ISSN:0975-9646

Enhancing Security of Traffic Redundancy and Elimination Approach in Cloud Data Storage using 3DES

Anisha Sarah Mathew[1], Mitha Rachel Jose[2]

[1] PG Scholar, [2]Assistant Professor, Department of Computer Science and Engineering, MG University
Mangalam College of Engineering, Kottayam, Kerala-India

Abstract— **Cloud computing refers to the delivery of computing resources over the internet. Cloud services are popular since they can reduce cost and complexity of owning and operating computers and networks. But it is important to provide the convenient pricing model for users of cloud. Here, bandwidth reduction is also an issue. Hence a new traffic redundancy elimination scheme was designed for reducing cloud bandwidth and costs. In this paper, Predictive acknowledgement (PACK) is presented which is a destination-to-destination traffic redundancy elimination system. PACK is based on a technique, which allows the client to use newly received chunks to identify previously received chunk chains. Cloud related TRE needs to apply a judicious use of cloud resources so that bandwidth cost reduction combined with the extra cost of traffic redundancy computation and data storage would be optimized. The non redundant data is identified and using triple DES (3DES) those data chunks are encrypted and sent to the cloud for storage. Finally, analysis and implementation of Predictive acknowledgement benefits for cloud users is determined.**

Keywords-Cloud Computing, PACK,TRE ,Bandwidth,3DES

I. Introduction

Cloud computing [5] is a type of computing that relies on sharing resources rather than having local servers or personal devices to handle applications. Cloud services allow individuals and businesses to use software and hardware that are managed by third parties at remote locations. The cloud computing allows access to information and computer resources from anywhere that a network connection is available. Success of cloud computing customers is mainly due to the ability of the customer to use on-demand services with a pay-as-you-go model. Reasonable price and high flexibility make customers to migrate to cloud compelling. Cloud computing [5] is the long dreamed vision of computing as a utility, where users can remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources.

Various traffic redundancy techniques are used for the judicious use of cloud's resources, of which Traffic Redundancy elimination is used for reducing bandwidth and costs. Traffic redundancy arises from commonly used end-user's activities [8] ,such as repeatedly accessing, uploading, downloading, distributing and modifying the same or similar information. Traffic Redundancy Elimination System [17] is used to eliminate the transmission of redundant content of data and information, mean while it significantly reduces the cost of network. In most common Traffic redundancy elimination scenarios [13][14], both sender and receiver examine the data chunk and find match between the signatures of all data chunks, parsed according to the content, prior to their transmission. If a redundant chunk is found, sender replaces the transmission of each redundant chunk with its signature.

Existing destination-to-destination Traffic Redundancy Elimination scenarios [3] are sender-based. In this case, the cloud server directly sends the data chunks, here the server continuously maintain client's status. It is believed that a destination-to-destination, software based TRE should meet the following desirable properties:

1) Standard: The traffic redundancy elimination scenario is supposed to work across all server and client platforms and operating systems. Regardless of client nature and location, this thus enables servers to reduce redundant traffic.
2) Application Independent: The traffic redundancy elimination should support majority of the applications that transmit redundant information.
3) High server performance: It should limit the size of a traffic redundancy elimination specific buffering and amount of processing overheads due to expensive lookups and data computations.
4) Minimum impact on end-to-end latency: TRE protocols introduce additional traffic latencies even when TRE is not in effect. The standard TRE should minimized additional latencies.

In this paper, a low latency, low overhead receiver-based destination-to-destination TRE scenario [1] is presented that uses a prediction based approach to eliminate redundantly occurring traffic between the cloud and its end-users. Here each receiver observes the incoming stream of chunk chain to match with a previously received chunk chain. The receiver sends to the server predictions that include chunk's length, chunk's hint of the sender's future data. The sender first examines the hint and then traffic redundancy elimination is performed based on hint-match basis [8].

The objective of this paper is to reduce the expensive traffic redundancy elimination(TRE) computation at the sender side by eliminating the traffic redundancy. To improve the efficiency of the security, an encryption scheme is being used. The proposed system deals with a secure prediction based system in which triple DES [12] algorithm is being used for encrypting and decrypting the data chunk. Thus a level of security is being achieved. Hence the original data cannot be seen, only an encrypted form is seen.

The paper is organized as follows: Related work is described in Section II. Sections III present a receiver-based TRE solution and describe the predicting process and prediction-based traffic redundancy elimination mechanism. The experimental evaluation is presented in Section IV

II. RELATED WORK

Various Traffic redundancy elimination techniques such as independent TRE, packet level TRE have been explored in recent years. All these techniques describe how to get away with the three-way handshake between the sender and the receiver if a full state synchronization is maintained.

A receiver oriented destination-to-destination Traffic Redundancy Elimination Predictive Acknowledgment solution is proposed for cloud computing application. The stream of data received at the PACK receiver is parsed to a sequence of variable size, content based signed chunks according to [1] .The match between the incoming chunk and the receiver chunk store takes place. If a matching is found, the local chunk sequence termed as a chunk chain is being identified by the receiver. This prediction consisting of the data in sequence is sent to the sender. If the resulted data yet to be sent matches the prediction, it continues to perform SHA-1operation.Then it confirms the match. If a successful match is found, the sender sends a confirmation message to the receiver. Finally enabling it to copy from the chunk store to the buffer.

To eliminate redundancy within individual objects, protocol-independent redundancy elimination techniques operated on individual packets [2]. The first sender based TRE was proposed in this paper. The amount of data to be transferred or stored is reduced by identifying both intra-object and inter-object redundant data element, which is replaced with a reference or pointer to the unique data copy. The objective of protocol-independent Data Redundancy Elimination is to encode the outgoing stream of data by identifying and replacing redundant data chunks with fixed-size Meta data which is done in our work.

A sender and receiver Cooperative end-to-end TRE solution (CoRE) for efficiently identifying and removing both short term and long term redundancy is proposed [3].This is done using a two-layer redundancy detection design and a single pass algorithm for chunking and fingerprinting [9] .The first layer TRE module detects long term redundancy by a prediction based Chunk Match approach like PACK [1]. A real life traffic redundancy is presented in [4].Redundant data is identified and eliminated using Redundancy suppression and data compression techniques [4].SHA-1 is a cryptographic hash function which outputs a 160-bit has value. The proposed design was captured using VHDL hardware language and also implemented on Xilinx FPGA [6].SHA-1 is implemented in the work to find the signature of a data. Hence a hashed value is obtained.

Backup application also exploits information redundancy to reduce the amount of information to be backed up [7]. Thus whatever processing takes place the data will be backed up in the cloud, which thus provides high availability [11]. Based upon the characteristics of the file a few works proposed to apply variable-size chunking and fixed size chunking. Based upon the metadata of individual files, Liu et al. proposed ADMAD scheme [10] which applies different chunking methods. Reference [15] presents a redundancy aware routing algorithm. These papers assume the routers are equipped with data caches and they search those routes that make a better use of the cached data.

To best of our knowledge none of the previous works have addressed the requirement for a secure [12], cost efficient end-to-end Traffic redundancy solved by Predictive Acknowledgement.

III. SECURE_PACK

In this paper, a security maintained receiver driven operation of Predictive acknowledgement protocol is described. The incoming stream of data received at receiver side is parsed to a sequence of variable-size, content based signed chunks. The comparison between the incoming chunk data and the previously arrived chunks which is present in the local storage termed as chunk store takes place. If a matching chunk is found in the chunk store, the receiver retrieves sequence of subsequent chunks referred to as a chain. This is done by traversing the sequence of LRU chunk pointers that are included in the chunk's metadata. Thus a chain of matched data will be constructed and using this chain the receiver sends a prediction to the sender for subsequent data. Part of each prediction termed as a hint, is an easy-to-compute function with a small enough false positive value. The prediction sent by the receiver which is present in the prediction queue includes the length of predicted data, the hint and the hashed value of the predicted data. The data owner identifies the length of the data and verifies the hint. If the result matches the received hint, it continues to perform SHA-1 [16] signature operation.

Depending on the signature match, data owner sends a confirmation message to the receiver, thus permitting it to copy the matched data from its chunk store. The system uses a new chunk chain scheme, in which chunks are linked to other chunks according to their last received order. To efficiently maintain and retrieve the stored chunks, their predicted chunk chain, caching and indexing techniques are used. Each chunk's signature is computed using SHA-1 when the new data are received and parsed to the chunks. At this point, the chunk is added to the chunk store. Thus the newly received chunk is placed after the previously received chunk in a least recently used manner. After the identification of the non redundant chunk, encryption of data takes place using triple DES algorithm.

A. Receiver Algorithm

When the new data arrives, respective signature for each chunk is computed by the receiver and then match is being looked out in its chunk store. If the chunk's signature is found, receiver finds out whether it is a part of previously received chunk chain using the chunk's metadata. Thus the receiver sends a prediction to the data owner indicating the next expected chunk chain. The prediction contains a starting point in the byte stream, the total length of the chunk and identity of subsequent chunks.

Fig 1 illustrates the receiver operation and describes how the data gets encrypted.

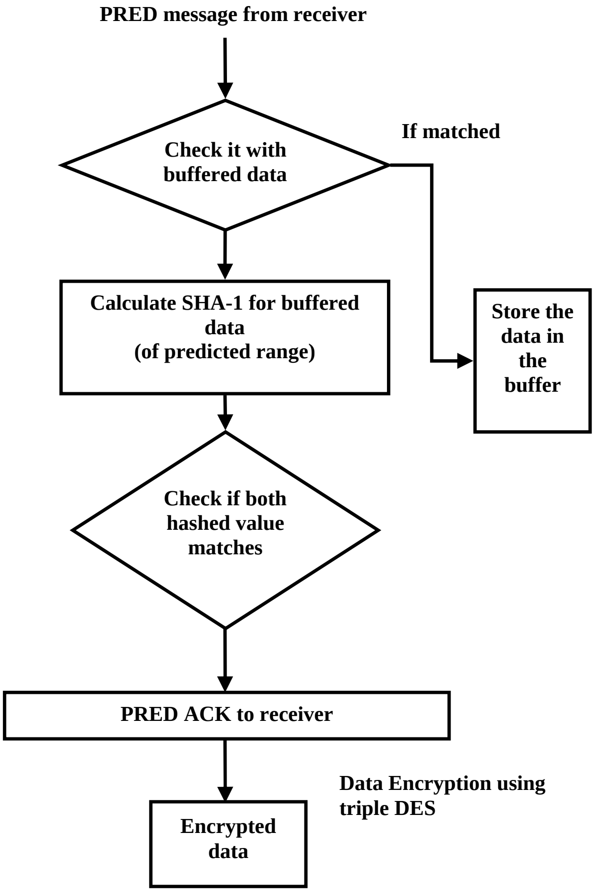
PRED message from receiver
Check it with buffered data
If matched
Calculate SHA-1 for buffered data (of predicted range)
Store the data in the buffer
Check if both hashed value matches
PRED ACK to receiver
Data Encryption using triple DES
Encrypted data

Fig 1 Receiver Algorithm

Upon a successful prediction, data owner sends a PRED-ACK confirmation message. Ones receiver confirms that the chunk is not redundant, the resulted data is encrypted using the triple DES algorithm and the encrypted data is sent to the cloud server by the receiver .Thus this improves the security level. The receiver copies the corresponding encrypted data from the chunk store to its cloud's buffer after the reception of the PRED-ACK message. If the data chunk is a redundant content, then the corresponding data ID is sent to the cloud. Thus traffic is avoided. At this point, receiver sends a TCP Acknowledgement with the next expected TCP sequence number.

B. Data Owner Algorithm

After the reception of PRED message from the receiver, it compares the prediction message with the data present in the data owner side. The data owner determines the range and verifies the hint for each prediction. Upon a hint match, the SHA-1 signature for the predicted data is being found out and the result is compared with the signature in the PRED message. If the both SHA-1 signature matches, data owner confirms that the receiver's prediction is correct. Thus the data owner sends a PRED-ACK message to the receiver. If suppose the hint does not match, then a computationally expansive operation is saved, which is thus used as a future predictor. The non redundant data is encrypted using triple DES algorithm and is sent to the cloud server.

Fig 2 illustrates the sender operation and describes how the sender tries to match a predicted range to its outgoing data.

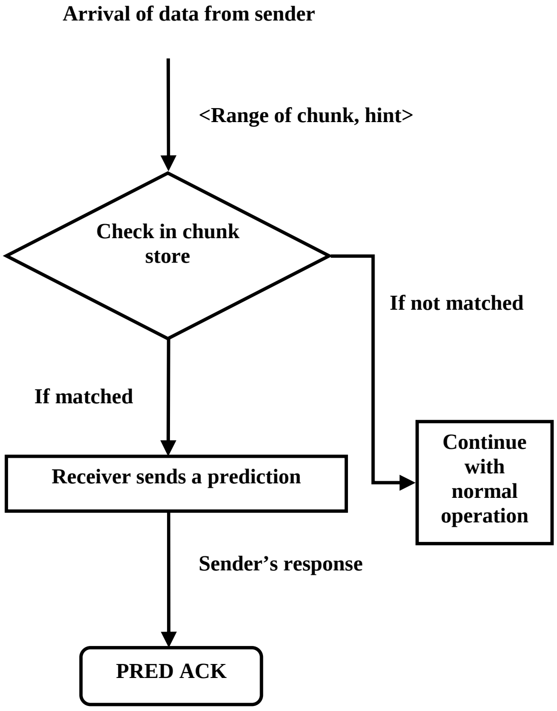

Fig 2 Data Owner Algorithm

C. Cloud Server Module

Cloud infrastructure is a "pay-as-you-go model". Data owner stores their applications on cloud. The delivery of computing service is over the internet. The data to be stored in the cloud is encrypted using the triple DES algorithm,which makes the data more secure. The procedure for encryption is exactly the same as regular DES ,except that it passes three times through the DES engine. The first pass is a DES encryption , the second pass is a DES decryption of the first DES Ciphertext result and the third pass is a DES encryption of the second pass result. This produces the resultant Triple DES Ciphertext. The encrypted data is finally placed in the cloud. Cloud server module can view the registered clients and their transaction details. Due to the existence of predictive acknowledgement, redundant traffic is eliminated. Thus non-redundant data is only being stored in the cloud. This reduces the cloud bandwidth , the overall operational cost and the security is also being maintained.

Encrypted Data System on the cloud

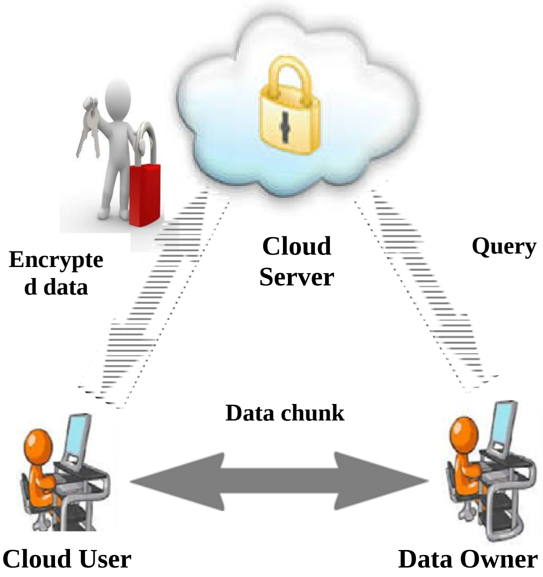

Fig. 3 Secure_PACK

Fig. 3 shows the system architecture of the Secure_PACK . After the non redundant data is being identified, the data is encrypted using the triple DES algorithm and is sent to the cloud server for storage. Since in cloud computing a distributed computing takes place, it is not secure to place the raw data in the cloud. Hence for maintaining security, data is encrypted using triple DES . Thus specific data owner can only view his data , which created privacy. In the previous work , even though bandwidth and cost were reduced , security level was not at all maintained. In our work the security level is maintained thus this overcomes the disadvantage of the existing system.

IV. Performance Evaluation

The section describes about the experimental evaluation. The SHA-1+Triple DES and SHA-1 are used to show the experimental results.

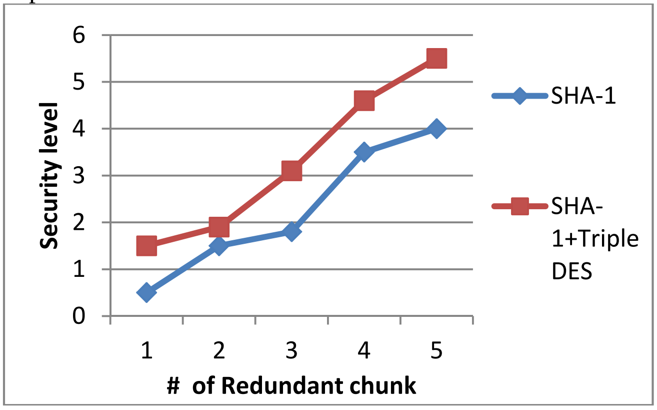

Fig. 3 Experimental Result.

The fig. 3 shows that security level increases even when the number of redundant chunk increases. The blue line shows the performance of the system when SHA-1 alone is applied. The red line shows the performance of the system when SHA-1+Triple DES is used. In the proposed system triple DES is being used for encrypting the data, hence security is maintained compared to the existing system. It clearly shows that the proposed system performs better than the previous methods in terms of security level.

V. Conclusion

In last year's for the shipment of large application files and rich data content, internet and intranet traffic has been evolved. Due to this shift evolution of proprietary, middle box based Traffic redundancy elimination scenarios came into existence which address the need of large corporations. Similar traffic characteristic trends continue to dominate the new generation mobile and wireless networks. Since most of the data redundancy exist at destination-to-destination exchange, a universal, standard, software based TRE is needed. Predictive Acknowledgement is a receiver based destination-to-destination TRE scheme which reduces the computation time and the cloud operational cost. Cloud elasticity and user mobility is being enabled since it doesn't require the server to continuously maintain client's status. Without applying a three-way handshake, traffic redundancy is eliminated. By maintaining the previously received chunk chain in a least recently used manner, the chunk chain is maintained in the receiver .Since the encrypted data is maintained in the cloud, thus this provides much more security to the previously existing system. Hence a secure, cost efficient and with reduced bandwidth cloud system will be obtained.

References

[1] Eyal Zohar, Israel Cidon, and Osnat Mokryn, "PACK: Prediction-Based Cloud Bandwidth and Cost Reduction System" *IEEE/ACM Transactions on Networking*, vol. 22, no. 1, February 2014,pp. 39-51.

[2] Yan Zhang and Nirwan Ansari, "On Protocol-Independent Data Redundancy Elimination", *IEEE Communications Surveys & Tutorials*, Vol. 16, no. 1, First Quarter 2014, pp. 455-472.

[3] Lei Yu, Karan Sapra, Haiying Shen and Lin Ye ,"Cooperative End-to-End Traffic Redundancy Elimination for Reducing Cloud BandwidthCost", *IEEE Communications Surveys & Tutorials*, Vol. 14, NO. 4, Fourth Quarter 2012.

[4] Yong Xu, Yin Liu, Yao Liu, "Algorithm for redundancy elimination in Network Traffic" , *2012 International Conference on Networking* ,pp. 1613-1617.

[5] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph,R. Katz, A. Konwinski , G. Lee, D. Patterson, A. Rabkin, I. Stoica, M.Zaharia,"A view of cloud computing", *Commun . ACM*, Vol.53, No. 4, pp. 50–58

[6] Anak Agung Putri Ratna, Ahmad Shaugi, Prima Dewi Purnamasari, "Analysis and comparison of MD5 and SHA-1 algorithm implementation in authentication based security system", *2013 International Conference on Computer Science and Electronics Engineering*, pp. 99-104

[7] Yingwu Zhu, Justin Masui, "Backing Up Your Data to the Cloud: Want to Pay Less?" *2013 42nd International Conference on Parallel Processing*,pp. 409-418.

[8] Suresh Chougala, Sharavana K., " Design of Traffic Redundancy and Elimination Approach for Reducing Cloud Bandwidth and Costs," *international journal of innovative research in science and engineering*, vol.2,Issue 2, pp.24-28, Feb 2014.

[9] B. Aggarwal, 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 Proc. of NSDI*, 2010.

[10] Chuanyi Liu, Yingping Lu, Chunhui Shi, Guanlin Lu, David H.C.Du, Dong-Sheng Wa ,"ADMAD: Application-Driven Metadata Aware Deduplication Archival Storage System", *Fifth IEEE International Workshop on Storage Network Architecture and Parallel I/Os* , pp. 29-35.

[11] J.Srinivasan,W.Wei, X.Ma, andt. Yu, "EMFS:Email-based personal cloud storage," *in Proc. NAS*, 2011, pp. 248–257.

[12] Sombir Singh, Sunil K. Maakar ,Dr.Sudesh Kumar., " Enhancing the Security of DES Algorithm Using Transposition Cryptography Techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol 3, issue 6, June 2013.

[13] A. Gupta, A. Akella, S. Seshan, S. Shenker, and J. Wang. "Understanding and Exploiting Network Traffic Redundancy". Technical Report 1592, UW-Madison, April 2007

[14] U. Manber. "Finding similar files in a large file system". *In Proc. of the USENIX winter technical conference*, pages 1–10, Berkeley, CA, USA, 1994. USENIX Association.

[15] Shu Yamamoto, Akihiro Nakao , "P2P Packet Cache Router for Network-Wide Traffic Redundancy Elimination", *International Conference on Computing, Networking and Communications, Network Architecture & P2P Protocol Symposium 2012*, pp. 830-834.

[16] Piyush Garg, Namita Tiwari. " Performance Analysis of SHA Algorithms," *International Journal of Computer Technology and Electronics Engineering* ,vol 2, issue 3, June 2012,pp. 130-132.

[17] A. Anand, C. Muthukrishnan, A. Akella, and R. Ramjee. "Redundancy in Network Traffic: Findings and Implications". *In Proc. of SIGMETRICS*, ACM New York, NY, USA, 2009, pp 37–48.