Fuzzy Search Over Encrypted Data in Cloud Using AES Algorithm

Asttle .J[1], Gurupandyan .K[2], Manikandan .T[3], Vijayanand .M[4], Mrs.Sheelavathi .A[5]
Dept of Information Technology
Saranathan College of Engineering.

Abstract- *High-speed networks and ubiquitous Internet access become available to users for access anywhere at any time. Cloud computing is a concept that treats the resources on the Internet as a unified entity, a cloud. Cloud storage is a model of networked online storage where data is stored in virtualized pools of storage which are generally hosted by third parties. Hosting companies operate large data centers, and people who require their data to be hosted buy or lease storage capacity from them.The data center operators, in the background, virtualize the resources according to the requirements of the customer and expose them as storage pools, which the customers can themselves use to store files or data objects. Physically, the resource may span across multiple servers.Data robustness is a major requirement for storage systems. There have been many proposals of storing data over storage servers. One way to provide data robustness is to replicate a message such that each storage server stores a copy of the message.A decentralized fuzzy code is suitable for use in a distributed storage system.We construct a secure cloud storage system that supports the function of secure data forwarding by using an AES and Proxy re encryption. In this model initial phase owner will upload the data with AES Encryption. Next phase, inside of cloud again the data has divided into small pieces, for this process we will apply a dividing key. Data will place in different storage lactations. The information of data storage will monitor by a unique data distributors. If the valid user accessing the data cloud will retrieve the data as reversible manner.*

Keywords- Cloud computing ,secret key, cloud service provider,Fuzzy search

I. INTRODUCTION

Cloud computing is an information technology (IT) paradigm that enables ubiquitous access to shared pools of configurable system resources and higher-level services that can be rapidly provisioned with minimal management effort, often over the Internet. Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility.Third-party clouds enable organizations to focus on their core businesses instead of expending resources on computer infrastructure and maintenance. Advocates note that cloud computing allows companies to avoid or minimize up-front IT infrastructure costs. Proponents also claim that cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and that it enables IT teams to more rapidly adjust resources to meet fluctuating and unpredictable business demand. Cloud providers typically use a "pay-as-you-go" model, which can lead toun expected operating expenses if administrators are not familiarized with cloud-pricing models.

II. TYPES OF CLOUD COMPUTING

Public Cloud: A public cloud is basically the internet. Service providers use the internet to make resources, such as applications (also known as Software-as-a-service) and storage, available to the general public, or on a ‘public cloud. Examples of public clouds include Amazon Elastic Compute Cloud (EC2), IBM’s Blue Cloud, Sun Cloud, Google AppEngine and Windows Azure Services Platform. For users, these types of clouds will provide the best economies of scale, are inexpensive to set-up because hardware, application and bandwidth costs are covered by the provider. It’s a pay-per-usage model and the only costs incurred are based on the capacity that is used.Thereare some limitations, however; the public cloud may not be the right fit for every organization. The model can limit configuration, security, and SLA specificity, making it less-than-ideal for services using sensitive data that is subject to compliancy regulations.

Private Cloud: Private clouds are data center architectures owned by a single company that provides flexibility, scalability, provisioning, automation and monitoring. The goal of a private cloud is not sell “as-a-service” offerings to external customers but instead to gain the benefits of cloud architecture without giving up the control of maintaining your own data center.Private clouds can be expensive with typically modest economies of scale. This is usually not an option for the average Small-to-Medium sized business and is most typically put to use by large enterprises. Private clouds are driven by concerns around security and compliance, and keeping assets within the firewall.

Hybrid Cloud: By using a Hybrid approach, companies can maintain control of an internally managed private cloud while relying on the public cloud as needed. For instance during peak periods individual applications, or portions of applications can be migrated to the Public Cloud. This will also be beneficial during predictable outages: hurricane warnings, scheduled maintenance windows, rolling brown/blackouts.The ability to maintain an off-premise disaster recovery site for most organizations is impossible due to cost. While there are lower cost solutions and alternatives the lower down the spectrum an organization gets, the capability to recover data quickly reduces. Cloud based Disaster Recovery (DR)/Business Continuity (BC) services allow organizations to contract failover out to a Managed Services Provider that maintains multi-tenant infrastructure for DR/BC, and specializes in getting business back online quickly.

Advantages of Cloud Computing:

Drive down costs: Avoid large capital expenditure on hardware and upgrades. Cloud can also improve cost efficiency by more closely matching your cost pattern to your revenue/demand pattern, moving your business from a capital-intensive cost model to an Opex model. Cope with demand: You know what infrastructure you need today, but what about your future requirements? As your business grows, a cloud environment should grow with you. And when demand is unpredictable or you need to test a new application, you have the ability spin capacity up or down, while paying only for what you use.Run your business: don’t worry about your IT: Monitoring your infrastructure 24/7 is time consuming and expensive when you have a business to run. A managed cloud solution means that your hosting provider is doing this for you. In addition to monitoring your infrastructure and keeping your data safe, they can provide creative and practical solutions to your needs, as well as expert advice to keep your IT infrastructure working efficiently as your needs evolve.Innovate and lead: Ever-changing business requirements mean that your IT infrastructure has to be flexible. With a cloud infrastructure, you can rapidly deploy new projects and take them live quickly, keeping you at the vanguard of innovation in your sector.Improved security and compliance: You have to protect your business against loss of revenue and brand damage.Your cloud hosting provider will build in resiliency and agility at an infrastructure-level to limit the risk of a security breach, and will work with you to help address compliance and regulatory requirements. Reduce your carbon footprint: Hosting in a data center rather than onsite allows you to take advantage of the latest energy-efficient technology. Additionally, as cloud service providers host multiple customers on shared infrastructure, they can drive higher and more efficient utilization of energy resources. Future-proof your business: There is unprecedented demand for access to data anywhere, any time and on any device. Don’t let your business fall behind. By embracing the cloud, you can handle emerging mobile, BYOD and wearable technology trends.

CLOUD COMPUTING SECURITY: Cloud computing security or, more simply, cloud security refers to a broad set of policies, technologies, and controls deployed to protect data, applications, and the associated infrastructure of cloud computing. It is a sub-domain of computer security, network security, and, more broadly, information security.

SECURITY ISSUES IN CLOUD: Cloud computing and storage provides users with capabilities to store and process their data in third-party data centers. Organizations use the cloud in a variety of different service models (with acronyms such as SaaSPaaS and IaaS) and deployment models (private, public, hybrid, and community). Security concerns associated with cloud computing fall into two broad categories: security issues faced by cloud providers (organization providing software-, platform-,or infrastructure-as-a-service via the cloud) and security issues faced by their customers (companies or organizations who host applications or store data on the cloud), The responsibility is shared, however. The provider must ensure that their infrastructure is secure and that their clients’ data and applications are protected, while the user must take measures to fortify their application and use strong passwords and authentication measures.When an organization elects to store data or host applications on the public cloud, it loses its ability to have physical access to the servers hosting its information. As a result, potentially sensitive data is at risk from insider attacks. According to a recent Cloud Security Alliance report, insider attacks are the sixth biggest threat in cloud computing. As a result, there is a chance that one user's private data can be viewed by other users (possibly even competitors). To handle such sensitive situations, cloud service providers should ensure proper data isolation and logical storage segregation. The extensive use of virtualization in implementing cloud infrastructure brings unique security concerns for customers or tenants of a public cloud service. Virtualization alters the relationship between the OS and underlying hardware – be it computing, storage or even networking. This introduces an additional layer – virtualization – that itself must be properly configured, managed and secured. Specific concerns include the potential to compromise the virtualization software, or "hypervisor". While these concerns are largely theoretical, they do exist. For example, a breach in the administrator workstation with the management software of the virtualization software can cause the whole datacentre to go down or be reconfigured to an attacker's liking.

III. EXISTING SYSTEM

we use a straightforward integration method. In straightforward integration method Storing data in a third party's cloud system causes serious concern on data confidentiality. In order to provide strong confidentiality for messages in storage servers, a user can encrypt messages by a cryptographic method before applying an fuzzy code method to encode and store messages. When he wants to use a message, he needs to retrieve theCodeword symbols from storage servers, decode them, and then decrypt them by using cryptographic keys. General encryption schemes protect data confidentiality, but also limit the functionality of the storage system because a few operations are supported over encrypted data. A decentralized architecture for storage systems offers good scalability, because a storage server can join or leave without control of a central authority.The user can perform more computation and communication traffic between the user and storage servers is high.The user has to manage his cryptographic keys otherwise the security has to be broken.

The data storing and retrieving, it is hard for storage servers to directly support other functions.

Searchable encryption: Traditional searchable encryption has been widely studied in the context of cryptography. Among those works, most are focused on efficiency improvements and security definition formalizations. The first construction of searchable encryption was proposed by Song et al. in which each word in the document is encrypted independently under a special two-layered encryption construction. Gol proposed to use Bloom filters to construct the indexes for the data files. To achieve more efficient search, Chang et al and Curtmola et al both proposed similar “index” approaches, where a single encrypted hash table index is built for the entire file collection. In the index table, each entry consists of the trapdoor of a keyword and an encrypted set of file identifiers whose corresponding data files contain the keyword. As a complementary approach, Boneh et al. presented a public-key based searchable encryption scheme, with an analogous scenario to that of Note that all these existing schemes support only exact keyword search, and thus are not suitable for Cloud Computing. Others. Private matching as another related notion, has been studied mostly in the context of secure multiparty computation to let different parties compute some function of their own data collaboratively without revealing their data to the others. These functions could be intersection or approximate private matching of two sets, etc. The private information retrieval is an often-used technique to retrieve the matching items secretly, which has been widely applied in information retrieval from database and usually incurs unexpectedly computation complexity.

ENHANCED FUZZY SEARCH SCHEME

In our proposed system, we address the problem of forwarding data to another user by storage servers directly under the command of the data owner. We consider the system model that consists of distributed storage servers and key servers. Since storing cryptographic keys in a single device is risky, a user distributes his cryptographic key to key servers that shall perform cryptographic functions on behalf of the user. These key servers are highly protected by security mechanisms. Here Storage system has allocates by different data container. Once owner uploads the data with AES encryption mechanism, system again takes the data and makes Secure Data segregation process. All the data pieces will be save in different location in cloud storage. Here public distributor monitors all the data and corresponding positions where it is saved. When a proper client asking the data, cloud system will provide the data in reversible manner. So our system will prevent our data from both Inside and Outside attackers.

Tight integration of encoding, encryption, and forwarding makes the storage system efficiently meet the requirements of data robustness, data confidentiality, and data forwarding.The storage servers independently perform encoding and re-encryption process and the key servers independently perform partial decryption process. More flexible adjustment between the number of storage servers and robustness.

ADVANCED ENCRYPTION STANDARD

The Advanced Encryption Standard (AES) is defined in each of FIPS PUB 197: Advanced Encryption Standard (AES) ISO/IEC 18033-3: Information technology – Security techniques – Encryption algorithms

DESCRIPTION OF CIPHER: AES is based on a design principle known as a substitution-permutation network, a combination of both substitution and permutation, and is fast in both software and hardware. Unlike its predecessor DES, AES does not use a Feistel network. AES is a variant of Rijndael which has a fixed block size of 128 bits, and a key size of 128, 192, or 256 bits. By contrast, the Rijndael specification per se is specified with block and key sizes that may be any multiple of 32 bits, with a minimum of 128 and a maximum of 256 bits.

AES operates on a 4 × 4 column-major order matrix of bytes, termed the state, although some versions of Rijndael have a larger block size and have additional columns in the state. Most AES calculations are done in a particular finite field.
For instance, if there are 16 bytes, b0,b1,...,b15

b_{0},b_{1},...,b_{15}} , these bytes are represented as this matrix:
[b0b4b8b12
b1b5b9b13
b2b6b10b14
b3b7b11b15]The key size used for an AES cipher specifies the number of repetitions of transformation rounds that convert the input, called the plaintext, into the final output, called the ciphertext.

The number of cycles of repetition are as follows:10 cycles of repetition for 128-bit keys-- 12 cycles of repetition for 192-bit keys-- 14 cycles of repetition for 256-bit keys.

Each round consists of several processing steps, each containing four similar but different stages, including one that depends on the encryption key itself. A set of reverse rounds are applied to transform ciphertext back into the original plaintext using the same encryption key.

High-level description of the algorithm

KeyExpansions—round keys are derived from the cipher key using Rijndael's key schedule. AES requires a separate 128-bit round key block for each round plus one more.

InitialRound: AddRoundKey—each byte of the state is combined with a block of the round key using bitwise xor.

Rounds: SubBytes—a non-linear substitution step where each byte is replaced with another according to a lookup table.

ShiftRows—a transposition step where the last three rows of the state are shifted cyclically a certain number of steps.

MixColumns—a mixing operation which operates on the columns of the state, combining the four bytes in each column.

Final Round (no MixColumns)

1. SubBytes
2. ShiftRows
3. AddRoundKey.

SECURITY

Until May 2009, the only successful published attacks against the full AES were side-channel attacks on some specific implementations. The National Security Agency (NSA) reviewed all the AES finalists, including Rijndael, and stated that all of them were secure enough for U.S. Government non-classified data. In June 2003, the U.S. Government announced that AES could be used to protect classified information:

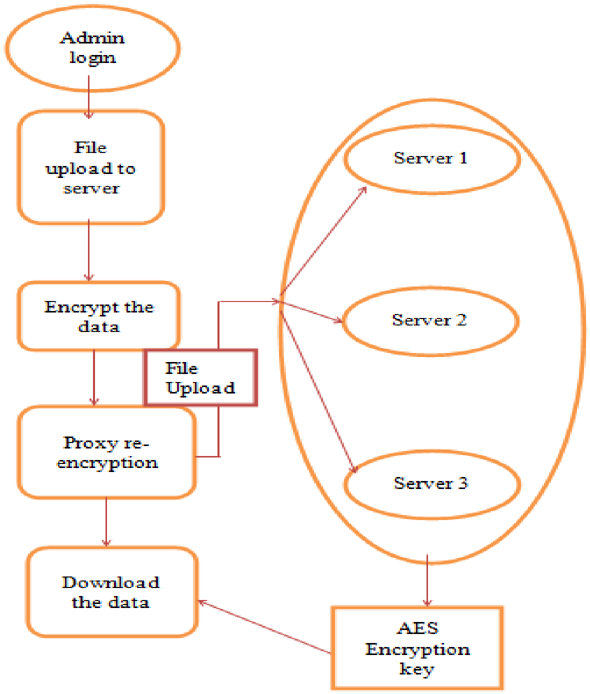
The design and strength of all key lengths of the AES algorithm (i.e., 128, 192 and 256) are sufficient to protect classified information up to the SECRET level. TOP SECRET information will require use of either the 192 or 256 key lengths. The implementation of AES in products intended to protect national security systems and/or information must be reviewed and certified by NSA prior to their acquisition and use. AES has 10 rounds for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for 256-bit keys.By 2006, the best known attacks were on 7 rounds for 128-bit keys, 8 rounds for 192-bit keys, and 9 rounds for 256-bit keys.

The features of AES are as follows

Symmetric key symmetric block cipher-- 128-bit data, 128/192/256-bit keys-- Stronger and faster than Triple-DES

Provide full specification and design details-- Software implementable in C and Java

IV. SYSTEM ARCHITECTURE



Registration:

For the registration of user with identity ID the group manager randomly selects a number. Then the group manager adds into the group user list which will be used in the traceability hase. After the registration, user obtains a private key which will be used for group signature generation and file decryption.

Sharing Data:

The canonical application is data sharing. The public auditing property is especially useful when we expect the delegation to be efficient and flexible. The schemes enable a content provider to share her data in a confidential and selective way, with a fixed and small cipher text expansion, by distributing to each authorized user a single and small aggregate key.

Secure Cloud Storage:

Data robustness is a major requirement for storage systems. There have been many proposals of storing data over storage servers. One way to provide data robustness is to replicate a message such that each storage server stores a copy of the message. A decentralized fuzzy code is suitable for use in a distributed storage system

Proxy re-encryption:

Proxy re-encryption schemes are crypto systems which allow third parties (proxies) to alter a cipher text which has been encrypted for one user, so that it may be decrypted by another user. By using proxy re-encryption technique the encrypted data (cipher text) in the cloud is again altered by the user. It provides highly secured information stored in the cloud. Every user will have a public key and private key. Public key of every user is known to everyone but private key is known only the particular user.

Data retrieval:

Reports and data are the two primary forms of the retrieved data from servers. There are some overlaps between them, but queries generally select a relatively small portion of the server, while reports show larger amounts of data. Queries also present the data in a standard format and usually display it on the monitor; whereas reports allow formatting of the output however you like and is normally retrieved.

Conclusion and Future works:

we tackled the challenging fuzzy search problem over the encrypted data. We proposed and integrated several innovative designs to solve the fuzzy search problems simultaneously with high efficiency. Our approach of giving privacy to the files is through advanced encryption technique . We proposed a basic scheme as well as an improved scheme in order to meet different security requirements. Thorough theoretical security analysis and experimental evaluation using real-world dataset were carried out to demonstrate the suitability of our proposed scheme for the practice usage.

The registered data users will provided with the security key to download the document or a datasets, By giving extra security to the key the data or a document can be securely protected.

REFERENCES

[1] E. Anderson. Simple table-based modeling of storage devices. Technical Report HPL-2001-04, HP Laboratories, 2001.

[2] J. Bruno, J. Brustoloni, E. Gabber, B. Ozden, and A. Silberschatz. Disk scheduling with quality of service guarantees. In Proc. of the IEEE International Conference on Multimedia Computing and Systems, 1999.

[3] J. Bent, G. Gibson, G. Grider, B. McClelland, P. Nowoczynski, J. Nunez, M. Polte, and M. Wingate. PLFS: a checkpoint _le system for parallel applications. In Proc. of the Supercomputing Conference, 2009.

[4] QoS Support for End Users of I/O-intensive ApplicationsUsing Shared Storage Systems. Author: Xuechen Zhang ECE Department Wayne State Universities Trans. Kei Davison Alamos National Laboratory Parallel and Distributed Systems, vol. 21, no. 11, pp. 1586-1594, Nov. 2010

[5] Repair Locality from a Combinatorial Perspective. Author:Anyu Wang and ZhifangZhangKey Laboratory of Mathematics Mechanization, IEEE Dec.2014.

[6] D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” S&P 2000, vol. 8, pp. 44–55, 2000.

[7] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, T. Hou, and H. Li, “Privacypreservingmulti-keyword text search in the cloud supporting similaritybasedranking,” in ASIACCS 2013, May 2013.

[8] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, “Public keyencryption with keyword search,” EUROCRYPTO 2004, pp. 506–522,
2004.

[9] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-preserving multi-keyword ranked search over encrypted cloud data,” INFOCOM

2011, pp. 829–837, 2011.

[10] S. Barua, R. Thulasiram, and P. Thulasiraman. Highperformance computing for a _nancial application using Fast Fourier Transform. In Proc. of EURO-PAR, 2005.

[11] D. Chambliss, G. Alvarez, P. Pandey, D. Jadav, J. Xu, R. Menon, and T. Lee. Performance virtualization for large-scale storage systems. In Proc. of the Symposium on Reliable Distributed systems, 2003.