

Robust Data De-duplication Security in Fog Computing

Deepika & Mrs.Kavita Rathi

M.Tech scholar DCRUST Murthal, Sonipat,

deepikaporia@gmail.com

Asstt. Prof. (CSE) DCRUST Murthal, Sonipat,

Kavita1217@gmail.com

Abstract-The Fog computing over the cloud data storage provides numerous benefits to their clients such as cost savings, accessibility, scalability etc., users around the world tend to shift their invaluable data to cloud storage. As the data generation rates are increasing, it is a tedious task for cloud storage providers to provide efficient storage. Cloud storage providers uses different techniques to improve storage efficiency and one of leading technique employed by them is deduplication, which claims to be saving 90 to 95% of storage. Data Deduplication technique evolved as an simple storage optimization technique in secondary then widely adapted in primary storage as well as larger storage areas like cloud storage area. Now, data deduplication is widely used by various cloud storage providers. Data once deployed to cloud servers, its beyond the security premises of the data owner, thus most of them prefer to outsource their in an encrypted format. In this paper we provide review of various data deduplication techniques used for fog computing. We also propose hybrid approach for encryption for providing security on fog computing.

Keywords: Cloud Computing, Fog computing, edge computing, data deduplication

I. INTRODUCTION

With the origination of Cloud computing [1], computation technology has entered to a new era. Many computation service providers including Google, Amazon, IBM, Microsoft, etc. are currently nurturing this popular computing paradigm as a utility. They have enabled “cloud based services such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS)”, etc. to handle numerous enterprise and educational related issues simultaneously. However, most of the Cloud datacentres are geographically centralized and situated far from the proximity of the end devices/users. As a consequence, real-time and latency-sensitive

computation service requests to be responded by the distant Cloud datacentres often endure large round-trip delay, network congestion, service quality degradation, etc.

To resolve these issues besides centralized Cloud computing, a new concept named “Edge computing” has recently been proposed [2]. The fundamental idea of Edge computing is to bring the computation facilities closer to the source of the data. Taking the notion of Edge and Cloud computing into account, several computing paradigms have already been introduced in computation technology. One of them is Fog Computing.

In Fog computing [3], services can be hosted at end devices such as set-top-boxes or access points. The infrastructure of this new distributed computing allows applications to run as close as possible to sensed actionable and massive data, coming out of people, processes and thing. Such Fog computing concept, actually a Cloud computing close to the ‘ground’, creates automated response that drives the value.

We adopt a simple three level hierarchy as in Figure 1.

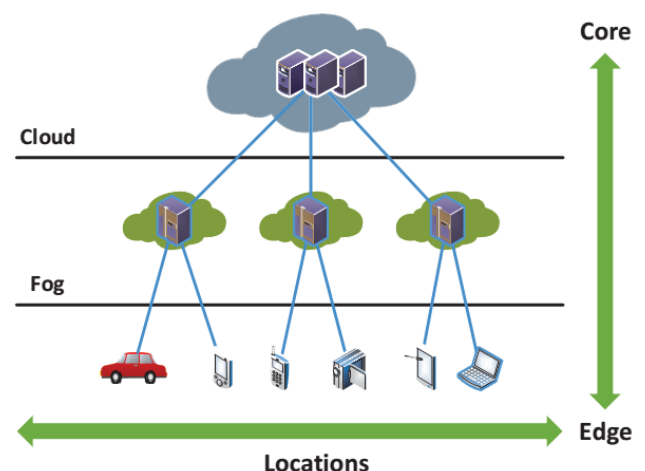


Fig 1: Fog between edge and cloud.

In this framework, each smart thing is attached to one of Fog devices. Fog devices could be interconnected and each of them is linked to the Cloud.

In some way fog computing behaves similar to cloud computing. Both computing technologies provide application, storage, data and computing services to their registered clients. But fog computing provides services close to its end users as compared to cloud computing that provides services remotely.

Both Cloud and Fog provide data, computation, storage and application services to end-users. However, Fog can be distinguished from Cloud by its proximity to end-users, the dense geographical distribution and its support for mobility [4].

In this paper we provide review of various data deduplication techniques used for fog computing. We also propose hybrid approach for encryption for providing security on fog computing.

II. DATA DEDUPLICATION

With numerous benefits of cloud storage such as cost savings, accessibility, scalability etc., users around the world tend to shift their invaluable data to cloud storage. As the data generation rates are increasing, it is a tedious task for cloud storage providers to provide efficient storage. Cloud storage providers uses different techniques to improve storage efficiency and one of leading technique employed by them is deduplication, which claims to be saving 90 to 95% of storage. Data Deduplication technique evolved as an simple storage optimization technique in secondary then widely adapted in primary storage as well as larger storage areas like cloud storage area.

Deduplication [5][6] can be categorized based on the location it takes place: server-side and client-side. In the former, remote server allows data owners to upload their contents and stores them in its storage. Then, duplicate identification and elimination occur in the background or periodically on remote storage. On the other hand, the remote server selectively allows data owners to upload their contents in the latter. Based on a small representative value (e.g., hash digest), data owners can upload their contents only when the server cannot discover the same value as a searching key on its storage. Otherwise, the upload-requesting user is registered as a valid data owner (after a certain kind of validation process if necessary). As regards fog computing environment, the latter is more preferable because it additionally reduces communication

overhead by avoiding redundant uploads for the same content and costs for communications are often more expensive than that for computations, especially for resource-constrained equipment such as IoT devices.

Deduplication [7] can also be assorted according to chunking level: block-level and file-level. Although the former could reduce more redundant blocks (by removing identical blocks composing different files) than the latter, it might increase the whole size of metadata in proportion to the number of blocks. This efficiency 150 degradation becomes even worse when the outsourced data is encrypted as long as data owner has to maintain a series of encryption keys for encrypted blocks of a file. Based on coverage of deduplication, it is sometimes classified as single-user and cross-user ones [8][9].

III. PROPOSED METHODOLOGY

The main objectives of research work are to Study Fog Computing and its data depuplication issues in detail. A new hybrid encryption based scheme for data deduplication in fog computing is to be propose. Finally we evaluate the performance of proposed scheme using various parameters. The proposed methodology is given in fig 2

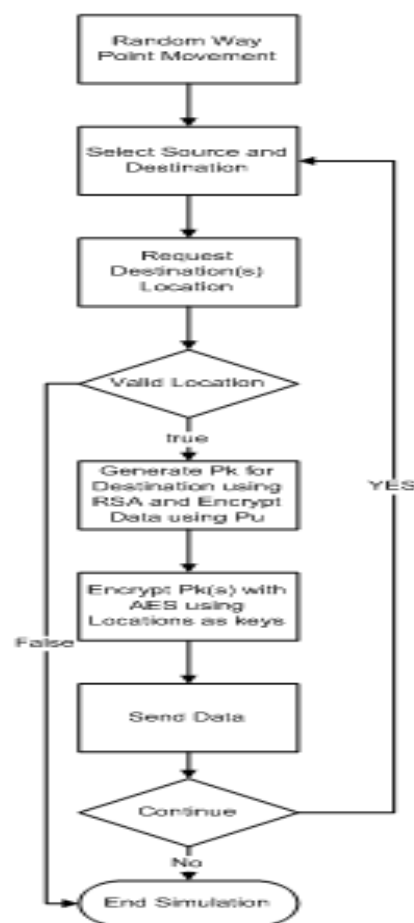


Figure 2:Workflow of Deduplication process

Following steps will explain the procedure of proposed work.

Step 1: Receiver R Requests data from the node S, only receiver knows the exact location of itself.

Step 2: On receiving request from the R node the sender initiate public key Exchange using RSA algorithm and stores keys $P_{R,S}$ as private key of the node R and $P_{S,R}$ acts as the public key for encryption of data.

Step 3: Node S then Request the Location $L_{R,S}$ of the node R.

Step 4: After receiving request from the node S the R sends its location $L_{R,S}$ to the node S.

Step 5: S then Encrypts the private key of the by using $L_{R,S}$ as Private Key using AES algorithm

Step 6: For multiple recipients the same process in repeated and a matrix is formed in following manner and forwarded to recipients.

Encrypted Data (RSA)	$L_{R,S}$ private key of the receiver 1 (AES)
	$L_{R,S}$ private key of the receiver 2 (AES)

	$L_{R,S}$ private key of the receiver N (AES)

Step 7: After receiving the packet the recipient decrypts the private key using its location and then decrypts the data sent by the sender.

IV. PROPOSED TOOL

MATLAB [10][11] is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. Typical uses include:

- Mathematical computation
- Algorithm development
- Data acquisition
- Modeling, simulation, and prototyping
- Data analysis, exploration, and visualization

- Scientific and engineering graphics
- Application development, including graphical user interface building

MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This allows you to solve many technical computing problems, especially those with matrix and vector formulations, in a fraction of the time it would take to write a program in a scalar non interactive language such as C or FORTRAN. The name MATLAB stands for matrix laboratory. MATLAB was originally written to provide easy access to matrix software developed by the LINPACK and EISPACK projects. Today, MATLAB engines incorporate the LAPACK and BLAS libraries, embedding the state of the art in software for matrix computation. MATLAB has evolved over a period of years with input from many users.

In university environments, it is the standard instructional tool for introductory and advanced courses in mathematics, engineering, and science. In industry, MATLAB is the tool of choice for high-productivity research, development, and analysis. MATLAB features a family of add-on application-specific solutions called toolboxes. Very important to most users of MATLAB, toolboxes allow you to learn and apply specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include signal processing, control systems, neural networks, fuzzy logic, wavelets, simulation, and many others.

V. IMPLEMENTATION

The implementation results for various parameters are performed as explained below.

Figure 3 below the first screen of our implementation. It asks for number of simulation nodes and simulation duration. Let number no. of nodes be 10 and simulation time be 10.

```

Command Window
Starting simulation for data deduplication security in Fog Computing
Enter number of nodes for simulation 10
fx Enter simulation duration 10
  
```

Figure 3: The first screen of our implementation

It then asks for the text file need for communication between two nodes.

Figure 4 below shows the list of connecting nodes for communication in fog network.

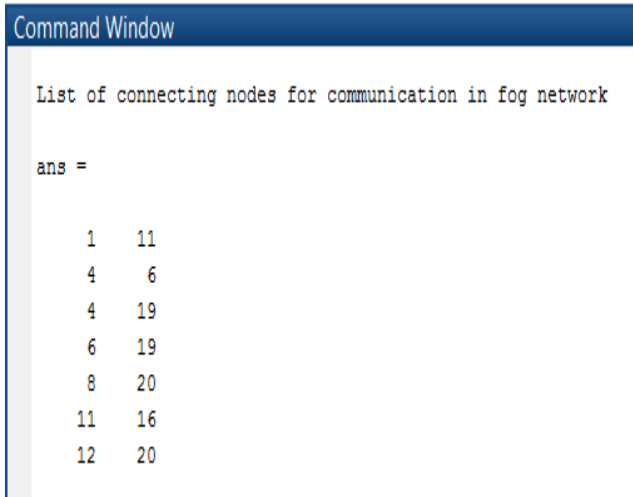


Figure 4: The list of connecting nodes using random waypoint

Figure 5 below shows the final result of fog computing networks. It shows simulation nodes in blue colour, node range in green colour and communication link between two nodes in red colour.

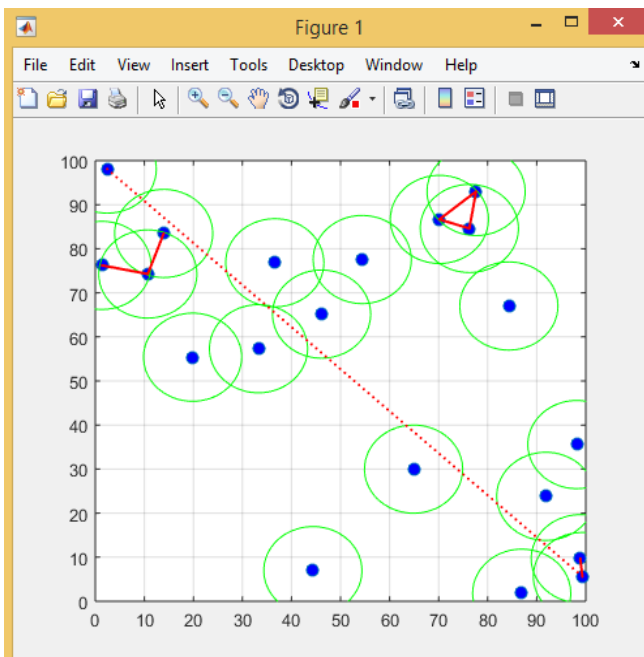


Figure 5: The final result of fog computing networks.

We have used RSA and AES encryption algorithms to encrypt the data before sending to the destination and decrypt at the destination location.

Figure 6 below shows the encryption time for fog computing network.

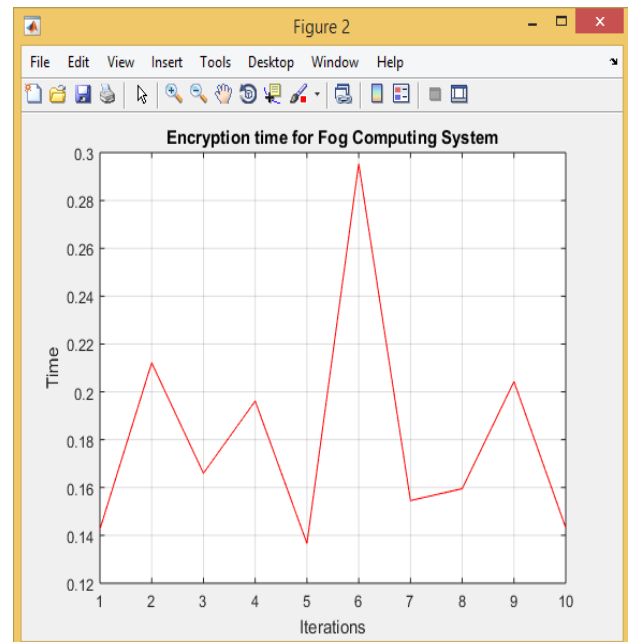


Figure 6: The encryption time for fog computing network.

Figure 7 below shows the decryption time for fog computing network.

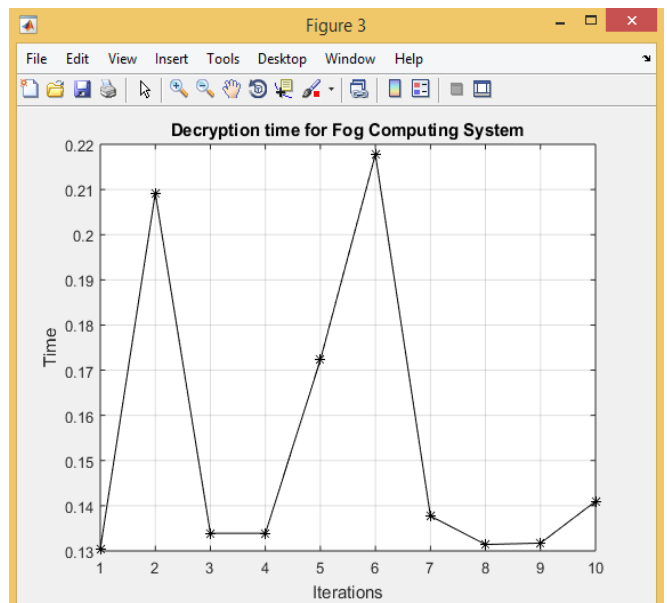


Figure 7: The encryption time for fog computing network.

As for each location a new private key is generated and attached to the packet, we require that there must not be a significant overhead as the Number of recipient grow. It can be seen from above figures that the encryption time & decryption time do not increases too much when number of nodes (iterations).

VI. CONCLUSION

In Fog computing, services can be hosted at end devices such as set-top-boxes or access points. The infrastructure of this new distributed computing allows applications to run as close as possible to sensed actionable and massive data, coming out of people, processes and thing. Both Cloud and Fog provide data, computation, storage and application services to end-users. However, Fog can be distinguished from Cloud by its proximity to end-users, the dense geographical distribution and its support for mobility. Data Deduplication technique evolved as an simple storage optimization technique in secondary then widely adapted in primary storage as well as larger storage areas like cloud storage area. Now, data deduplication is widely used by various cloud storage providers. Data once deployed to cloud servers, its beyond the security premises of the data owner, thus most of them prefer to outsource their in an encrypted format. In this paper we provided review of various data deduplication techniques used for fog computing. We also proposed hybrid approach for encryption for providing security on fog computing.

VII. REFERENCES

- [1]. P. Mell and T. Grance, "The NIST definition of cloud computing (draft), *NIST special publication*, vol. 800, no. 145, p. 7, 2011.
- [2]. Varghese, B., Wang, N., Barbhuiya, S., Kilpatrick, P., Nikolopoulos, D.S.: "Challenges and opportunities in edge computing", *Proceedings of the IEEE International Conference on Smart Cloud (2016)* 20–26.
- [3]. Dastjerdi, A., Gupta, H., Calheiros, R., Ghosh, S., Buyya, R., "Fog computing: principles, architectures, and applications", Morgan Kaufmann (2016)
- [4]. Bonomi, F., Milito, R., Zhu, J., Addepalli, S., "Fog computing and its role in the internet of things", *Proceedings of the first edition of the MCC workshop on Mobile cloud computing, ACM (2012)* 13–16
- [5]. Zhe Sun, Jun Shen, Jianming Young, "A novel approach to data deduplication over the engineering-oriented cloud systems", *Integrated Computer Aided Engineering*, 20(1), 2013, 45-57.
- [6]. Ni, Jianbing, et al. "Secure and Deduplicated Spatial Crowd sourcing: A Fog-Based Approach." *Global Communications Conference (GLOBECOM), 2016 IEEE*, 2016.
- [7]. Dutch T. Meyer and William J. Bolosky, "A Study of Practical Deduplication", 2011.
- [8]. Yitao Yang, Xiaolin Qin, Guozi Sun, Yong Xu, Zhongxue Yang and Zhiyue Zu, "Data Deduplication in Wireless Multimedia Monitoring Network", *International Journal of Distributed Sensor Networks*, 2013
- [9]. Bhushan Choudhary, Amit Dravid, "A Study on Authorized Deduplication Techniques in Cloud Computing", *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 12, April 2014* 4191.
- [10]. MATLAB Primer, MathWorks Inc, 2014.
- [11]. MATLAB Applications for the Practical Engineer by Kelly Bennett, InTech, 2014