

Cloud Computing Security Mechanism Analysis: Encryption and Compression to enhance Security and save Cost in Cloud- A study

Sajan Dung
Post Graduate Student
Department of
Computer Engineering
and Technology,
Guru Nanak Dev University,
Amritsar, Punjab, India

Sandeep Sharma
Professor
Department of
Computer Engineering and
Technology, Guru Nanak
Dev University, Amritsar,
Punjab, India

ABSTRACT

Transfer of data from source to destination is of prime concern in case of advanced computing. Virtual machines in terms of nodes exist in advance computing. Users of advance computing could be fair or malicious in nature. In case of malicious nodes some defence mechanism is needed. This defence mechanism is provided in terms of encryption. In advance computing, cost is encounter on the basis of pay per use. So, security along with space conservation is issues to tackle in case of advance computing. Both of these issues are tackled by the use of data deduplication. Propose literature is concerned with analysing techniques which are used in existing literature to provide data deduplication for security and compression. The comparative analysis is provided to give worth of the study.

Keywords

chunk based deduplication, data deduplication, data reduction, redundant data.

1. INTRODUCTION

Recently the international data corporation has been studied that the data that are stored in data centers has been increased day to day and it must be reached to 7.9 zettabyte at the end of this year. Many organisations have been striving with this growth of information and also its protection. To solve this problem data warehouse and data centre has been utilized that contain huge data in terabytes, petabytes. In these centres there is a chance of duplication of data which affect the performance, storage inconsistencies etc. For solving this problem many techniques can be utilized like data compression, data deduplication etc.

Data deduplication provides mechanism to ensure the safe passage of data from cloud resources. Cloud computing is

described in this section. Next section describes related work. Next section describes comparison table and last section gives summary in terms of conclusion.

2. CLOUD COMPUTING

Resources on physical machine re constrained hence forth may not be available as and when required. This initiate need of cloud computing. Cloud computing transformed the services in a manner which are delivered to the destination just like normal services including water supply, telephony and gas. The applications of cloud computing allow the virtual systems to provide physical services. Cloud computing comes out with risk free supra system in which any one can have control. It also allows distinct users to participate into it. The user applications frequently require resources. These resources may be computational in nature. The service required under this scheme of thing is known as Infrastructure as a Service(IaaS)[1]. Application resource may be prerequisite for some of the applications. Services in that case required is known as Software as a Service(SaaS)[2]. Construction, development and management of resources is critical for the success of services. These management of service is catered with the nourishment of Platform as a Service(PaaS)[3].

Distinct vendors claim to provide services along continents. Services include storage, computation and application hosting. In order to attain these services SLA(Service level agreement) need to be satisfied. Customer needs to pay only for those services which they currently utilized. Hence pay per use facility is provided to the user[4]. Two distinct parties exist in this case.

- a) Cloud Service Provider
- b) Cloud service Consumer

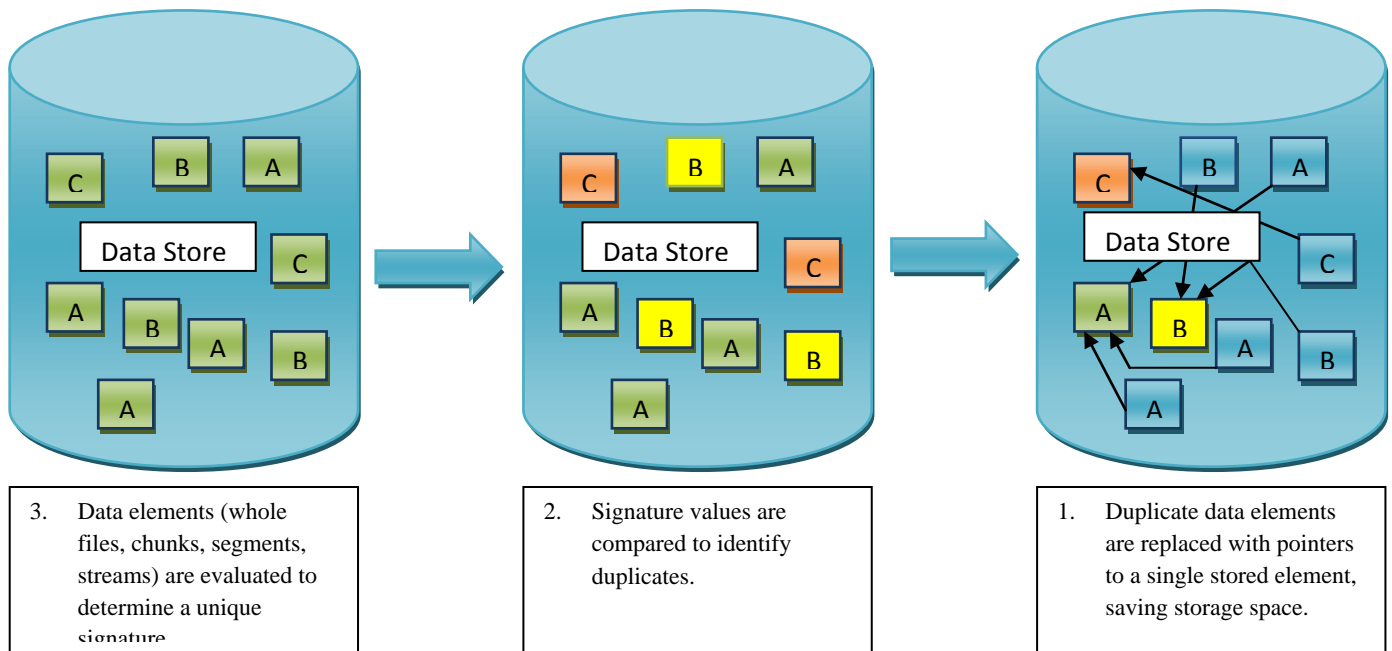


Figure -1 Basic concept of Data DeDuplication

Cloud service provider becomes the source of provider of services and clients act as consumer of services provided by the cloud. [1]

Scalability of services is big enumeration given by cloud to user. Any number of users can participate in cloud. Critical property of cloud is lack of centralized ownership. So elasticity of service is also present in this case. [5]

Cloud can be classified into further category based on scope of services. These services are as listed below

- a) Public Cloud
- b) Private Cloud
- c) Hybrid Cloud

Public cloud services are provided to general public on the pay per use basis. The access is not restricted in this case. Private cloud on the other hand provides services only to members, employees and partners. The effective features are collaborated together to form hybrid cloud. In hybrid cloud public and private clouds are utilized in a seamless manner. [3] Legion of services in addition to above listed services are discovered and presented as follows

Cloud services are commonly listed in terms of visualization. Visual computing and gaming services are unique feature listed in the armoury of cloud. Cloud computing services are enhanced version of the normal services. Physical machines may not be capable enough to handle and provide services to the users. Heavy games required hardware as well as software resources. Older platform physical machine may or may be capable enough to support such services. In order to tackle such situations cloud services becomes critical. Hence real time gaming services are provided by the use of cloud

computing. The user machines act as a source of input and output only. All the processing is done at cloud end. The game events are captured from the users. These captured events are then transferred to the cloud. These events are then processed. The obtained result is then transferred to the user

machine. Powerful computation services and platform is required in this case. Both 2D and 3D graphics can be handled using the visual environment provided with the help of cloud computing. The platform provided is cost effective solution to gaming environment. [6]

Cloud computing involves large number of computers connected over high speed network. The adoption rate is enhanced to 40% in 2013 indicating utilization of cloud resources has been growing in leaps and bounds. Infrastructure as a service(IaaS) is utilized in situation where visual services are required. Visual services are widely utilized in gaming services provided with the help of cloud computing. In order to provide such services caching, transcoding and bandwidth cost is encountered. Energy consumption has to be minimized in order to provide optimal services to the user. In fact more and more users are utilizing these services since service provided is cost effective. These optimal cost effective strategies are implemented by the use of media cloud[7]. All types of services require storage facility. The storage facility is generally provided with the help of storage services through cloud computing[8].

The execution concurrency level is rising continuously because of introduction of multi core processors. In such system buffer pool management with high scalability is required. The scalability indicates that machines are not restricted in number. Any number of resources can join the network. It is also possible to remove the resource if not required. Lock based environment is presented in order to control concurrency in such system. The problem of deadlock may incur as transaction acquiring the lock may not wish to release the resource within the buffer. The common solution is to utilize replacement policy in order to tackle the issues of deadlocks. This becomes integral part of buffer pool management. Least recently used paging schemes are used to enhance performance of buffer pool management. Traffic and contention problems are tackled through optimized buffer management with LRU methodology. [9]Services provided by

cloud can be implemented through hardware or software platforms. These platforms require separate identifications.

Software defined networking is used in order to provide loosely coupled mechanism for control plane and data plane. Network function visualization is another technique to make network functions available to open software environment.

Volume is enhancing in cloud since legions of users are preferring cloud for storing and retrieving data. Since physical machines are not capable enough to provide growing needs of the users. Information and resource require are distinct hence varsities feature requirements are always the issue. Cloud is capable enough to handle diverse information and resources so versatility is also provided using cloud computing. Computing is automated hence velocity at which processing is performed is high. Mobile application generally utilized these features. Mobile mobility is recommended in this case[9]. As more and more data is stored within cloud volume increases and Legions of techniques has been devised in order to analyze and transfer the data to the cloud. One such mechanism is known as IoT. Internet of Things provides cost effective mechanism in order to handle such volume of data.[7]

3. LITERATURE SURVEY

[11]In this paper the deduplication techniques which is based on block level elimination of redundancy has to be given. In this to save space in live file system the whole file has to be analysed and metadata has been studied. By studying this metadata new dataset are developed and evaluation of this data are done. This evaluation finds the lower storage consumption in the whole file. And provide an efficient deduplication at much lower cost.

[12] In this paper the data deduplication technology which is utilized has been used to optimize the storage system. In this data compression also reduced the number of disks energy consumption costs. In this the strategy which has been followed can reduce the amount of duplicate data.

[13] In this paper the techniques that are used to manage data over the clouds are to analysed. This describes the deduplication methods like Extreme binding , multilevel deduplication. In this the data backup services also introduced, all the techniques that has used also eliminates the duplication. In this original data has been reconstructed by using pointers of data.

[14] Here in this the use of deduplication in cloud computing for storage resources is analysed. In the cloud computing private cloud storage are to be used to store data that may have redundant data. For better resource utilization the organisations has used the deduplication techniques. It provide an optimize way to utilize the storage resources of an organisation. The organisation have utilized the chunk level and file level deduplication methods.

[15] In this paper the cluster deduplication methods has been utilised for optimal storage. Here in the method detects the redundant chunks in the system and it has two key elements. The first element contains the chunk index and second one contains the combination of communication protocol. This evaluation combines the small chunks within an environment.

[16] This paper describes about the deduplication technique that is based on the address of the data stored. It eliminates the duplicate copies of the data for efficient storage capacity and save bandwidth. The security of data during the application of deduplication methods is a major concern which has been proposed in this paper. It provides security to the data by using encryption techniques.

[17] In this paper the study of various compression techniques are to be done. The data compression is to be utilized in distributed system to decrease the duplicate data. In this paper the evaluation of different algorithms of compression is done. The evaluation is based on various factors like efficiency, susceptibility of algorithm.

[18] Here in this paper the various lossy compression techniques are to be examined on various physics factors. These techniques reduce the impact of low bandwidth. It also describes the simulation strategy for data compression in the systems. It also introduce that lossy is much more effective than other compression techniques because it uses less time and are more efficient.

4. COMPARATIVE ANALYSIS OF PARAMETERS IN EXISTING LITERATURE

This table present comprehensive comparison of parameters used in existing literature. Parameters strength and shortcomings are also highlighted so that optimal technique can be selected for future work.

Table 1: Comparison of techniques used in Data Deduplication

Author	Parameters Method Used	Encryption technique	Complexity	Storage	Security Strength	Performance enhanced
R. Chen, Y. Mu, G. Yang, and F. Guo[1]	BL-MLE	RSA Technique	High, In terms of iterations required	Repeated bits present hence high storage requirements	Least since no security slandered are followed	Better Handling of redundancy
R. Miguel[2]	Homomorphic Deduplication	Homomorphic Encryption	High, in terms of length of code	Bit level duplication showing high storage	Key is used hence medium security is present	Better data Handling

G. Zhu, X. Zhang, L. Wang, Y. Zhu, and X. Dong[3]	Intelligent Back Up system	RSA Technique	Length of code ensure complexity	Multiple copies of data ensures high storage	No security slandered established	Back is improved
H. Nagarajaiah, S. Upadhyaya, and V. Gopal[4]	Embedded Processor Dedupliation	AES Algo	Calculations in AES makes it complex	Repeated data elimination techniques ensures less storage requirements	Private and public keys ensure high security	Security is improved
S. C. Satapathy, P. S. Avadhani, S. K. Udgata, and S. Lakshminarayana [5]	Critical Infratructure is considered	None	Length of code in terms of LOC is less	Redundancy is handled hence space requirement is low	No security standards are used	Infrastructure capabilities are considered
J. J. Park, A. Zomaya, H.-Y. Jeong, and M. Obaidat[6]	Frontier Technique	None	Multiple algorithms ensure complexity	Length of code required maximum storage	Security standards ensure high security	A new Technique where storage compression is considered
K. He, C. Huang, H. Zhou, J. Shi, X. Wang, and F. Dan[7]	Public auditing technique	RSA algorithm	Complexity in terms of calculations is high	Storage requirements is high since LOG is maintained	Security Mechanisms are used	Public auditing for encrypted data with client-side deduplication in cloud storage
X. Li, J. Li, and F. Huang[8]	Fuzzy deduplication	Fuzzy Algorithm	Low complexity since logical values in terms of 0 and 1 is used	Low storage requirements since result is stored in terms of Boolean values	Fuzzy storage has least security associated with it	A secure cloud storage system supporting privacy-preserving fuzzy deduplication
N. Christin and R. Safavi-Naini, Eds[10]	Financial cryptography	Cost Based Encryption	Complexity in terms of calculations is high	Redundancy is handled hence space requirement is low	Security Mechanisms are used	A cost is a factor on which deduplication is considered
F. Rashid, A. Miri, and I. Woungang[11]	Secure Enterprise Data Deduplication	AES Algo	Complexity in terms of calculations is high	Redundancy is handled hence space requirement is low	Security Mechanisms are used	High security in data deduplication
W. K. Ng, Y. Wen, and H. Zhu[12]	Private cloud data deduplication	None	Length of code in terms of LOC is less	Storage requirements are high in terms of multiple data	Security is low since no security slandered s are used	Only private cloud is considered
C. Wang, Q. Wang, K. Ren, and W. J. Lou[13]	Ensuring Data Storage Security in Cloud Computing,	AES Encryption	Complexity in terms of calculations is high	Redundancy is handled hence space requirement is low	Security Mechanisms are used	Data security is high
Y. Yuan, X. Wu, and Y. Lu, Eds[14]	<i>Trustworthy Computing and Services</i>	None	None	None	None	Trust parameter is considered

C.-I. Fan, S.-Y. Huang, and W.-C. Hsu[15]	Hybrid data deduplication in cloud environment	Cost based Encryption	Complexity in terms of calculations is high	Redundancy is handled hence space requirement is low	Security based standards are used	Hybrid deduplication is considered
F. Rashid, A. Miri, and I. Woungang[16]	A secure data deduplication framework for cloud environments	Encryption based on cost	Complexity is high in terms of LOC	Storage requirements are high since code is complex	Data security mechanism are required	A security is provided
X. Zhang and J. Zhang[18]	Data Deduplication Cluster Based on Similarity-Locality Approach	Cluster based encryption is used	Complexity is high since cluster of information is present	High storage requirements in terms of clusters	Encryption standards ensure high	Cluster based approach is used

Following table present comparison of strategies in terms of parameters. Work satisfying most of the parameter list is optimal one. The table listed below indicates that [8] is optimal in nature.

Table 2: Comparison in terms of parameters

Paper	Encryption	Compression	Minimum Storage	Performance
X. Li, J. Li, and F. Huang[8]	✓	✓	✓	✓
W. K. Ng, Y. Wen, and H. Zhu[12]	X	✓	X	X
C. Wang, Q. Wang, K. Ren, and W. J. Lou[13]	✓	✓	X	✓
F. Rashid, A. Miri, and I. Woungang[16]	✓	✓	X	X
X. Zhang and J. Zhang[18]	✓	X	X	X

5. CONCLUSIONS

The various deduplication techniques have been surveyed in this paper which are used to handle the redundancy in data. By closely observing that we have concluded that the only that technique is good for data deduplication if it is variable sized as compared to other techniques. This strategy improves the performance and storage efficiency of data centres that hold the data and the storage resources can maximise their capacity to hold the data by removing redundant data. In future more research work can be done on the variable sized deduplication techniques to develop an efficient method for high throughput.

6. REFERENCES

- [1] A. Celesti, F. Tusa, M. Villari, and A. Puliafito, "An Approach to Enable Cloud Service Providers to Arrange IaaS, PaaS, and SaaS Using External Virtualization Infrastructures," 2011.
- [2] S. Lee, J. H. Song, and I. K. Kim, "CDA Generation and Integration for Health Information Exchange Based on Cloud Computing System," vol. 1374, no. c, 2014.
- [3] R. Buyya, "Introduction to the IEEE Transactions on Cloud Computing," vol. 1, no. 1, pp. 3–21, 2013.
- [4] K. Bousselmi, "Energy efficient partitioning and scheduling approach for Scientific Workflows in the Cloud," 2016.
- [5] U. Wajid, C. Cappiello, P. Plebani, B. Pernici, N. Mehandjiev, M. Vitali, M. Gienger, K. Kavoussanakis, D. Margery, D. G. Perez, and P. Sampaio, "On Achieving Energy Efficiency and Reducing CO₂ Footprint in Cloud Computing," vol. 7161, no. c, 2015.
- [6] I. Transactions, O. N. Circuits, S. For, and V. Technology, "Introduction to the Special Section on Visual Computing in the Cloud: Cloud Gaming and Virtualization," vol. 25, no. 12, pp. 1955–1959, 2015.
- [7] A. On and Y. For, "Introduction to the Special Section on Visual Computing in the Cloud: Fundamentals and Applications," vol. 25, no. 12, pp. 1885–1887, 2015.
- [8] J. H. Abawajy, S. Member, and A. Kelarev, "Large Iterative Multitier Ensemble Classifiers for Security of Big Data," vol. 2, no. 3, 2014.
- [9] X. Ding, J. Shan, and S. Jiang, "A General Approach to Scalable Buffer Pool Management," vol. 9219, no. c, 2015.
- [10] N. Kaurav, "An Investigation on Data De-duplication Methods And it's Recent Advancements De-duplication De-Duplication Types," 2014.

- [11] D. T. Meyer and W. J. Bolosky, "A Study of Practical Deduplication."
- [12] Q. He, Z. Li, and X. Zhang, "Data Deduplication Techniques," pp. 430–433, 2010.
- [13] T. Y. J. Nagamalleswari, "Deduplication Techniques : A Technical Survey," vol. 1, no. 7, pp. 318–325, 2014.
- [14] A. Info, "PERFORMANCE COMPARISON OF DEDUPLICATION TECHNIQUES FOR STORAGE IN," vol. 5, pp. 42–45, 2014.
- [15] "ADVANCED DATA DEDUPLICATION TECHNIQUES AND THEIR," 2013.
- [16] J. Li, Y. K. Li, X. Chen, P. P. C. Lee, and W. Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication," pp. 1–12, 2014.
- [17] G. Sethi, S. Shaw, K. Vinutha, and C. Chakravorty, "Data Compression Techniques," vol. 5, no. 4, pp. 5584–5586, 2014.
- [18] D. Laney, S. Langer, C. Weber, P. Lindstrom, and A. Wegener, "Assessing the effects of data compression in simulations using physically motivated metrics 1," vol. 22, pp. 141–155, 2014.