Prediction-Based Traffic Redundancy Elimination

MADHAVI.M
M.Tech Student, Deptof CSE
Siddartha Educational Academy Group Of Institutions
C.Gollapalli,Tirupathi

P.RAJASEKHAR, M.TECH
Associate Professor,Dept of CSE
Siddartha Educational Academy Group Of Institutions
C.Gollapalli,Tirupathi

Abstract:
Cloud computing is an emerging technology trend, which provides an economical pay-for-use service model. Cloud customers pay only for the actual use of computing resources, storage, and bandwidth, which changes continuously according to ones requirements. In particular, a data-transfer charge like bandwidth utilization is one of the important issues to minimize the cloud costs for an end-user. By reducing the bandwidth utilization, it is possible to reduce the cloud cost for end-users. In order to save the bandwidth, it is needed to eliminate the traffic redundancy over the network. In this paper a hybrid sender-receiver approach is proposed to provide solutions for above problems by eliminating traffic redundancy. In this system, prevention of multiple uploads or downloads of a single file can be validated, which saves the bandwidth utilization to reduce the cloud cost. The system also provides an authorized accessing of the content, in which the data owner needs to provide permission for any user, who requests to download the same.

Keywords: *Bandwidth, Elimination, Prevention, Trafic redundancy*

1. Introduction

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, there-fore, 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. When redundant chunks are detected, the sender replaces the transmission of each redundant chunk with its strong signature [3–5]. 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 and branch offices, eliminating repetitive traffic between them (e.g., Cisco [6], Riverbed [7], Quantum [8], Juniper [9], Blue Coat [10], Expand Networks [11], and F5 [12]).

In the proposing system a hybrid sender-receiver approach is proposed to provide solutions for above problems by eliminating traffic redundancy. In this system, prevention of multiple uploads or downloads of a single file can be validated, which saves the bandwidth utilization to reduce the cloud cost. The system also provides an authorized accessing of the content, in which the data owner needs to provide permission for any user, who requests to download the same.

2. Existing System

In the existing system the receiver based end-to-end TRE solution is used to eliminate the traffic redundancy by using predictions for future received chunks. 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' meta-data 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. The sender first examines the hint and performs the TRE operation only on a hint- match. The purpose of this procedure is to avoid the expensive TRE computation at the sender side in the absence of traffic redundancy. When redundancy is detected, the sender then sends to the receiver only the ACKs to the predictions, instead of sending the data.

3. Disadvantages of Existing System

- It does not provide authenticated users only to access the file and thus many users can access it, which raises the privacy issue and also load will be increased.
- The experimental results show that only a particular amount of traffic redundancy only eliminated by using this approach.

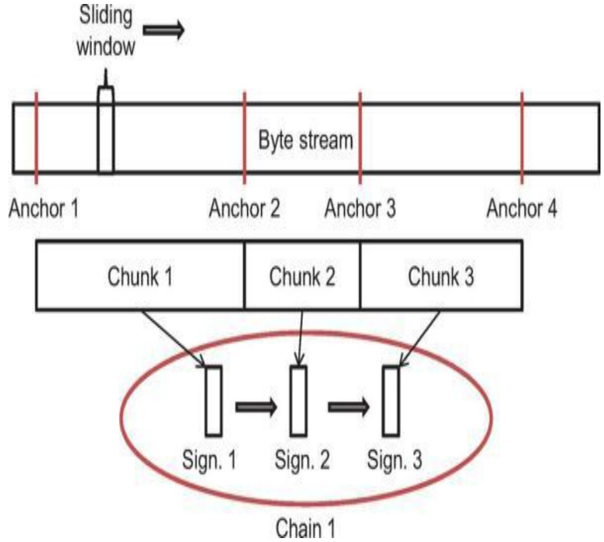
4. Proposed System

In the proposing system a hybrid sender-receiver approach is proposed to provide solutions for above problems by eliminating traffic redundancy. In this system, prevention of multiple uploads or downloads of a single file can be validated, which saves the bandwidth utilization to reduce the cloud cost. The system also provides an authorized accessing of the content, in which the data owner needs to provide permission for any user, who requests to download the same.

5. Advantages of Proposed System

- Provides authorization to the data user , which provides privacy and also load will be decreased.
- Eliminates multiple uploads and downloads of a single file which saves the bandwidth utilization to reduce the cloud cost

6. System Architecture



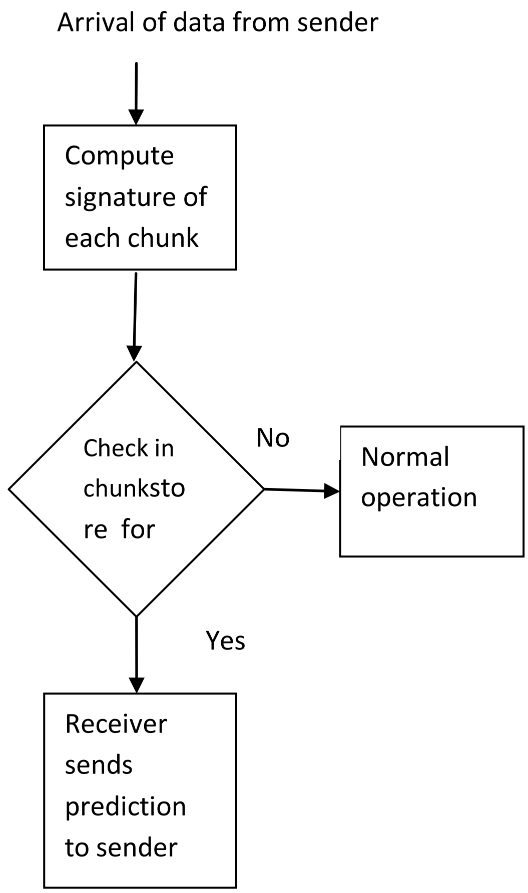
The stream of data received at the PACK receiver is parsed to a sequence of variable-size, content-based signed chunks similar to [3] and [5]. The chunks are then compared to the receiver local storage, termed *chunk store*. If a matching chunk is found in the local chunk store, the receiver retrieves the sequence ofsubsequent chunks, referred to as a chain, by traversing the sequence of LRU chunk pointers that are included in the chunks metadata.

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

8. Receiver Chunk Store

The receiver maintains a *chunk store*, which is a large size cache of chunks and their associated metadata.

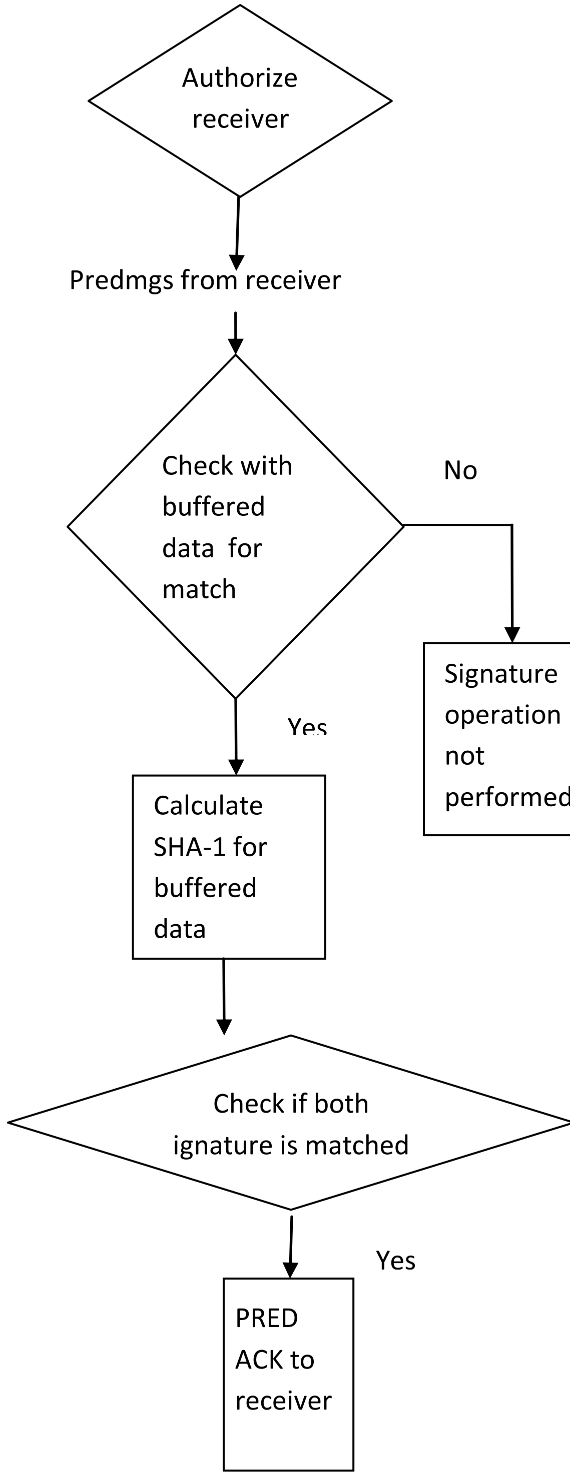


Receiver Algorithm

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.

9. Sender Algorithm.

only get predictive acknowledgement from the sender, thus increasing privacy .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.



Sender Algorithm

10. 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 operational cost while accounting for application latencies, user mobility, and cloud elasticity.

References

1. Armbrust, M. A. Fox, R. Griffith, A. D. Joseph,R.Katz, A. Konwinski, G.Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia (2010). A view of cloud computing, Commun.ACM, vol. 53, no. 4, pp. 50–58.
2. Lev-Ran, E. I. Cidon, and I. Z. Ben-Shaul (2009). Method and apparatus for reducing network traffic over low bandwidth links, US Patent 7636767, Nov.
3. Manber, U. (1994). Finding similar files in a large file system,‖ in Proc. USENIX Winter Tech. Conf., pp. 1–10.
4. Mccanneand, S. M. Demmer (2004). Content-based segmentation scheme for data compression in storage and transmission including hierarchical segment representation, US Patent 6828925, Dec.
5. Muthitacharoen, A. B. Chen, and D. MaziEres (2001). A low-bandwidth network files system, in Proc. SOSP, 2001, pp. 174–187.
6. Spring , N. T. and D. (2000). Wetherall, A protocol-independent technique for eliminating redundant network traffic, in Proc. SIGCOMM, 2000, vol.30, pp. 87–95.
7. Zohar, E. I. Cidon, and O. Mokryn (2011). The power of prediction: Cloud bandwidth and cost reduction, in Proc. SIGCOMM, pp. 86–97.