Cloud Computing: A Survey on Security Issues and DNA, ID-base Cryptography

R. Thilagavathy* and A. Murugan

Department of Computer Science, Dr. Ambedkar Government Arts College (Autonomous), University of Madras, Chennai - 600039, Tamil Nadu, India; thilagmca@gmail.com, amurugan1972@gmail.com

Abstract

Background: Cloud computing is a model whereby users access compute resources from a service provider through internet. The Cloud Computing has great capability to boost productivity and minimize cost, hence many organizations are migrating to the new technology, but at the same time it constitutes many security risks and challenges. Methods: DNA and ID base cryptography methods are used for cloud computing security. Findings: Cloud computing can address the problems when user moves from single cloud to multi-clouds. In this paper, we provide an extensive survey of cloud computing security challenges and different technologies taken to tackle the issues. Improvements: Highlighting the security issues of cloud computing, the proposed system is, DNA technology and JPEG Zigzag coding with an encryption scheme. We conclude the paper with a critical analysis of challenges that have not yet been fully met, and highlight directions for future work.

Keywords: Cloud Computing, Cloud Security Risks and Challenges, CSP, Cryptography, DNA, Encryption Scheme, Privacy, Security

1. Introduction

Cloud Computing is the consumption model required for easy resources and content sharing. NIST\textsuperscript{1} defines cloud is a utility computing model used for convenient on demand network based access. It is used to a shared pool of resources that can be rapidly provisioned and realized with minimal management effort of service provider interaction. The utility model proposed by NIST is given in the Figure 1.

Cloud computing has many advantages but at the same time it compromise security challenges and possess risks. This model satisfies a broad range of purposes and meets various requirements. The goal of cloud computing is resource sharing and low cost facilities.

Today many service providers are using this model for convenient delivery of content. The Google mail and web based document management are the good examples for mail service providers. Cloud computing applications are possible in internet and web engineering\textsuperscript{3}.

The main advantages of cloud are the creation of healthy information technology infrastructure and providing maximum flexibility, efficiency, reliability and more user convenience. Many author’s discussed about the security related risk factors in cloud computing\textsuperscript{2,3} and following are the common security challenges discussed in\textsuperscript{1},

- Data
- Infrastructure
- platform and Hosted Code
- Resource Access
- Compliance
- Third-party control of data
- Multi-party processing
- Content sharing

Various approaches that have been used to mitigate security issues in cloud computing will be discussed in the Literature Review section. The paper is organized as follows: Section 1, discussion about the cloud definition

*Author for correspondence
Cloud Computing: A Survey on Security Issues and DNA, ID-base Cryptography

and cloud security concerns. Section 2 provides survey of cloud computing security challenges and approaches. We conclude by summarizing the paper’s contribution and scoping further work.

Many models and architecture are found in the Literature Review. We discuss about the technological implications of each of these concerns and related research issues.

The authors proposed an identity based cryptographic approach that prefers other identity based cryptography system over trust hierarchy system.

In order to avoid the problems arising from certificate based infrastructure, this model uses a identity based cryptographic (IBC) system. It uses any string as a public key. This model removes the need of heavy weight certificate infrastructure and provides intuitive security environment.

The proposed system would enable cloud resource providers (CSP) to have multiple security domains rather than single root trust, which would operate independently to each other. This system can be applied to service-oriented cloud computing environment. This approach also saves bandwidth by using two approaches,

1. Certificate approach
2. Identity-based cryptography

Since the security domains are completely independent to each other and does not have a trust hierarchy, this system addresses the problems of service-oriented applications of cloud computing. This model can handle network load efficiently.

The authors describe a Signal Processing in the Encrypted Domain (SPED) technique. The SPED is used for efficient storage with privacy. Author introducing the architecture called virtualized cryptoDSPs for cloud scenarios to perform needed operations.

They said that the Virtual CryptoDSP is used for signal processing and materializing the signal processing using virtual DSP (Digital Signal Processor).

Figure 2 presents the proposed architecture of Virtual CryptoDSP, This Virtual CryptoDSP architecture for the final user is transparent. This architecture has three main blocks, they are

1. Virtualized CryptoDSP core
2. Client Plug-In
3. Virtualized Coded Storage Module

This technology is used in cloud infrastructure (IaaS) to produce maximum versatility. These three above mentioned elements are implemented as middleware of IaaS. Finally, author said, the cryptography and signal processing is a novel interdisciplinary field for providing efficient privacy preserving signal processing in the Encrypted Domain.

The SCMCS model is used for cost effective storage of data in multi cloud environment with high security. This model used by service providers to provide better data availability to the customer in the secured manner.

The customers can also take better decisions about secure storage according to their available budgets.

Author says that preservation of data privacy and data integrity are important security concerns regarding user data. The cloud service provider can get new approaches for hiding the data and for data privacy. All these approaches are based on cloud data with centralized distribution.

In their proposed model they used redundant distribution scheme, such as, in which at least a threshold number of pieces of data are required out of the entire distribution range, for successful retrieval.

Figure 1. Cloud computing model by NIST.

Figure 2. Diagrammatic representation of Virtual CryptoDSP.

---

1. [Figure 1. Cloud computing model by NIST]
2. [Figure 2. Diagrammatic representation of Virtual CryptoDSP]
Finally author said that SCMCS model is used for the multi cloud storage in cloud computing. This model is used to distribute the economical based data distribution. This provides customer a high security and availability.

The authors described an algorithm which introduces data hiding concepts in DNA sequences in order to decrease the rate of complexity and to increase the confidentiality in cloud environments\textsuperscript{15}. This proposal is based on complimentary pair rules and binary coding rules. This work is based on DNA sequences used in the previous research\textsuperscript{16-20}.

The biological characteristics of DNA sequences are utilized in the proposed work. By using binary pairing rules the binary data is converted in to amino acids as a DNA sequences. This work has two phases. First phase is embedding phase and the next is extracting phase. In Embedding phase, the secret data is embedded with base paring rule. In Extracting phase, the original data is extracted from the embedded form.

The author says that each violator must be aware of the following information.

1. DNA reference sequence
2. Binary coding rule
3. Complimentary pairing rule

Without the above mentioned information, the possibility of stealing the original data is zero. They said that the probability of making a successful guess by violator is,

\[
\frac{1}{163} \ast \frac{1}{24} \ast \frac{1}{24}
\]

Finally, the author concludes that getting original data in the DNA sequence is not easy because of the visibility of the sequence is very low.

The author describes technique using DNA sequence with numbers\textsuperscript{21}. The numbers are random numbers which is used to provide high data security. Author said function is a random element and it is a single outcome. This single element is selected from some families of functions. The family should consist of some class of all maps. This class may be from the domain to the co domain.

The random function gives the value which is from the same realization and evaluated at different points. The values stated to be independent which is depending on the model\textsuperscript{22}.

Author suggests researchers to reduce computation complexity and to avoid unwanted intruders to hack data they should concentrate on crypto system, because cryptography is the active research area which is used for security.

They said for each cipher that crypt-analyst breaks, a few more seem to sprout up in its place. Even though the recent development in algorithmic might obsolete the cryptosystem overnight, this field will continue to survive due to its breath and diversity\textsuperscript{23}.

The important part of their work is utilizing the biological properties of DNA sequence. They use the amino acids to convert binary data and the converted amino acids are taken as a DNA sequence and synthesizing nucleotides in real environment using base pair rule. This pair rule is also called constant rules\textsuperscript{15}.

Here author explains that, when binary representation of A, T, G, and C are used finding possibility is difficult. They used random number key value, binary arithmetic operations to find exact probability.

Finally, author concludes that their work focuses on the data security issues and usage of encryption and decryption method. They used random number keys to provide a secure and effective encryption and decryption methods.

DNA cryptography using chip based micro array technology is introduced\textsuperscript{24} in cloud computing the cryptography algorithms are used to provide data security. For Cryptography the DNA sequences are used to store and transmit the information as well as for computation. The author recommends that the theoretical analyzes should be performed before its real applications because it requires high tech lab requirements and computational limitations.

The author says that chip based micro array technology is generally expressed with two conversions.

- Plain text to cipher text conversion called Encryption.
- Recovery of plain text from Cipher text called decryption.

The author uses 2D image encryption for their process and also suggests that the Steganography is the best technique when using DNA computing. The improved security is also suggested by the researchers\textsuperscript{25,26}. They used symmetric key algorithm key principally.

The author says cryptography method is not a constraint only to the encryption and decryption area, but also to the message authentication. If hash function
method is used, the powerful encryption scheme is created. This scheme is used when generating message authentication code (MAC).

The author says that the security is satisfactory when multiple rounds are used. But the problem is, length of the authentication code is hard to control, when multiple rounds are used.

The author proposed a solution to the problem; getting MAC with fixed length by performing multiple rounds of process together with padding. The Steganography can be implemented when real data is used. This is specially used for image Steganography.

An efficient program can be created in C or assembly language as only four nucleotides are involved in encipher and decipher process of plaintext.

In author's point of view, the hardware can be easily implemented based on these simple principles. This model is very efficient and the technique can be implemented for confidentiality, integrity and authentication in cloud systems.

The program creates DNA sequence using user id and shift key. This mechanism will assign four characters DNA key “ACTG” for each user id and it will produce shift key randomly from 1 to 16. It can produce 16 combinations that is shifted based on the key.

They said that the system can generate more than 163 million unique DNA reference sequences. Now the conjecture by attacker is 1/24. The binary code rules are 4x3x2x1=24. So correct guess by attacker is 1/24 and the possibility of getting a correct guess by intruder is same as the probability which is mentioned in 15.

The authors explain that when dividing data into several pieces, the availability and better privacy can be achieved. After that the data pieces are distributed among different Service Providers (SP). The whole data block can be retrieved only when minimum threshold Service Providers take part in the retrieval process.

This model provides customer to make better cloud data storage decision according to their budget concern.

They say customer's data stored at cloud service provider is vulnerable to various threats. They suggest two types of threat models. They are,

1. Single point of failure- When the cloud service provider's server crashes, it would be difficult to retrieve the customer's data. One of the solutions for this threat is to store the data at multiple service providers to ensure availability of data.

2. Colluding service providers- User stored data is accessed and reconstructed by cloud service provider when they collude together.

Finally, this model describes the solution to provide the customer's data with a secured storage economically, by dividing and distributing customer’s data. Secured cost-effective multi cloud storage (SCMCS) enables customer to make better data storage decisions. This system allow user to consider about the budget as well as quality of service.

The authors defined DNA based cryptography using Watson- Crick Base Pairing rules. They demonstrate powerful security strategy of using DNA cryptography. Many companies prefer Multi-Cloud usage as the usage of single cloud is intolerable because of risks such as service outage, data leakage, theft of data and the chances of malicious inside attacker.

The following universal rule explains the basic synthesis of nucleotides in real environment.

- Purine Adenine (A) always pairs with Pyrimidine Thymine (T)
- Pyrimidine Cytosine (C) always pairs with Purine Guanine (G).

In this model, the critical data is uploaded to cloud by the client within a company using cloud environment.

Critical data can be processed in two phases,

- Data Embedding.
- Data Extracting.

In Data Embedding, the original data should be uploaded to cloud. Binary code rule is applied and executed using DNA sequence. Binary data will then be converted into DNA nucleotides. The base pair rule is applied to DNA nucleotides and those indexes of Nucleotides in DNA reference sequence are called Cipher text.

In Data Extracting, the Cipher text conversion is carried out by applying base pair rule in reverse way and after that the code will be converted into DNA sequence. Later, using binary coding rule the code will be converted to original binary data.

The authors said that in multithread implementation, the data will be initially divided into four pieces and stored in four cloud service providers and then work can be extended to 8 and 12 CSP's. The system performance is analyzed in terms of data size and execution time with graphical representation. The authors mentioned that the
Table 1. Technologies and approaches used in cloud computing security

<table>
<thead>
<tr>
<th>S.no</th>
<th>Author's Name</th>
<th>Month &amp; Year</th>
<th>Technology Used</th>
<th>Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Christian Schridde et al.⁴</td>
<td>June 2010</td>
<td>Identity-Based Cryptography (IBC) system.</td>
<td>In certificate based security infrastructure it is used to avoid the management problem.</td>
</tr>
<tr>
<td>2.</td>
<td>Juan Ramon Troncoso-Pastoriza et al.⁵</td>
<td>Dec 2010</td>
<td>Novel discipline of Single Processing out Sourcing through a Virtual CryptoDsp.</td>
<td>Provides efficient privacy-preserving signal processing.</td>
</tr>
<tr>
<td>3.</td>
<td>Yashawi Singh et al.⁷</td>
<td>May 2011</td>
<td>Multi-Cloud Storage (SCMS) Model with security and minimum cost.</td>
<td>This model gives high data availability and security.</td>
</tr>
<tr>
<td>4.</td>
<td>Mohammad Reza Abbasy et al.¹²</td>
<td>June 2011</td>
<td>Data hiding in DNA sequence using binary coding and paring rule.</td>
<td>Used to increase the confidentiality and reduce the complexity.</td>
</tr>
<tr>
<td>5.</td>
<td>Siddaramappa V ¹⁸</td>
<td>July 2012</td>
<td>Data hiding in DNA sequence DNA Substitution and One-time pads.</td>
<td>To provide data security random numbers are used in encryption and decryption scheme.</td>
</tr>
<tr>
<td>7.</td>
<td>D. Sureshraj et al.²³</td>
<td>Nov 2013</td>
<td>User id and Shift key are used.</td>
<td>Used to provide a user better decision about data storage and quality of service.</td>
</tr>
</tbody>
</table>

possibility of making a correct guess by intruder is same as the probability which is mentioned in¹⁵. They developed the model using Microsoft Visual Studio 2010, MS SQL 2008 tools on Windows8 operating system. Programming language used is ASP C#.Net. Minimum hardware requirement is Processor Pentium-IV, RAM of 2GB and 80 GB HDD.

Finally, they concluded that splitting users critical data into different parts and then applying powerful DNA based Cryptography to each part of data will provide cloud user with more secured storage and security. An overview of technologies used and their advantages, of different approaches in cloud computing security is illustrated in Table 1.

2. Conclusion and Future Work

Cloud Computing provide us cheaper, faster, flexible, effective environment. The government, universities, educational institute, research center can get benefit from this model. But security has become a major issue. We have given an extensive survey of cloud computing security issues. Highlighting the security issues of cloud computing, the propose system is, DNA technology and JPEG Zigzag coding ³⁰ with an encryption scheme. This encryption scheme is used for secure Transmission of images with privacy and security.

3. References


