International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072

Selective Encryption and Component-Oriented De-duplication for Cloud Data Computing

Aishwarya.N[1], Gouri Mangalgi[2], Nilafar[3], Anitha.R[4]

[1,2,3] Student, Dept. of Computer Science and Engineering, National Institute of Engineering, Mysuru, India.
[4]Associate Professor, M.Tech & B.E, Dept. of Computer Science and Engineering, National Institute of Engineering, Mysuru, India.

Abstract —*as smart devices gain their popularity and usage applications become versatile, the users are also hoping to perform resource intensive tasks at anywhere and anytime as conveniently as using their static computers. To overcome the smart device's intrinsic resource limitations in processing, storage, and power, emerging collaborative cloud technologies such as Mobile Cloud Computing (MCC), Mobile-Edge Computing (MEC), and Fog Computing (FC) augment the smart device's capabilities by leveraging distributed and remote cloud resources. However, in collaborative computing environments, the demand for big data processing and exchanges among smart devices is considered as a significant challenge. An effective technique to reduce data at a source device is essential to save network bandwidth and storage spaces. It, in turn, improves the data processing overhead as well as reduces the security vulnerability caused by data movement among the smart devices. In this paper, we design and develop a novel Selective Encryption and Component-Oriented De-duplication (SEACOD) application that achieves both fast and effective data encryption and reduction for Cloud services. Specifically, SEACOD efficiently de-duplicates redundant objects in files, emails, as well as images exploiting object-level components based on their structures. It also effectively reduces the overall encryption overhead on the smart devices by adaptively applying compression and encryption methods according to the decomposed data types. Our evaluation using real datasets of structured files shows that the proposed scheme accomplishes as good of storage savings as a variable-block de-duplication, while being as fast as a file-level or a large fixed-size block-level de-duplication.*

Key Words: **Amazon SimpleDB, Amazon S3, SHA-256, Rijndael algorithm, Open xml, Visual Studio, Amazon Wb Services.**

1. INTRODUCTION

Cloud computing provides seemingly unlimited storage resources to users. But one of the critical challenges of cloud storage services is the management of the ever-increasing volume of data. To make data management scalable in cloud computing, de-duplication has been a well-known technique and has attracted more and more attention recently. Also, recent years have witnessed the trend of leveraging cloud-based services for large scale content storage, processing, and distribution. Security and privacy are among top concerns for the public cloud environments. Towards these security and storage challenges, we propose and implement a new client-side de-duplication scheme for securely storing and sharing outsourced data via the public cloud.

Data de-duplication is a specialized data compression technique for eliminating du-plicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, de-duplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Although data de-duplication brings a lot of benefits, security and privacy concerns arise as users sensitive data are susceptible to both inside and outside attacks. Traditional encryption, while providing data confidentiality, is incompatible with data de-duplication. Specifically, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different cipher texts, making de-duplication impossible. Convergent encryption has been proposed to enforce data confidentiality while making de-duplication feasible.

2. PROPOSED SYSTEM

We use an algorithm called SEACOD (Selective encryption and component oriented de-duplication). Our secure client-side data de-duplication scheme is based on an original use of the convergent encryption. That is, on one hand, when a data owner wants to store a data file in remote storage servers, he has first to generate the data identifiers from the extracted document components. These component data identifiers are derived by applying a one way hash function to each data components. Then, for subsequent data outsourcing, at first the client is not required to send the data. However he sends only the data identifiers for the selected document. Component level de-duplication check happens at cloud service followed by selective encryption at client.

The SEACOD framework consists of a light-weight smart phone application (SEACOD client) and a server middleware (SEACOD server) on cloud platform. The SEACOD client consists of four modules which includes

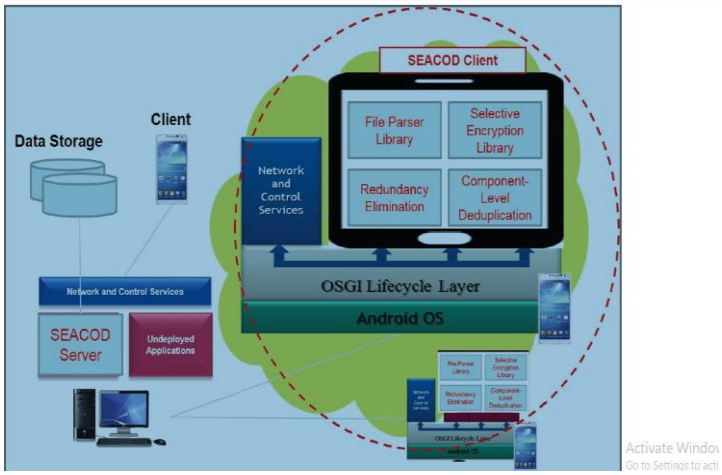

Fig-1: Proposed system model

File parser library module: This includes various basic types of file parsers (i.e., docx, pptx, and pdfs) .They decomposes an original file into potentially many smaller sized objects according to the file structure policy. The file parser also can combine several small objects into a compound object based on the file parsers policy

Selective encryption library module: This maintains the overall encryption overhead for different settings including different sizes of data blocks and different data types (text, images, and audio files) for each encryption algorithm.

Component-level de-duplication manager: This receives object indexes and decomposed objects of a file from the file parser and checks the uniqueness of an index by checking the object index table.

Redundancy elimination protocol module: This uploads the component indexes created by the SEACOD clients De-duplication parser to the SEACOD server via a batch packet so that the SEACOD server can correlate the indexes to eliminate the redundant data exchange.

Efficient non server side data reduction techniques are essential to save data on the path from a user to cloud servers or storage spaces. It in turn, expedites the data processing and transmission speed as well as reduces data vulnerability in the platform. Although traditional server-side data De-duplication techniques tend to achieve a high data reduction rate, as they require high processing overhead due to data chunking, index processing, and data fragmentation, they cannot be directly used in capacity limited mobile devices.

3. ARCHITECTURE

Architectural design depicts the functionalities of the modules of the system and the interaction between the individual modules of the system. It shows the flow of the information or execution process in the system. A major task of the design is to spell out in detail, the input, output and functionality of each module of the system. This forms the design document. The design document is the developers blue print. It provides precise direction to software programmers about how basic control and data structures will be organized. The designed document is usually written before the programming starts. It describes how the software will be structured, and what functionality will be included. This document forms the basis for all future design and coding.

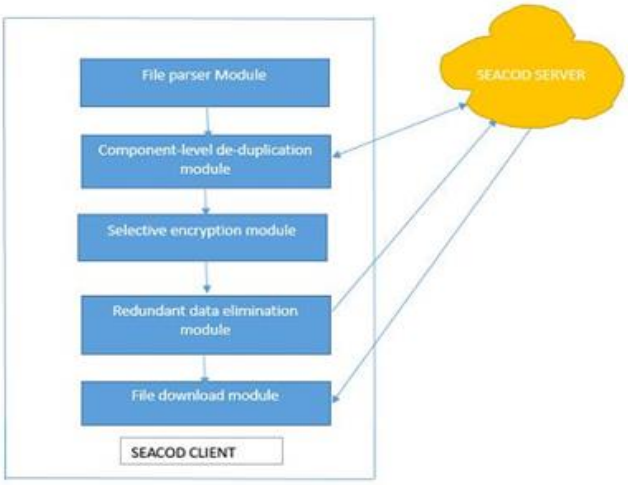

Fig-2: The design of the system

Module 1: File Parser Module, MS Word file (*.docx) follows the Office Open XML format (called Open XML). Texts of a word file are contained in a document.xml object, and image objects are under a media directory, while other directories contain meta-data objects. Word file is parsed and decomposed into smaller sized objects according to the provided file structure policies. The parser also can combine several small objects into a compound object based on the le

Parsers policy. The small sized objects are called components in our case.

Fig-3: Flow Chart for File Parser Module

Module2: Component Level De-duplication Manager Module First, it indexes the components by their Hash Value (SHA256 Algorithm).Without sending the data over the network, it uploads indexes for the cloud server to look up the same data (i.e. checks the uniqueness of an index by checking the Object Index Table).Cloud server in return sends acknowledge on only the non-redundant component indexes. By sending the hash values instead of original full component, the network over-load will be reduced.

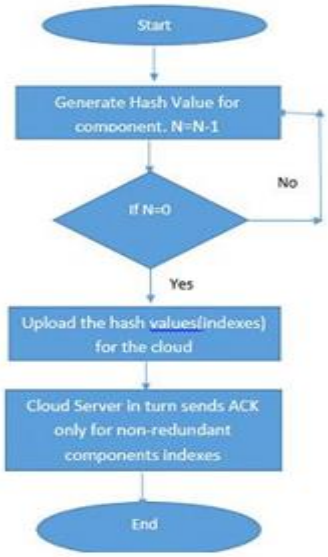

Fig-4: Flow Chart for Component-Level De-duplication Manager Module

Module 3: Selective Encryption Module: Choose Components for Encryption. Generates key for data encryption using the computed data index (Hash Value) Performs data encryption (Symmetric Encryption Algorithms AES). For Encryption, we are using rijndael encryption algorithm.

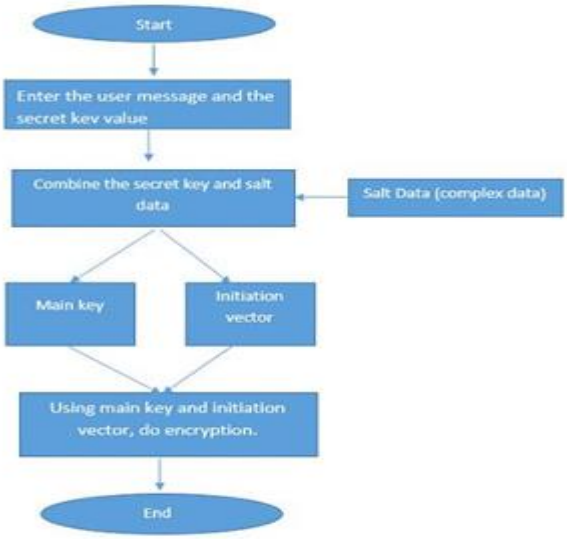

Fig-5: Flow Chart for Selective Encryption Module

Module 4: Redundant Data Elimination Module: Unique (non-redundant) encrypted data components are uploaded to cloud server for storage. Cloud stores all the received data components and maintains an Object Index Table. Amazon Simple DB will be used to hold file informations, object indexes, file owner, date and time etc. Amazon S3 will be used to hold physical encrypted file objects (text/images).

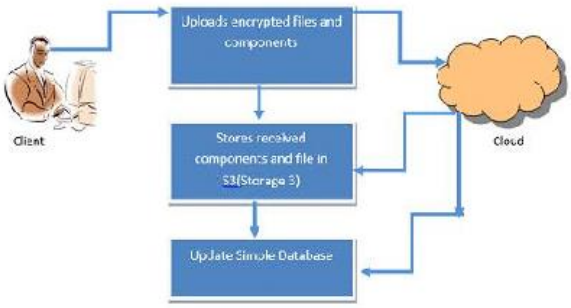

Fig-6: Flow Chart for Redundant Data Elimination Module

Module 5: File Download Module: Unique (non-redundant) encrypted data components are uploaded to cloud server for storage to client machine. Encrypted data components are decrypted using respective cryptography keys. Word File is re composed with all decrypted components.

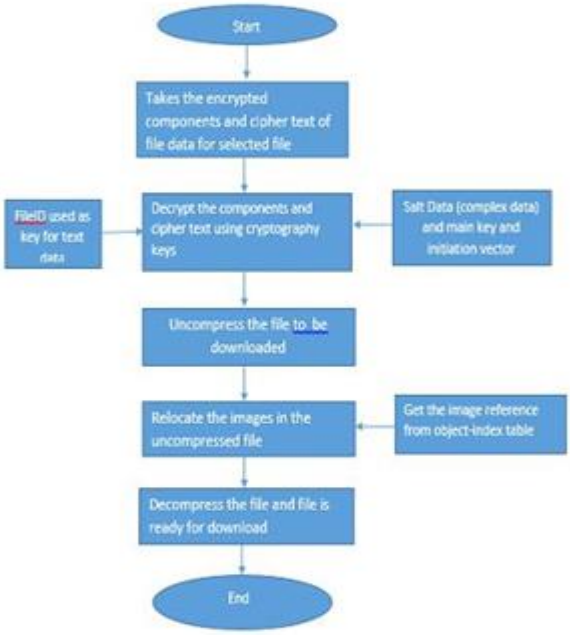

Fig-7: Flow Chart for File Download Module

4. EXPERIMENTAL OBSERVATIONS AND RESULTS

The modules implementation of the Selective encryption and component-oriented de-duplication (SEACOD) are as follows. The MS Word file follows the Office Open XML format(called Open XML).Texts of a word file are contained

in a document .xml object , and image objects are under a media directory , while other directories contain meta-data objects. In the File parser module Word file is parsed and decomposed into smaller sized objects according to the provided file structure policies .In the Component Level De-duplication Manager Module ,components are indexed with their unique hash values and are sent through the network. The Cloud server in return sends the acknowledgement of those indexes which are not present in the cloud. In the Selective Encryption Module, Components are selectively encrypted. In the Redundant Data Elimination Module, Unique encrypted components are uploaded to the cloud. At last in File Download Module, the encrypted components are decrypted, formatted to original document and downloaded.

5. CONCLUSION

SEACOD application has been developed and is successfully able to make optimize use of remote storage (cloud). SEACOD application that achieves effective data reduction, better encryption, and data-oriented collaboration control for resource intensive mission-oriented cloud computing services. We designed efficient methods to reduce the overall encryption overhead on the system by selectively applying encryption methods according to the decomposed data types. We developed an intelligent mechanism to avoid the unnecessary data exchanges by exploring the collaborating members data processing and transfer capability and existing data components. The results as observed from the testing and result analysis are found to be satisfactory meant to demonstrate the possibility and methodology for the implementation. In the end our study creates a solution on the de-duplication problem based on the components (considering images in the docx file).However, it could be further extended out in a few ways. For instance, de-duplication based on audio and videos files. This application can be created for a domain specific purpose. For instance, Electronic Health Record (EHR) files such as a DICOM format for the Hospitals, where Physicians can access, update and store records(EHR files) in the remote server in efficient and optimize way.

REFERENCES

[1]Song, Sejun, Baek-Young Choi, and Daehee Kim. "Selective encryption and component-oriented de-duplication for cloud data computing." Computing, Networking and Communications (ICNC), 2016 International Conference on. IEEE, 2016.

[2] Li, Jin, et al. "A hybrid cloud approach for secure authorized De-duplication." IEEE Transactions on Parallel and Distributed Systems 26.5 (2015): 1206-1216.

[3] Kawtikwar, Namrata P., and M. R. Joshi. "Data Deduplication in Cloud Environment using File-Level and Block-Level Techniques." Imperial Journal of Interdisciplinary Research 3.5 (2017)

[4] Kaaniche, Nesrine, and Maryline Laurent. "A secure client side deduplication scheme in cloud storage environments." New Technologies, Mobility and Security (NTMS), 2014 6th International Conference on. IEEE, 2014.

[5] Alatorre, Gabriel, et al. "On selective compression of primary data." Network and Service Management (CNSM), 2015 11th International Conference on. IEEE, 2015.