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

A Traffic Redundancy Elimination System for Improving the Cloud Elasticity by Estimating Bandwidths

G.Yashwanth
M.Tech Student
Department of CSE
B V Raju Institute of Technology
Narsapur, India.

T.Bharath Kumar
Assistant Professor
Department of CSE
B V Raju Institute of Technology
Narsapur, India.

Abstract:

Cloud computing, also known as on-demand computing, is a kind of internet-based computing, where shared resources and information are provided to computers and other devices on-demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in third-party data centers. It relies on sharing of resources to achieve coherence and economies of scale, similar to a utility (like the electricity grid) over a network. At the foundation of cloud computing is the broader concept of converged infrastructure and shared services. In cloud computing, computing resources square measure provided as services over the web and users will access resources on supported their payments. But for server specific TRE approach it's tough to handle the traffic efficiently and it doesn't suites for the cloud setting due to high process prices. During this paper we provide a survey on the new traffic redundancy technique called novel-TRE conjointly called receiver based TRE. This novel-TRE has important options like detective work the redundancy at the customer, randomly rotating appear chained, matches incoming chunks with a antecedently received chunk chain or native file and sending to the server for predicting the long run information and no would like of server to unceasingly maintain consumer state.

Keywords: ***Cloud Computing, chunking, TRE, novel-TRE, computing paradigm.***

Introduction:

Cloud computing, or in simpler shorthand just "the cloud", also focuses on maximizing the effectiveness of the shared resources. Cloud resources are usually not only shared by multiple users but are also dynamically reallocated per demand. This can work for allocating resources to users. For example, a cloud computer facility that serves European users during European business hours with a specific application (e.g., email) may reallocate the same resources to serve North American users during North America's business hours with a different application (e.g., a web server). This approach helps maximize the use of computing power while reducing the overall cost of resources by using less power, air conditioning, rack space, etc. to maintain the system. With cloud computing, multiple users can access a single server to retrieve and update their data without purchasing licenses for different applications.

Cloud computing is growing type of delivery during which applications, information and resources are speedily provisioned as standardized offerings to users with a versatile worth. The cloud computing paradigm has achieved widespread adoption in recent years. Its success is due mostly to customers' ability to use services on demand with a pay-as-you go evaluation model, which has proven convenient in several respects. Low prices and high flexibility create migrating to the cloud compelling. Cloud computing is that the long unreal vision of computing as a utility, wherever users will remotely store their information into the cloud thus on get pleasure from the on demand top quality applications and services from a shared

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

pool of configurable computing resources. By information outsourcing, users are all aviated from the burden of native information storage and maintenance. Traffic redundancy and elimination approach is employed for minimizing the value.

PACK (traffic redundancy elimination):

Cloud customers utilize the pliability of pay-per-use Bandwidth value reduction(judicious use of cloud resources) cut back the extra value of TRE computation and storage industrial TRE solutions are fashionable at enterprise networks(not in cloud environments) enterprise networks:-middle boxes at the entry purpose of information centers and branch offices cloud settings:-cloud supplier cannot get benefit(reduce bills) mounted client-side and serverside TRE is inefficient for a mixture of mobile environment Sender primarily based end-to-end TRE Load equalization, power optimization, information migration => hefty masses to server Receiver primarily based end-to-end TRE

Related Work:

Finding Similar Files In A Large File System

This paper presents 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.

A Low-bandwidth Network File System

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 voids sending data over the network when the same data can already be found in the server's file system or the client's cache.

Redundancy in Network Traffic: Findings and Implications

A large amount of popular content is transferred repeatedly across network links in the Internet. In recent years, protocol-independent redundancy elimination, which can remove duplicate strings from within arbitrary network flows, has emerged as a powerful technique to improve the efficiency of network links in the face of repeated data. Many vendors offer such redundancy elimination middle boxes to improve the effective bandwidth of enterprise, data center and ISP links alike. This paper conduct a large scale trace-driven study of protocol independent redundancy elimination mechanisms, driven by several terabytes of packet payload traces collected at 12 distinct network locations, including the access link of a large US-based university and of 11 enterprise networks of different sizes. Based on extensive analysis, it presents a number of findings on the benefits and fundamental design issues in redundancy elimination systems.

Packet Caches on Routers: The Implications of Universal Redundant Traffic Elimination

Many past systems have explored how to eliminate redundant transfers from network links and improve network efficiency. Several of these systems operate at the application layer, while the more recent systems operate on individual packets. A common aspect of these systems is that they apply to localized settings, e.g. at stub network access links. This paper explores the benefits of deploying packet-level redundant content elimination as a universal primitive on all Internet routers. Such a universal deployment would immediately reduce link loads everywhere. However, we argue that far more significant network-wide benefits can be derived by redesigning network routing protocols to leverage the universal deployment.

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.
- 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:

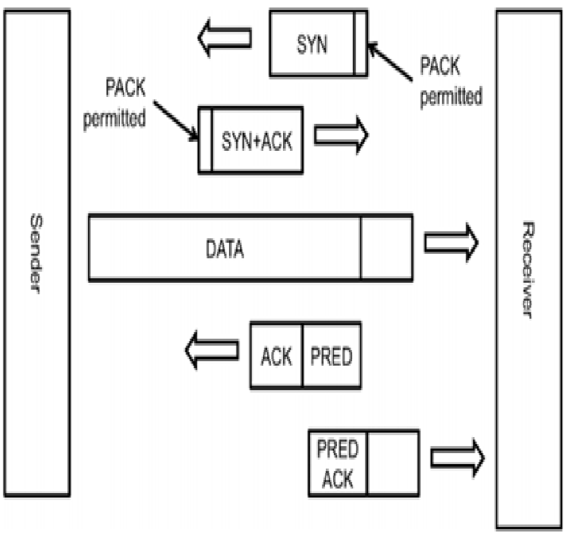
- PACK's main advantage is its capability of offloading the cloud-server TRE effort to end-clients, thus minimizing the processing costs induced by the TRE algorithm.
- Unlike previous solutions, PACK does not require the server to continuously maintain clients' status. This makes PACK very suitable 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.
- 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 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.

System Architecture:



Modules:

1. User Connect with Bandwidth and Storage Cost to Cloud
2. Upload File
3. Cloud Segment Processing
4. Read & Download File
5. Delete File from Cloud

User Connect with Bandwidth and Storage Cost to Cloud

- The Data owner wants to connect with cloud server and upload his file, then access this file anywhere, anytime.
- So, he connect with bandwidth and storage cost to cloud server. Bandwidth and Storage Cost are used for upload his file and access this file.
- Depending on the upload files, his bandwidth and storage cost are reduced.

Upload File:

- In computer networks, to upload can refer to the sending of data from a end user to a remote system such as a cloud server or another client with the intent that the remote system should store a copy of the data being transferred, or the initiation of such a process.
- Where the connection to the remote computers is via a dial-up connection, the transfer time required to download locally and then upload again could increase from seconds to hours or days.
- Here, the user uploads his file in cloud. The cloud maintains a chunk store. It store the chunks and its signature.

Cloud Segment Processing:

- Upon the arrival of new data, the cloud 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 cloud determines whether it is a part of a formerly

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

received chain, using the chunks' metadata. If affirmative, the cloud sends a prediction to the sender for several next expected chain chunks.

- The prediction carries a starting point in the byte stream (i.e., offset) and the identity of several subsequent chunks (PRED command).
- Upon a successful prediction, the user 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.

Read & Download File:

- First, the user sent a filename as request to cloud. Then the server will match this filename and get this file.
- The Server encrypts this file using user's public key. Then it sends a cipher text to user.
- Finally user will decrypt this cipher text and get the original file.
- Then download this file.

Delete File from Cloud:

- The user wants to delete his file. So he sent the filename as request to cloud. Then the server will match this filename and get this file.
- Then it will delete this file and its chunks.
- The server calculates this deleted files bandwidth and cost. Then add his bandwidth and cost.

Conclusion:

Cloud computing is expected to trigger high demand for TRE solutions as the amount of data exchanged between the cloud and its users is expected to dramatically increase. The cloud environment redefines the TRE system requirements, making proprietary middle -box solutions inadequate. Consequently, there is a rising need for a TRE solution that reduces the cloud's op-erational cost while accounting for application latencies, user mobility, and cloud elasticity. In this paper, we have presented PACK, a receiver-based, cloudfriendly, end - to-end TRE that is based on novel specula-tive principles that reduce latency and cloud operational cost. PACK does not require the server to continuously maintain clients' status, thus enabling cloud elasticity and user mobility while preserving long -term redundancy

References:

[1] E. Zohar, I. Cidon, and O. Mokryn, "PACK: Prediction-Based Cloud Bandwidth and Cost Reduction System. IEEE/ACM Transactions on Networking, 01 February 2013.

[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] 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.

[4] B. Aggarwal, A. Akella, A. Anand, A. Balachandran, P. Chitnis, C. Muthukrishnan, R. Ramjee, and G. Varghese, "EndRE: An endsystem redundancy elimination service for enterprises," in Proc. NSDI, 2010,pp. 28–28.

[5] Ben-Shaul, "Method and apparatus for reducing network traffic over low bandwidth links," US Patent 7636767, Nov. 2009.

[6] Ihm, K. Park, and V. Pai, "Wide-area network acceleration for the developing world," in Proc. USENIX ATC, 2010, pp. 18–18.

[7] U. Manber,"Finding similar files in a large file system", in Proc. USENIX Winter Tech. Conf., 1994, pp. 1–10.

[8] N. T. Spring, D. Wetherall,"A protocol-independent technique for eliminating redundant network traffic", In Proc. SIGCOMM, 2000, Vol. 30, pp. 87–95.

[9] A. Muthitacharoen, B. Chen, D. Mazières,"A low-bandwidth net-work file system", In Proc. SOSP, 2001, pp. 174–187.

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

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

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