Traffic Redundancy Elimination and Bandwidth Optimization Over Cloud

Iti Jain
M. Tech. Research Scholar
Lakshmi Narain College of Technology,
Bhopal (M.P.), [INDIA]
Email: itijain390@gmail.com

Sweta Gupta
Senior IT Faculty
iNurture Education Solutions
Jagran Lakecity University
Bhopal (M.P.), [INDIA]
Email: 6.shwetagupta@gmail.com

Hitesh Gupta
Assistant Professor
Lakshmi Narain College of Technology,
Bhopal (M.P.), [INDIA]
Email: hitesh034@gmail.com

Dr. Vineet Richhariya
Head of Department
Department of Computer Science
Lakshmi Narain College of Technology,
Bhopal (M.P.), [INDIA]
Email: vineetrich100@gmail.com

Abstract—*Cloud computing or on-demand computing offers various on demand services to the user. User can simultaneously access these services. That increases the bandwidth cost, thus TRE (Traffic Redundancy Elimination) techniques are used to reduce redundant data from the cloud and reduce cost of the bandwidth. Generally there are two type of techniques called sender based redundancy elimination and receiver based redundancy elimination are used to eliminate redundant data from the cloud. Thus mutual cooperation of the sender and receiver can be used to provide enhanced functionality to remove redundant data and optimize the cloud bandwidth. But security is also important in cloud. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments using our prototype. A new technique which uses convergent encryption and mutual cooperation of the sender and receiver to provide a secure data redundancy elimination is proposed in this paper.*

Keywords:—*Cloud Computing, TRE (Traffic redundancy Elimination), PACK (Predictive ACKs).*

1. INTRODUCTION

Cloud computing is a computing in which various resources are provided to the user. By the use of these resources user can access these resources as per their use. A huge amount of data is flooded over the internet that data contains redundant data. Thus due to that huge data problem of traffic congestion can be occurs. To provide an optimized traffic management to the user elimination of that redundant data is required.

There are various cloud service provider which used to provide various on demand services to the user. A description over that is presented in the figure 1.1. in that there are vendors like Google, Amazon, Microsoft, VMware, AT&T etc. which used to provide services to the user.

Figure 1.1: Cloud Service Providers.

To provide uninterrupted services to the user, a traffic management is required. Because there a large amount of data is flooded over the internet which requires proper management to remove redundant data. There are various eliminating techniques like in [1] a predictive ACK based technique is presented which provides prediction about the redundant data to eliminate that data. In [2] a receiver based selection strategy is presented which provides a traffic redundancy elimination at receivers end. A SHA-1 based technique is used to provide redundancy reduction to the user, in [3] a cooperative redundancy reduction based technique is presented which provides an enhanced framework to reduce redundancy in the data. Because a sender based or a receiver based approach is not enough to provide solution for such problems. A mutual cooperation between sender and receiver is used to provide a redundancy mechanism to the user.

But in cloud security of the data also one of the prime concern thus a convergent encryption based technique which uses mutual cooperation of sender and receiver to remove duplicate data communication channel. And optimize the bandwidth and cost in cloud.

Redundancy in cloud computing can be defined as the supplying of duplicate copies of various data, equipment, systems, or the like, to be used in the event that part of one's cloud computing system fails or cannot be accessed. This redundancy is made available by having fully replicated data several times on multiple computers or units involved in the same data center.

2. LITERATURE REVIEW

Eyal Zohar, Israel Cidon, and Osnat Mokryn [1]

Pack (Predictive ACKs), a traffic redundancy elimination solution is presented. A receiver based, end to end traffic redundancy elimination mechanism is provided to optimize bandwidth in cloud. Prediction at the receiver's end is made to eliminate traffic between the cloud and end to end users. A monitoring of the input stream is conducted then matching of the chunk is performed. Result of this long time chunk matching is sent to the sender's end. That content helps to make the decision about TRE operation. A hint about the TRE is sent to the sender. Sender perform TRE operation when that hint matched found. After the whole operation a prediction ACK is send to the receiver's end. At receiver's end a lightweight chunking scheme is used, which enhances the performance of computation speed and provide better performance to eliminate redundancy in the traffic.

Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou [2]

Data deduplication is a technique which used to eliminate redundant data from the database and reduce storage space and bandwidth. An authorized deduplicate scheme is presented. That technique used an authorized redundancy elimination technique to optimize the bandwidth in cloud. In convergent encryption scheme user's original copy of the data is used generate convergent key and that convergent key used to encrypt the original copy. User also derive a tag for data copy, such that if two data copies are same then the tags of these copies are also same. That tag further used to remove duplicate copies of the data. That technique used to provide an authorized deduplication mechanism which used a token based scheme to eliminate duplicate data in a hybrid cloud scenario. That a secure outsourcing mechanism is provided to the user which enhances the functionality of the whole deduplication mechanism[7].

Lei Yu, Haiying Shen, Karan Sapra, Lin Ye and Zhipeng Cai [3]

Cloud computing provides various on-demand services to the user. In that various user access these services simultaneously that generate congestion in data transmission channel. Traffic redundancy elimination can be used to enhance the bandwidth capacity and also reduce cost of the bandwidth. But to eliminate TRE there either a sender based TRE

elimination technique or receiver based TRE technique can be used. But these techniques are not efficient to provide better solution for TRE. CORE (Cooperative end to end traffic redundancy elimination) is presented, that technique uses both a sender and receiver based traffic redundancy elimination technique to provide an enhanced mechanism for TRE. In that technique cooperative operations between the sender and receiver is used to remove the redundant data from the bandwidth.

Swathi Kurunji, Tingjian Ge, Benyuan Liu, Cindy X. Chen [4]

A read optimized databases provides better performance for various read intensive data warehouse applications. In these applications data increases rapidly a flexible and dynamically environment like cloud is required to provide better performance to the user. But there is an efficient query mechanism is required to provide better performance to store data in these databases. In cloud scenario nodes are increases rapidly thus that generate inter-node communication cost that also generate a huge amount of data which requires an enhanced functionality to handle that data. A storage structure called PK-map and query processing technique is used to reduce the inter node communication cost. In that structure primary key and foreign key of the tables can be used to reduce the inter node communication from the various databases.

Lluis Pamies-Juarez, Pedro Garcia-Lopez, Marc Sanchez-Artigas, Blas Herrera [5]

"Towards the Design of Optimal Data Redundancy Schemes for Heterogeneous Cloud Storage Infrastructures" analyze how distributed redundancy schemes can be optimally deployed over heterogeneous infrastructures. Specifically, they are interested in infrastructures where nodes present different online availabilities. Considering these heterogeneities, they present a mechanism to measure data availability more precisely than existing works. Using this mechanism, they infer the optimal data placement policy that

reduces the redundancy used, and then its associated overheads up to 70%.

A. Gupta, A. Akella, S. Seshan, S. Shenker, and J. Wang, [6]

Cloud computing brings significant benefits for service suppliers and users due to its characteristics: e.g., on demand, gets use, scalable computing. Virtualization management may be an important task to accomplish effective sharing of physical resources and scalability. Transmission price plays a crucial role once attempting to minimize cloud price. But for server specific TRE approach it's troublesome to handle the traffic with efficiency and it doesn't suites for the cloud atmosphere due to high process prices. During this paper we have a tendency to provide a survey on the new traffic redundancy technique called novel-TRE conjointly called receiver based mostly TRE. This novel-TRE has vital options like police investigation the redundancy at the shopper, repeats seem enchained, matches incoming chunks with a antecedently received chunk chain or native file and causing to the server for predicting the longer term information and no want of server to ceaselessly maintain shopper state, our implementation maintains chains by keeping for Associate in Nursing chunk solely the last discovered ensuant chunk in an LRU fashion. So on the receiver aspect we are able to refresh the chunk store for incoming chunks.

Zhifeng Xiao and Yang Xiao,IEEE June 2013 conference – "Security and Privacy in Cloud Computing" [7]

They have worked on various attribute confidentiality, integrity, availability, accountability, and privacy-preservability and performed the various security concern issues in aspects, authors have systematically studied the security and privacy issues in cloud computing based on an attribute-driven methodology, We have identified the most representative security/privacy attributes (e.g., confidentiality, integrity, availability, accountability and privacy- preservability), as

well as discussing the vulnerabilities, which may be exploited by adversaries in order to perform various attacks. Defense strategies and suggestions were discussed as well, thus this is the paper included the security and study aspects in cloud computing, the data integrity verification made dealing with encryption algorithm and the audit was performed with the help of hashing algorithm available in order to verify the value generated again while checking the data integrity available with the associated file, here they have worked on different aspects such as user account access approach, availability of data, data changing or integrity verification and the technique should be privacy preserving so that the data should not be leak during the cloud execution.

A comparison is performed in Table1 below.

3. PROPOSED WORK

Traffic redundancy elimination (TRE) is one of the technique which provides solution to reduce redundant data from the bandwidth which reduce the cost of the cloud band width. Techniques like PACK (Predictive ACKs) [4] which uses a receiver based mechanism to provide traffic redundancy mechanism in cloud. Some sender based TRE techniques can be used to provide traffic redundancy elimination mechanism in cloud. But only sender based technique or receiver based technique is not able to provide better solution to the user. Technique like CORE (cooperative end to end Traffic Redundancy Elimination)[6] which uses mutual cooperation between the sender and receiver to eliminate the traffic redundancy in the cloud data. A redundancy prediction and topical and characteristics based

Table 1- Comparison Analysis of different approach.

Technique	Avg. Chunk size	Sender chunking	Sender signing	Receiver Chunking	Receiver Signing	File type	Remark
P A C K (Existing)	No limit	None	SHA-1	P A C K - c o n t e n t based	SHA-1	Limited to t e x t u a l content	Less efficient for large data, limited and high band-width con-sumption.
End RE chunk matching	Limited to 32-64 byte	S a m p l e byte	SHA-1	None	None	Only text data	Limited to content and No security Mecha-nism.
E P A C K (Proposed)	No Limit	Chunking with meta data ex-traction+ Similarity m e a s u r e c o m p u t a-tion	SHA-2	E P A C K - C o n t e n t b a s e d + Meta based approach + similarity score based approach	SHA-2	Applicable to all file formats	E f f i c i e n t with multi-ple file over simi-l a r i t y m e a s u r e technique, E f f i c i e n t with large h a s h i n g value.

system is going to have following characteristics:

- The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.

- We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys.

- Reduce the storage size of the tags for integrity check. To enhance the security of DE duplication and protect the data confidentiality.

The table 1, above demonstrate how the mechanism differentiate in themselves for operations.

But security of the data is also one of the biggest concern in cloud computing. Thus a convergent encryption [5] based technique which uses mutual cooperation of the sender and receiver to eliminate redundancy in cloud is proposed in this paper. That technique provides a secure redundancy elimination mechanism in cloud.

4. CONCLUSION

TRE (Traffic Redundancy Elimination) is one of the effective solution to optimize the cost of the bandwidth in cloud. Various techniques like PACK (Predictive ACKs) a receiver based traffic redundancy elimination technique and some sender based traffic redundancy elimination technique are used to optimize the bandwidth of the cloud. Thus a mutual cooperation of the sender and receiver can be used to provide batter solution to remove redundant data from the cloud bandwidth and reduce the cost of the bandwidth. A secure TRE technique which uses convergent encryption and mutual cooperation of the sender and receiver to remove redundant data from the bandwidth proposed in this paper. A new secured, cost effective, highly available multi-cloud architecture for enabling privacy preserving outsourced storage of data has been introduced.

REFERENCES:

[1] Eyal Zohar, Israel Cidon, and Osnat Mokryn "PACK: Prediction-Based Cloud Bandwidth and Cost Reduction System" IEEE, 2014.

[2] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou "A Hybrid Cloud Approach for Secure Authorized Deduplication" IEEE, 2014.

[3] Lei Yu, Haiying Shen, Karan Sapra, Lin Ye and Zhipeng Cai "CoRE: Cooperative End-to-End TrafficRedundancy Elimination for Reducing Cloud Bandwidth Cost" IEEE, 2016.

[4] Swathi Kurunji, Tingjian Ge, Benyuan Liu, Cindy X. Chen "Communication Cost Optimization for Cloud Data Warehouse Queries" IEEE, 2012.

[5] Lluis Pamies-Juarez, Pedro Garc__a-L_opez, Marc S_anchez-Artigas, Blas Herrera, "Towards the Design of Optimal Data Redundancy Schemes for Heterogeneous Cloud Storage Infrastructures" Computer Networks, 2011.

[6] A. Gupta, A. Akella, S. Seshan, S. Shenker, and J. Wang, "Understanding and exploiting network traffic redundancy" UWMadison, Madison, WI, USA, Tech. Rep. 1592, Apr. 2007.

[7] Zhifeng Xiao and Yang Xiao, Senior Member, IEEE, "Security and Privacy in Cloud Computing", IEEE 2013.