ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Novel TRE Framework Based Cloud Bandwidth and Cost Reduction System

Punuru Sindhu,
M.Tech(CSE)
Department of CSE,
Visvodaya Engineering College,
Kavali, Nellore (dist), India.

E.Dayakar, M.Tech
Associate Professor
Department of CSE,
Visvodaya Engineering College,
Kavali, Nellore (dist), India.

D.Srujan Chandra Reddy,
HoD
Department of CSE,
Visvodaya Engineering College,
Kavali, Nellore (dist), India.

ABSTRACT

Cloud is changing our life by giving clients new sorts of administrations. a novel end-to-end movement repetition disposal (TRE) framework, intended for distributed computing clients. Cloud-based TRE needs to apply a wise utilization of cloud assets so that the transfer speed expense decrease consolidated with the extra cost of TRE. Be that as it may, for server particular TRE methodology it's hard to handle the activity proficiently and it doesn't suites for the cloud setting because of high process costs. Amid this paper we give a study on the new movement excess procedure called novel-TRE conjointly called collector based TRE. This novel-TRE has vital choices like criminologist work the repetition at the client, arbitrarily pivoting seem anchored, matches approaching pieces with a prior got lump chain or local record and sending to the server for anticipating the long run data and no would like of server to persistently keep up buyer state.

Keywords— ***Caching, cloud computing, network optimization, traffic redundancy elimination.***

INTRODUCTION

Cloud computing is emerging style of delivery in which applications, data and resources are rapidly provisioned as standardized offerings to users with a flexible price. The cloud computing paradigm has achieved widespread adoption in recent years. Its success is due largely to customers' ability to use services on demand with a pay-as-you go [2] pricing model, which has proved convenient in many respects. Low costs and high flexibility make migrating to the cloud compelling. Cloud computing 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. By data outsourcing, users can be relieved from the burden of local data storage and maintenance. Traffic redundancy and elimination approach is used for minimizing the cost. Cloud applications that offer data management services are emerging. Such clouds support caching of data in order to provide quality query services. The users can query the cloud data, paying the price for the infrastructure they use.

Cloud management necessitates an economy that manages the service of multiple users in an efficient, but also, resource economic way that allows for cloud profit. Naturally, the maximization of cloud profit given some guarantees for user satisfaction presumes an appropriate price-demand model that enables optimal pricing of query services. The model should be plausible in that it reflects the correlation of cache structures involved in the queries. Optimal pricing is achieved based on a dynamic pricing scheme that adapts to time changes.

Protocol-independent redundancy elimination [3] works on Member to detect similar, but not necessarily identical, information transfers. In terms of improving Web performance, it has the potential to exceed the benefits of other approaches such as delta coding and duplicate suppression. This is because the similarity algorithms on which it is based include as a subset both exact matches (duplicate suppression) and differences between versions of the same document

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

(delta coding).A distinguishing feature of our system is that it is protocol independent. It makes no assumptions about the syntax or semantics of HTTP. This has two distinct advantages. It is able to identify fine grained sharing, as may be common with dynamically generated or personalized pages, as well as inter-protocol sharing. It does not need to be updated to take advantage of new types of content, such as streaming media, as they emerge or delivery protocols are revised

RELATED WORK

Many Redundancy Elimination techniques [3] have been explored in recent years. A protocol freelance Redundancy Elimination was planned in This paper was describes a sender packet-level Traffic Redundancy Elimination, utilization of the rule given in several industrial Redundancy Elimination answers that delineate in and have combined the sender based TRE concepts with the rule and implement approach of PACK and on with the protocol specific optimizations technique for middle box solution. In necessary have to be compelled to describe the way to escape with this tripartite hand shake between the sender half and additionally the receiver half if any full state synchronize is maintain.TRE system for the developing world wherever storage and WAN information measure are scarce. It's a application primarily based and connected middle-box replacement for the overpriced industrial hardware. During this kind, the sender middle-box holds back the TCP stream and sends data signatures to the receiver middle-box. The receiver verifies whether or not the info is found in its native cache. Information chunks that are not found in the cache are fetched from the sender middle-box or a near receiver middle-box. Naturally, such a theme incurs a three-way-handshake (3WH) latency for no cached information.

LITERATURE SURVEY

a) The power of prediction: Cloud bandwidth and cost reduction:

In this paper we present PACK (Predictive ACKs), a novel end-to-end Traffic Redundancy Elimination (TRE) system, designed for cloud computing customers. Cloud-based TRE needs to apply a judicious use of cloud resources so that the bandwidth cost reduction combined with the additional cost of TRE computation and storage would be optimized. PACK's main advantage is its capability of offloading the cloud-server TRE effort to end-clients, thus minimizing the processing costs induced by the TRE algorithm. Unlike previous solutions, PACK does not require the server to continuously maintain clients' status. This makes PACK very suitable for pervasive computation environments that combine client mobility and server migration to maintain cloud elasticity.

PACK is based on a novel TRE technique, which allows the client to use newly received chunks to identify previously received chunk chains, which in turn can be used as reliable predictors to future transmitted chunks. We present a fully functional PACK implementation, transparent to all TCP-based applications and network devices. Finally, we analyze PACK benefits for cloud users, using traffic traces from various sources.

b) Finding similar files in a large file system:

We present a tool, called sif, for finding all similar files in a large file system. Files are considered similar if they have significant number of common pieces, even if they are very different otherwise. For example, one file may be contained, possibly with some changes, in another file, or a file may be a reorganization of another file. The running time for finding all groups of similar files, even for as little as 25% similarity, is on the order of 500MB to 1GB an hour. The amount of similarity and several other customized parameters can be determined by the user at a post-processing stage, which is very fast. Sif can also be used to very quickly identify all similar files to a query file using a preprocessed index. Application of sif can be found in file management, information collecting (to remove duplicates), program reuse, file synchronization, data compression, and maybe even plagiarism detection

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

c) A protocol-independent technique for eliminating redundant network traffic:

We present a technique for identifying repetitive information transfers and use it to analyze the redundancy of network traffic. Our insight is that dynamic content, streaming media and other traffic that is not caught by today's Web caches is nonetheless likely to derive from similar information. We have therefore adapted similarity detection techniques to the problem of designing a system to eliminate redundant transfers. We identify repeated byte ranges between packets to avoid retransmitting the redundant data.We find a high level of redundancy and are able to detect repetition that Web proxy caches are not. In our traces, after Web proxy caching has been applied, an additional 39% of the original volume of Web traffic is found to be redundant. Moreover, because our technique makes no assumptions about HTTP protocol syntax or caching semantics, it provides immediate benefits for other types of content, such as streaming media, FTP traffic, news and mail.

d) A low-bandwidth network file system:

Users rarely consider running network file systems over slow or wide-area networks, as the performance would be unacceptable and the bandwidth consumption too high. Nonetheless, efficient remote file access would often be desirable over such networks---particularly when high latency makes remote login sessions unresponsive. Rather than run interactive programs such as editors remotely, users could run the programs locally and manipulate remote files through the file system. To do so, however, would require a network file system that consumes less bandwidth than most current file systems.This paper presents LBFS, a network file system designed for low-bandwidth networks. LBFS exploits similarities between files or versions of the same file to save bandwidth. It avoids sending data over the network when the same data can already be found in the server's file system or the client's cache. Using this technique in conjunction with conventional compression and caching, LBFS consumes over an order of magnitude less bandwidth than traditional network file systems on common workloads.

e) Method and apparatus for reducing network traffic over low bandwidth links:

A method is disclosed for reducing network traffic. At a sender, a data chunk is identified for transmission to a receiver, which is connected to the sender over a communication link. The sender computes a signature of the data chunk and determines whether the data chunk has been previously transmitted by looking up the signature in a sender index table. The sender index table associates the signatures of previously transmitted data chunks with unique index values. A message is transmitted to the receiver, where if the data chunk has previously been transmitted then the message includes an index value from the sender index table that is associated with the signature of the data chunk. At the receiver, the data chunk is located in a receiver cache that stores the previously transmitted data chunks by looking up the index value included in the message in a receiver index table. The receiver index table associates the unique index values with the locations in the receiver cache of the previously transmitted data chunks.

EXISTING SYSTEM

- Several commercial TRE solutions have combined the sender-based TRE ideas with the algorithmic and implementation approach along with protocol specific optimizations for middle-boxes solutions.
- In particular, how to get away with three-way handshake between the sender and the receiver if a full state synchronization is maintained.
- Traffic redundancy stems from common end-users' activities, such as repeatedly accessing, downloading, uploading (i.e., backup), distributing, and modifying the same or similar information items (documents, data, Web, and video).
- TRE is used to eliminate the transmission of redundant content and, therefore, to significantly reduce the network cost.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

- In most common TRE solutions, both the sender and the receiver examine and compare signatures of data chunks, parsed according to the data content, prior to their transmission.
- Current end-to-end solutions also suffer from the requirement to maintain end-to-end synchronization that may result in degraded TRE efficiency.
- When redundant chunks are detected, the sender replaces the transmission of each redundant chunk with its strong signature.
- Cloud providers cannot benefit from a technology whose goal is to reduce customer bandwidth bills, and thus are not likely to invest in one.

- Commercial TRE solutions are popular at enterprise networks, and involve the deployment of two or more proprietary-protocol, state synchronized middle-boxes at both the intranet entry points of data centers.

- The rise of "on-demand" work spaces, meeting rooms, and work-from-home solutions detaches the workers from their offices. In such a dynamic work environment, fixed-point solutions that require a client-side and a server-side middle-box pair become ineffective.

- Disadvantages: Cloud load balancing and power optimizations may lead to a server-side process and data migration environment, in which TRE solutions that require full synchronization between the server and the client are hard to accomplish or may lose efficiency due to lost synchronization.

PROPOSED SYSTEM

- PACK's main advantage is its capability of offloading the cloud-server TRE effort to end-clients, thus minimizing the processing costs induced by the TRE algorithm.
- Unlike previous solutions, PACK does not require the server to continuously maintain clients' status. This makes PACK very suitable for pervasive computation environments that combine client mobility and server migration to maintain cloud elasticity.
- PACK is based on a novel TRE technique, which allows the client to use newly received chunks to identify previously received chunk chains, which in turn can be used as reliable predictors to future transmitted chunks.

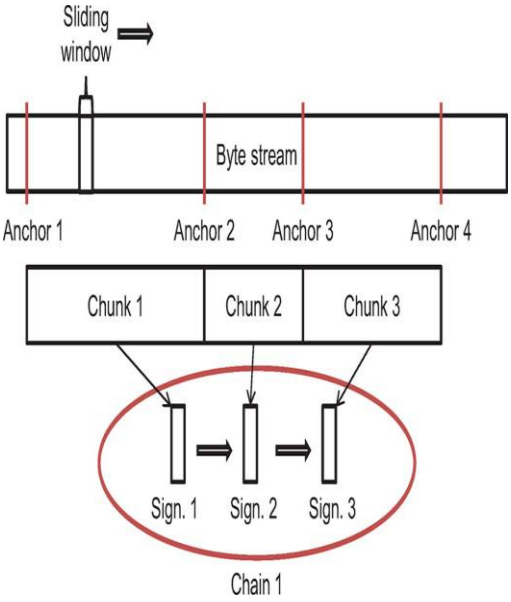

Fig. . From stream to chain.

- In this paper present a fully functional PACK implementation, transparent to all TCP based applications and network devices.
- In this paper, we present a novel receiver-based end-to-end TRE solution that relies on the power of predictions to eliminate redundant traffic between the cloud and its end-users. In this solution, each receiver observes the incoming stream and tries to match its chunks with a previously received chunk chain or a chunk chain of a local file. Using the long-term chunks' metadata information kept locally, the receiver sends to the server predictions that include chunks' signatures and easy-to-verify hints of the sender's future data. On the receiver side, we propose a new computationally lightweight

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

chunking (fingerprinting) scheme termed PACK chunking. PACK chunking is a new alternative for Rabin fingerprinting traditionally used by RE applications.

- Advantages: This solution achieves 30% redundancy elimination without significantly affecting the computational effort of the sender, resulting in a 20% reduction of the overall cost to the cloud customer.

IMPLEMENTATION

a) Receiver Chunk Store:

PACK uses a new chains scheme. Which chunks are linked to other chunks according to their last received order. The PACK receiver maintains a chunk store, which is a large size cache of chunks and their associated metadata. Chunk's metadata includes the chunk's signature and a (single) pointer to the successive chunk in the last received stream containing this chunk. Caching and indexing techniques are employed to efficiently maintain and retrieve the stored chunks, their signatures, and the chains formed by traversing the chunk pointers.

When the new data are received and parsed to chunks, the receiver computes each chunk's signature using SHA-1. At this point, the chunk and its signature are added to the chunk store. In addition, the metadata of the previously received chunk in the same stream is updated to point to the current chunk. The unsynchronized nature of PACK allows the receiver to map each existing file in the local file system to a chain of chunks, saving in the chunk store only the metadata associated with the chunks.

b) Receiver Algorithm:

Upon the arrival of new data, the receiver computes the respective signature for each chunk and looks for a match in its local chunk store. If the chunk's signature is found, the receiver determines whether it is a part of a formerly received chain, using the chunks' metadata. If affirmative, the receiver sends a prediction to the sender for several next expected chain chunks. Upon a successful prediction, the sender responds with a PRED-ACK confirmation message. Once the PRED-ACK message is received and processed, the receiver copies the corresponding data from the chunk store to its TCP input buffers, placing it according to the corresponding sequence numbers. At this point, the receiver sends a normal TCP ACK with the next expected TCP sequence number. In case the prediction is false, or one or more predicted chunks are already sent, the sender continues with normal operation, e.g., sending the raw data, without sending a PRED-ACK message.

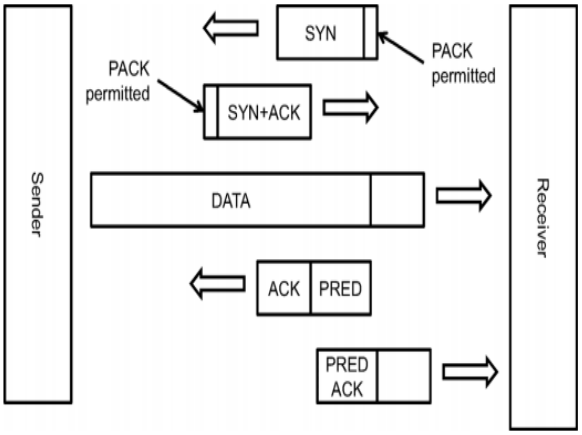

Fig: PACK wire protocol in a nutshell

c) Sender Algorithm

When a sender receives a PRED message from the receiver, it tries to match the received predictions to its buffered (yet to be sent) data. For each prediction, the sender determines the corresponding TCP sequence range and verifies the hint. Upon a hint match, the sender calculates the more computationally intensive SHA-1 signature for the predicted data range and compares the result to the signature received in the PRED message. Note that in case the hint does not match, a computationally expansive operation is saved. If the two SHA-1 signatures match, the sender can safely assume that the receiver's prediction is correct. In this case, it replaces the corresponding outgoing buffered data with a PRED-ACK message.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

d) Wire Protocol

The existing firewalls and minimizes overheads; we use the TCP Options field to carry the PACK wire protocol. It is clear that PACK can also be implemented above the TCP level while using similar message types and control fields. The PACK wire protocol operates under the assumption that the data is redundant. First, both sides enable the PACK option during the initial TCP handshake by adding a PACK permitted to the TCP Options field. Then, the sender sends the (redundant) data in one or more TCP segments, and the receiver identifies that a currently received chunk is identical to a chunk in its chunk store. The receiver, in turn, triggers a TCP ACK message and includes the prediction in the packet's Options field. Last, the sender sends a confirmation message (PRED-ACK) replacing the actual data.

CONCLUSION

In this paper, we proposed a novel-TRE approach for eliminating redundancy in the cloud environment. Our proposed scheme has significant features like reduces the transmission cost by predicting chunks, redundancy detection by the client, does not require the server to continuously maintain clients' status. Our receiver based end-to-end TRE suites for cloud environment. Distributed computing offers its clients a conservative and helpful pay-as-you-go administration model, referred to likewise as use based evaluating. Cloud clients pay just for the genuine utilization of processing assets, stockpiling, and transfer speed, as per their evolving needs, using the cloud's adaptable and versatile computational capacities. Specifically, information exchange costs (i.e., data transfer capacity) are a paramount issue when attempting to minimize costs. Cloud-based TRE needs to apply an astute use of cloud resources with the goal that the exchange rate cost decline united with the additional expense of TRE estimation and limit would be overhauled. PACK's essential purpose of investment is its ability of offloading the cloud-server TRE effort to end clients, in this way minimizing the taking care of costs incited by the TRE.

REFERENCES

[1] E. Zohar, I. Cidon, and O. Mokryn, "The power of prediction: Cloud bandwidth and cost reduction," in Proc. SIGCOMM, 2011, pp. 86–97.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph,R.Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Commun. ACM, vol. 53, no. 4, pp. 50–58, 2010.

[3] U. Manber, "Finding similar files in a large file system," in Proc. USENIX Winter Tech. Conf., 1994, pp. 1–10. [4] N. T. Spring and D. Wetherall, "A protocol-independent technique for eliminating redundant network traffic," in Proc. SIGCOMM, 2000, vol. 30, pp. 87–95.

[5] A. Muthitacharoen, B. Chen, and D. Mazières, "A low-bandwidth network file system," in Proc. SOSP, 2001, pp. 174–187.

[6] E. Lev-Ran, I. Cidon, and I. Z. Ben-Shaul, "Method and apparatus for reducing network traffic over low bandwidth links," US Patent 7636767, Nov. 2009.

[7] S.Mccanne andM. Demmer, "Content-based segmentation scheme for data compression in storage and transmission including hierarchical segment representation," US Patent 6828925, Dec. 2004.

[8] R. Williams, "Method for partitioning a block of data into subblocks and for storing and communicating such subblocks," US Patent 5990810, Nov. 1999.

[9] Juniper Networks, Sunnyvale, CA, USA, "Application acceleration," 1996 [Online]. Available: http://www.juniper.net/us/ en/products-services/application-acceleration/

[10] Blue Coat Systems, Sunnyvale, CA, USA, "MACH5," 1996 [Online]. Available: http://www.bluecoat.com/products/mach5

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[11] Expand Networks, Riverbed Technology, San Francisco, CA, USA, "Application acceleration and WAN optimization," 1998 [Online]. Available: http://www.expand.com/technology/ application-acceleration.aspx

[12] F5, Seattle,WA, USA, "WAN optimization ," 1996 [Online]. Available: http://www.f5.com/solutions/acceleration/wan-optimization/

[13] A. Flint, "The next workplace revolution," Nov. 2012 [Online]. Available: http://m.theatlanticcities.com/jobs-and-economy/2012/11/ nextworkplace-revolution/3904/

[14] A. Anand, C. Muthukrishnan, A. Akella, and R. Ramjee, "Redundancy in network traffic: Findings and implications," in Proc. SIGMETRICS, 2009, pp. 37–48.

[15] B. Aggarw al, 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. NSDI, 2010, pp. 28–28.

[16] "PACK source code," 2011 [Online]. Available: http://www.eyalzo.com/projects/pack

[17] A. Anand, A. Gupta, A. Akella, S. Seshan, and S. Shenker, "Packet caches on routers: The implications of universal redundant traffic elimination," in Proc. SIGCOMM, 2008, pp. 219–230.

[18] A. Anand, V. Sekar, and A. Akella, "SmartRE: An architecture for coordinated network-wide redundancy elimination," in Proc. SIGCOMM, 2009, vol. 39, pp. 87–98.

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

[20] M. Zink, K. Suh, Y. Gu, and J. Kurose, "Watch global, cache local: YouTube network traffic at a campus network—Measurements and implications," in Proc. MMCN, 2008, pp. 1–13.

Author details:

Punuru Sindhu, M.Tech in Computer Science and Enginnering, Visvodaya Engineering College, Kavali, Nellore (dist), India.

E.Dayakar, M.Tech, Associate Professor in Computer Science and Enginnering, Visvodaya Engineering College, Kavali, Nellore (dist), India.

D.Srujan Chandra Reddy, HoD of Computer Science and Enginnering, Visvodaya Engineering College, Kavali, Nellore (dist), India.