© 2018 IJSRST | Volume 4 | Issue 8 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Optimized and Secure Multiple Predictions Based Traffic Redundancy Elimination - A Survey

A. Anto Viji[*1], Dr. J. Jasper.
[*1]M.E Reseach Scholar, CSI Institute of Technology, Thovalai, Tamil Nadu, India
[2]Ph.D, Department of Electrical and Electronics Engineering, Ponjesly College of Engineering, Nagercoil, Tamil Nadu, India

ABSTRACT

Cloud computing is a specific term that involves delivering hosted services over the Internet, Cloud offers various types of on demand services that enable users to access simultaneous computing capabilities. Traffic Redundancy elimination is the most common factor in cloud computing. Cloud Computing involves mostly based on the concepts of high performance computing. Cloud Storage delivers virtualized storage on demand, over a network based on a request for a given quality of service. Redundancy elimination can be enhanced by mutual perceptive between the sender and receiver. Using either sender-based TRE or receiver-based TRE cannot simultaneously capture traffic redundancy in both short-term) and long-term data redundancy, which concurrently appear in the traffic. Cloud security is a major risk factor in cloud computing. In order to enhance the security and elimination of redundant data can be done by a novel proposed algorithm based on neural network schema.

Keywords : Cloud Computing, Neural Networks, Traffic redundancy Elimination, Optimization

I. INTRODUCTION

Cloud Computing is a way of using a network of remote servers hosted on the Internet for storage, data processing and to manage large rather than a local server or a personal computer. A large amount of data is flooded over the internet contains redundant data. Large amount redundant data storages leads to inefficient use of precious resources in cloud. Several cloud providers are available to provide services to users. Examples of cloud providers are amazon, google, Microsoft, VM ware etc. Traffic redundancy leads to the interruption of data present in the cloud. To provide un interrupted services in cloud traffic redundancy elimination is a major task.

There are several techniques are used to eliminate traffic redundancy in cloud. Techniques like PACK[[1] uses prediction based technique to eliminate redundant data with the help of SHA-1 algorithm, ENDRE[2] provides receiver based elimination technique to eliminate redundancy. WANAX [9] uses Middle box based algorithm to eliminate redundant data. Such sender or receiver based elimination of redundant data does not provide complete solution for such problems. Mutual co-relation between sender and receiver along with neural network scheme is used to provide a redundancy elimination mechanism to clients.

Cloud security is also a major factor for data storage in cloud environment, convergent encryption based technique used in mutual cooperation of sender and receiver to remove replica data in communication channels and optimize the bandwidth and cost in cloud. Redundancy in cloud is the process of sending replicated data stored in various parts of cloud, when a cloud computing system fails to send or receive data

the entire files in cloud 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 leads to higher costs.

II. LITERATURE SURVEY

Bandwidth cost of data transfer reduces from the cloud using Traffic Redundancy Elimination (TRE) technologies has been moralized [10]. Redundancy has been found in the network traffic [11], [12], due to common accesses to the same or similar data objects from the Internet end-users, by eliminating the transmission of replica information can reduce the bandwidth usage cost. A number of Traffic redundancy elimination techniques has been introduced for WAN optimization that provided an end to end solution EndRE [2] deployed at client and server side. Caches play an important role in order to synchronize data at sender and receiver side. Sender compares data from its local cache and detects redundant data available by sending the reference of the replicated data to the receiver instead of transmitting the entire data. It has been inefficient for cloud platforms that EndRE require stretched synchronization for maintaining caches at two ends, which is costly to maintain.

Additional usage of computation and storage resources in cloud platforms. Recently, a receiver-based TRE solution named PACK [1] is proposed to address the above issues arising in cloud environment. In PACK, once a client receives a data chunk that already available in its local cache, it is expected that the future coming data are also matched with its cached data. The client makes predictions for future coming data chunks and notifies the cloud server. PACK maintain cloud Elasticity and reduces computation, storage and traffic redundancy. It Uses the scale of time persistent data storage. Since long term persistent data storage make way for the long term usage of disks in which the data are stored which requires a huge amount of space. EndRe reduce short term redundant data elimination and PACK reduces Long Term redundant data elimination. PACK cannot exploit full term of network traffic redundancy.

III. OBJECT LEVEL APPROACHES.

Object level approaches uses [13] web objects in nature. Web objects have an impact on the access patterns and caching. Most of the objects required to access the objects at the same time which leads to network traffic and higher usage of bandwidth. During the file downloading all the data stored in the cache memory leads to the redundancy of data and shared among network. Cache storage in client or receiver side has a limit in the network access. Redundancy elimination during file downloads has also received a large amount of attention. Content based approach is used to divide the data into chunks and download only that chunk which are already present in the cache. Object level approaches are purely based on the data centric approaches which do not focus on the repetition of contents within the data chunks. Packet level content shares different data pool of users which results in the higher usage of caching and sharing of packet level.

Data de duplication scheme used to eliminate redundant data from the data base to ensure minimum amount of storage and bandwidth usage. Jin LI etal [3] proposed an authorized de duplication scheme for eliminating redundancy and to optimize bandwidth in cloud. Authorized de duplication scheme uses a convergent key to encrypt original copy of data from user. Original copy of data is used to generate convergent key and convergent key is used to encrypt the original data. Client derives a tag for data copy, in order to ensure that if data copy are same then tag are also same. Client tag is used to remove the duplicate copy of data. A token based mechanism is used to eliminate duplicate copy of data in hybrid clouds.

Lei Yu, et al [4] proposed a traffic redundancy scheme that provides services simultaneously and generate congestion in data transmission channel. TRE based elimination techniques can be done at both the sender and receiver based on the redundancy schemes. Co-operative based redundancy schemes uses both the sender and receiver based traffic redundancy elimination and to remove redundant data from the bandwidth itself.

Swathi Kurunji et al [5] proposed a technique for communication cost optimization which involves a read optimized data bases for providing better performance of data base ware house applications. In data warehouse applications data increases rapidly and a flexible, dynamically environment was required to provide better performance to the user. A strong query mechanism is required to provide better performance to store data in data bases. Due to the rapid increase in nodes an enhanced inter communication cost is required to improve the functionality of each nodes. Inter node communication cost increases when a large amount of data is handled. Primary Key map used to reduce inter node communication cost but the use of primary key and foreign key increases the delay in the network from collecting data from various database tables.

Lluis Pamies-Juarez, Pedro et al [6] proposed Towards the Design of Optimal Data Redundancy Schemes for Heterogeneous Cloud Storage Infrastructures described about distributed redundancy schemes over heterogeneous infrastructures. Heterogeneous cloud networks interested in infrastructures different type of nodes present different online capabilities. An optimal data placement policy is required to present a mechanism to measure data availability more precisely than existing works. Optimal data placement policy that reduces the redundancy used and overhead increase up to 70%.

Gupta, et al [7] proposed network redundancy traffic method in cloud computing provides significant benefits for service suppliers and users due to its characteristics: e.g., on demand, gets use, scalable computing. Virtualization management is an important concept of establishing effective shared physical resources and scalability. Server specific TRE approach is inefficient of handling troublesome 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 TREand Sender based TRE. This novel-TRE has a concept of redundancy detection and reduction at the sender and receiver the server for predicting higher term of information.

Zhifeng Xiao et al [8] proposes Security and Privacy in Cloud Computing‖ [7] They have worked on various attribute confidentiality, integrity, availability, accountability, and privacy. Preservability in privacy provides a higher range of security aspects and privacy issues in cloud computing based on an attribute-driven methodology. Most representative security and privacy attributes are confidentiality, integrity, availability, accountability and privacy preservability.

IV. PACKET-LEVEL APPROACHES

EndRE [2] manily focused on sender-based end-to-end TREfor large enterprise based networks. A new was chunking scheme was implemented faster than commonly-used Rabin fingerprint Algorithm, and restricted to chunks as small as 32-64 bytes. PACK, EndRE requires the server that has to maintain a fully and reliably synchronized cache for each and every client. To remain with the server memory requirements cache size are kept small, leads the client to be inadequate for medium to large content data or long term redundancy. EndRE is server specific scheme and not suitable for cloud

environment to provide best redundancy reduction method.

Wanax [17] is a TRE system where storage and WAN bandwidth are insufficient to send and receive data. Wanax is software based approach on which the middle box replacement is expensive commercial hardware sectors. Sender middle box holds TCP stream for each and every data transmission and sends data signatures to the receiver middle box. The receiver checks the entire data is available in its local cache, and if available it sends the data to client. Data from the sender to receiver are splits in to small portions chunks. Data chunks those are not present in the cache memory are collected from the sender middle box or nearby middle box which increases the latency for non cached data.

N. T. Spring et al [11] proposed independent packet level TRE solution uses rabins finger print by applying hash function to each 64 byte sub string of the packet content and selects a subset of fingerprint representing the recently transferred packets. Sender checks whether its current fingerprints have appeared in earlier stored packets. If present the sender identifies maximal overlap region around each and every matched fingerprint and replaces the finger print with a fixed size pointer value into the cache region. The sender is decoded with the fingerprint and compresses the data before sending. In order to decode compressed data, receiver has to replace the corresponding pointer data to local cache. Local Cache stores the currently received packets in both sender and receiver. Traffic redundancy in independent protocol packet level is about 75% and implies end to end elimination which results in the usage of bandwidth.

V. PROPOSED

Traffic redundancy is one of the most common end user activities in retrieving data. A distributed optimized dynamics of mutually co related connected neural networks are used to select heterogeneous cloud storages. Cloud networks optimize the network traffic without any centralized computation. Mutually connected network can be used for large scalable networks without compromising the network security. Sender based elimination or receive based redundancy elimination is not able to provide better solution for network traffic reduction and redundancy elimination. But a combination of both sender and receiver based elimination provide a better solution for network traffic and redundancy elimination.

Redundancy Elimination based on neural networks has the following characteristics

Users are allowed to perform security privileges and check the replicated data files in the network. Security on cloud networks is high when compared to the existing approaches. Storage cost and computation cost reduces on neural network based approach. Integrity and Confidentiality of Data present in cloud are maintained Security is the most important factor in cloud computing and also in a major concern. Mutual cooperation concern and neural network scheme increases the redundancy elimination and Security in homogeneous as well as heterogeneous cloud. Comparisons of the existing schemes are given in the table 1.1

Table 1.1 Comparison of Existing Approaches

Method	Data Size	Approach	Signature Type	File Type	Advantage	Disadvantage
PACK	Any Size Data	TRE Based Approach	SHA - 1	Text	Reduces Computation cost, Eliminating Redundancy based on the content from the sender	Limited Latency on cloud mobility and elasticity, Less efficient on large data, Bandwidth consumption is higher
ENDRE	32-64 byte of limited data	Sample Byte based approach	SHA-1	Text	Sender based solution, Cache synchronization is high	Suitable for enterprise networks, Not suitable for long term traffic redundancy, Processing and Memory Cost is high on servers
WANAX	Limited Data	Middle Box TCP stream approach		Text	Reduces latency for cached data	Expensive on commercial servers, Higher latency on non cached data, Limited security on contents
Rabins Fingerprint	64 Bye Data	Signature Based Approach	SHA-1	Text	Compress the data sizes, Security on data Contents, Independent of protocol	Only a limited amount of data is transferred, Computation cost is high on servers
Neural Network Schema (Proposed)	Any Size of Data	Fuzzy logic based Approach	SHA - 2	All files	Uses simple fuzzy sets to retrieve data from cache as well as from the servers, Highly secured and traffic redundancy elimination is high, Minimal Computation cost	

VI. CONCLUSION

Applied distributed optimization dynamics of mutually connected neural network to select optimal heterogeneous type cloud and homogeneous cloud storages to evaluate the performance of the proposed approach. Neural network based approach mutually connected neural network directly optimizes objective functions for the entire networks by distributed computation on each terminal of wireless cloud storages. The proposed algorithm can optimize various kinds of complicated objective functions without any centralized computation. Centralized Computation cost of cloud storages is reduced. New secured, cost effective, highly reliable heterogeneous multi cloud architecture for enabling privacy preserving outsourced storage of data has been introduced.

VII. REFERENCES

1. Eyal Zohar, Israel Cidon, and Osnat Mokryn "PACK: Prediction-Based Cloud Bandwidth and Cost Reduction System" IEEE, 2014.
2. B Agarwal, A. Akella, A. Anand, A. Balachandran, P. Chitnis, C. Muthukrishnan, R. Ramjee, and G. Varghese, "Endre: An end-system redundancy elimination service for enterprises," in NSDI, 2010, pp. 419– 432.
3. Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou "A Hybrid Cloud Approach for Secure Authorized Deduplication" IEEE, 2014.
4. 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.
5. Swathi Kurunji, Tingjian Ge, Benyuan Liu, Cindy X. Chen "Communication Cost Optimization for Cloud Data Warehouse Queries" IEEE, 2012.
6. 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.
7. 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.
8. Zhifeng Xiao and Yang Xiao, Senior Member, IEEE, "Security and Privacy in Cloud Computing", IEEE 2013.
9. S Ihm, K. Park, and V. Pai. Wide-area Network Acceleration for the Developing World. 2010.
10. E. Zohar, I. Cidon, and O. O. Mokryn, "The power of prediction: cloud bandwidth and cost reduction," in ACM SIGCOMM, 2011, pp. 86–97.
11. N. T. Spring and D. Wetherall, "A protocol-independent technique for eliminating redundant network tra_c," in ACM SIGCOMM, 2000, pp. 87–95.
12. A. Anand, C. Muthukrishnan, A. Akella, and R. Ramjee, "Redundancy in network traffic: findings and implications," in SIGMETRICS /Performance, 2009, pp. 37–48.
13. L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web caching and zipf-like distributions: Evidence and implications. In IEEE Infocom, 1999.