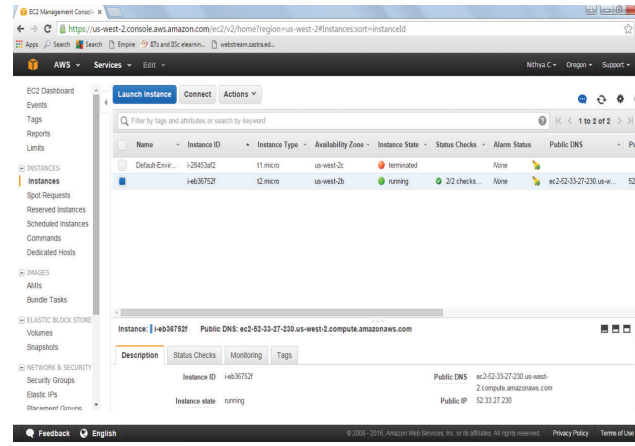

FIGURE 2: EC2 dashboard in Amazon Web Services (AWS).

(5) There are four main functions which use the above state variables and input message to produce the message digest. The functions are as follows:

$F(B, C, D) = (B \ \& \ C) \mid (\sim B \ \& \ D)$,

$G(B, C, D) = (B \ \& \ D) \mid (C \ \& \ \sim D)$,

$H(B, C, D) = B \wedge C \wedge D$,

$I(B, C, D) = C \wedge (B \mid\sim D)$.

Note: &, |, $\wedge$, and $\sim$ denote bitwise AND, OR, XOR, and NOT operations, respectively.

The above four functions are applied to all the individual 512-bit blocks. Finally, the digest is stored in the variables A, B, C, and D.

4. Results and Analysis

Various services to the registered users are offered by Amazon Web Services (AWS). In this proposed approach, EC2 service is used after registering in AWS. A remote machine was launched with the help of the Amazon EC2 service for the required configuration as shown in Figure 2.

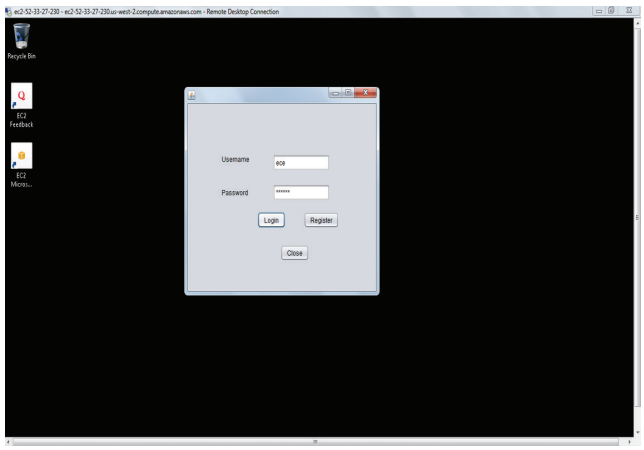

Figure 3: Application run on remote machine.

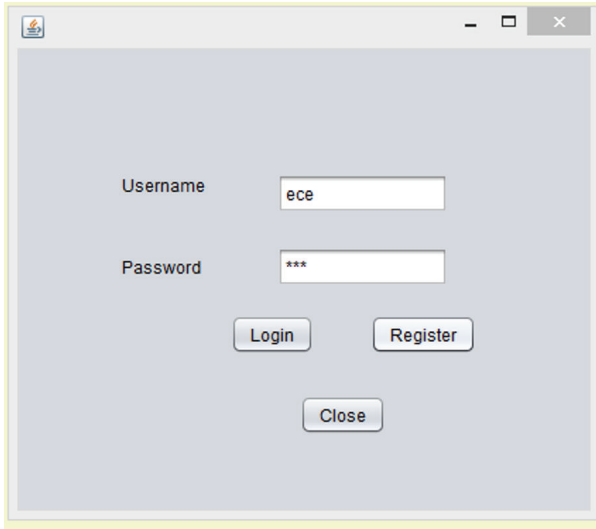

Figure 4: Registration.

The process is carried out in seven steps:

(1) User login with registered credentials.

(2) Encryption.

(3) Message digest generation.

(4) Saving the encrypted file.

(5) Retrieving the encrypted file.

(6) Checking integrity using message digest.

(7) Decryption.

4.1. User Registration. Using remote desktop connection, the remote machine was accessed and the developed application was implemented using JAVA platform as shown in Figure 3.

After giving username and password register request should be made. Before inserting the username and password in the database, a check is run in the database for any account with the mentioned username. This step is done to ensure unique usernames for every account as this column acts as the primary key in the database table. After logging into the database using the username and password, a message will be displayed as shown in Figure 4.

4.1.1. User Login. A user has to register with the system database to become eligible to store data. This prevents unauthorized access. While registering the user has to select a username and password which had not been taken before and can set any password of his choice as shown in Figure 5. After registering with a valid username, it allows the user to do the desired task in the future such as selecting the file from the drive for the encryption process.

4.1.2. Saved in Database. Figure 6 shows the username and the entered password's hash function, which is stored on the server. The entered password will be converted into its hash value using the MD5 algorithm.

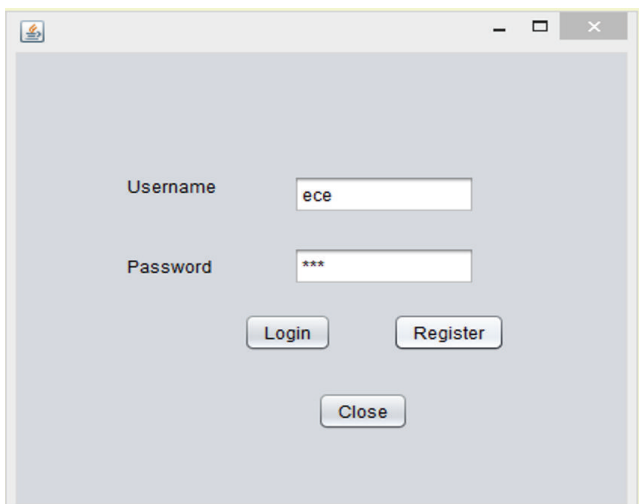

Figure 5: Login.

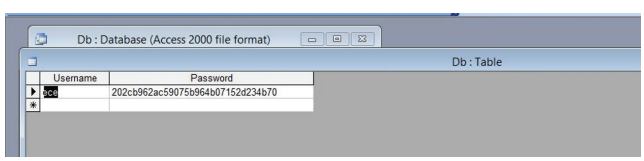

Figure 6: Login details in database.

4.2. Selecting a File. As a next step the user can select the file to be stored in the cloud. The file must be encrypted to maintain its confidentiality using RSA. From Figure 7 it is clear that the user can select a file by browsing and then encrypt it.

4.3. Encrypt and Save. Once the file is selected the user needs to click the encrypt button to encrypt the file using RSA

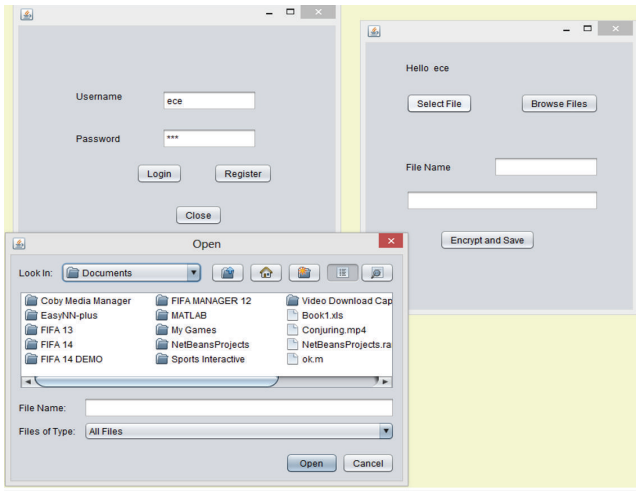

Figure 7: Selecting a file.

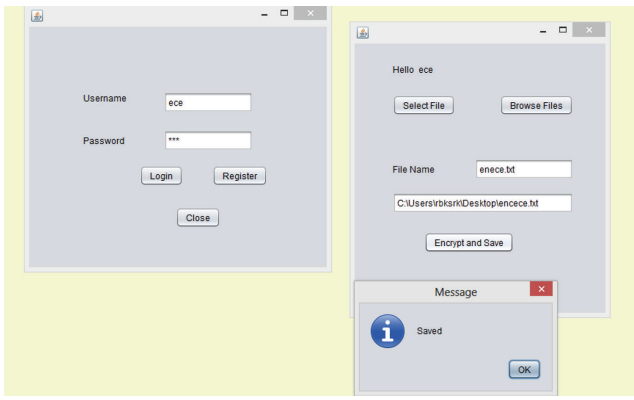

Figure 8: Encryption.

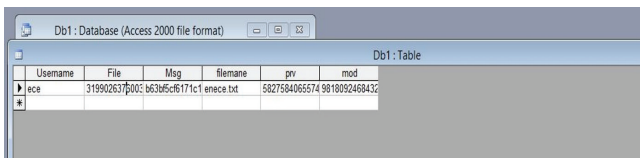

Figure 9: File in database.

algorithm and message digest function. After encryption, the file is saved in the database available in the cloud as shown in Figure 8.

4.3.1. Saved in Database. Now in the next database, along with the username, hash function of the password, file data (encrypted), keys used for encryption and decryption are saved as shown in Figure 9.

4.4. Decryption. Supposing that the user needs to view the file saved under the login provided, the user can select the file and download it to the specified place. The message digest is generated and at the same time it is being verified for ensuring the file integrity. The downloaded file is decrypted one as shown in Figure 10.

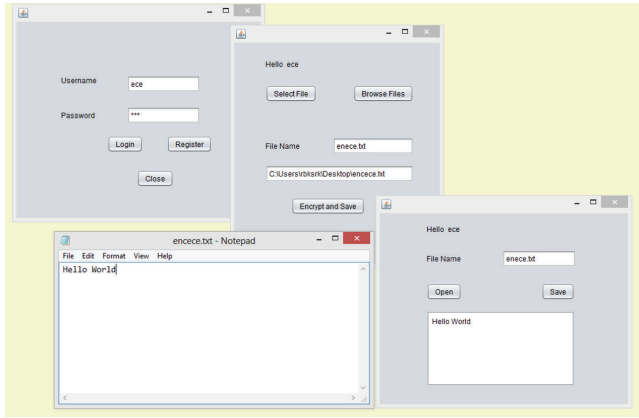

Figure 10: Decryption and saving the file.

4.5. Performance Analysis. In this proposed method public key infrastructure (RSA) is used where two different keys are used one for encryption and the other for decryption. To break the crypt system factoring "n" is needed, where n is the product of two large primes. RSA system is difficult to hack because guessing of two large prime numbers in the key space is complex.

With MD5 algorithm faster avalanche effect is achievable; that is, small change in the message leads to the predominantly different hash. MD5 even produces a hash for zero-length string.

Here the length of hash is 128 bits, so, for birthday attack, 264 random documents need to be tried [14].

In this proposed approach both RSA and MD5 features were combined together so complexity increased towards hacking. So, compared with the available literature, the maximum level of privacy with tamper-resistance for file storage in the cloud was achieved.

The proposed approach is compared with the existing system based on functions and the findings were shown in Table 1.

5. Conclusion

Thus, we have designed a secure file storage system which incorporates authenticity, ingenuousness, and confidentiality. It is implemented in almost everything which required granting permission to only authorized credentials. The password is stored in the database as a message digest. This kind of password storage is really tamperproof. The encryption process makes the data secure; it prevents readability by unauthorized personnel and also establishes a framework to remove imposture. The system has been designed in such a way that this one-way hash function, if even cracked, will lead to getting encrypted data only. The application of hash function makes it impossible for any changes on data to go unnoticed. After the application of RSA and MD5 before

Table 1: Performance analysis.

Functions	Guo et al. (2013) [9]	Han et al. (2013) [10]	Sheu et al. (2014) [1]	Park (2015) [4]	Proposed approach
Identification & authentication	No	Yes	No	Yes	Yes
Authorization	No	Yes	No	Yes	Yes
Confidentiality	Yes	Yes	No	Yes	Yes
Integrity	No	No	Yes	No	Yes
Antiattack capability	Weak	Medium	Weak	Medium	Strong

storage, the data becomes resistant to access or modifications by any third party and to the storage system.

Competing Interests

The authors of the paper do not have a direct financial relation with the commercial identity mentioned in their paper that might lead to competing interests for any of them.

References

[1] R.-K. Sheu, S.-M. Yuan, W.-T. Lo, and C.-I. Ku, "Design and implementation of file deduplication framework on HDFS," *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 561340, 11 pages, 2014.

[2] Y. Sun, J. Zhang, Y. Xiong, and G. Zhu, "Data security and privacy in cloud computing," *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 190903, 9 pages, 2014.

[3] S. Iyer, "Cyber security for smart grid, cryptography, and privacy," *International Journal of Digital Multimedia Broadcasting*, vol. 2011, Article ID 372020, 8 pages, 2011.

[4] S. B. Park, "Security requirements for multimedia archives," *Advances in Multimedia*, vol. 2015, Article ID 956416, 5 pages, 2015.

[5] S. Balasubramaniam and V. Kavitha, "A survey on data encryption tecniques in cloud computing," *Asian Journal of Information Technology*, vol. 13, no. 9, pp. 494–505, 2014.

[6] C.-L. Chen, T.-T. Yang, and T.-F. Shih, "A secure medical data exchange protocol based on cloud environment," *Journal of Medical Systems*, vol. 38, no. 9, article 112, 2014.

[7] D. Zissis and D. Lekkas, "Addressing cloud computing security issues," *Future Generation Computer Systems*, vol. 28, no. 3, pp. 583–592, 2012.

[8] N. Gonzalez, C. Miers, F. Redígolo et al., "A quantitative analysis of current security concerns and solutions for cloud computing," *Journal of Cloud Computing*, vol. 1, no. 1, pp. 1–18, 2012.

[9] P. Guo, L. Su, L. Ning, and G. Dan, "Hybrid encryption algorithms in cloud computing," *Information Technology Journal*, vol. 12, no. 14, pp. 3015–3019, 2013.

[10] J. Han, W. Susilo, and Y. Mu, "Identity-based data storage in cloud computing," *Future Generation Computer Systems*, vol. 29, no. 3, pp. 673–681, 2013.

[11] J. Li, J. Li, Z. Liu, and C. Jia, "Enabling efficient and secure data sharing in cloud computing," *Concurrency Computation Practice and Experience*, vol. 26, no. 5, pp. 1052–1066, 2014.

[12] S. K. Sood, "A combined approach to ensure data security in cloud computing," *Journal of Network and Computer Applications*, vol. 35, no. 6, pp. 1831–1838, 2012.

[13] S. Tan, L. Tan, X. Li, and Y. Jia, "An efficient method for checking the integrity of data in the cloud," *China Communications*, vol. 11, no. 9, pp. 68–81, 2014.

[14] B. Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, John Wiley & Sons, 2nd edition, 2001.