International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 19 (2017) pp. 8436-8442

An Enhanced Search Optimization Proposal Based on Gossip Protocol for the Cloud

Sudipta Sahana
Assistant Professor, Department of Computer Science and Engineering
JIS College of Engineering, Kalyani, Nadia, West Bengal, India.
Orcid Id: 0000-0002-9694-6399

Rajesh Bose
Research Scholar, Department of Computer Science and Engineering
University of Kalyani, Kalyani, Nadia, West Bengal, India.

Debabrata Sarddar
Assistant Professor, Department of Computer Science and Engineering
University of Kalyani, Kalyani, Nadia, West Bengal, India.

Abstract

The world that we inhabit today is vastly dependent on digital data than it ever was before. The need to process and store digital information is of paramount interest. This, therefore, has given rise to techniques and processes that can considerably reduce access times of data stored. It is important to understand though, that despite recent advancements made in storage hardware, the vast quantities of data that is generated every minute cannot be stored in single or even small clusters of devices. The rise of Cloud Computing offers a unique and generally efficient means to harbor the humongous quantities of data. This, however, is not without a price. Although, storing data on Cloud is accepted to be secure and efficient, the associated cost of storage is quite high in comparison to traditional storage methods. In this paper, we examine a technique that can reduce the cost of storing data in Cloud using Gossip protocol. The Gossip protocol has been in existence for quite sometime and is used to distribute and update information across wireless networks and large-scale database systems. We posit that using our method, users will be able to reduce time required to search for data. We demonstrate how our proposed model is more efficient in extracting information through a novel routing process. Our results also demonstrate that our proposed system exhibits efficiency and saves a lot more energy. The objective of a green cloud computing infrastructure can be achieved using our proposed model.

Keywords: Cloud Storage, Efficient routing, Search Optimization Data Center, Energy consumption.

INTRODUCTION

Cloud computing is a fruitful addition to the modern IT industry that deals with a considerable amount of data on daily basis. The various kind of services provided by cloud computing includes On-demand self-service, location sovereign basis assembling, widespread system access, pay per practice and conveyance of hazard. The cost of managing data constitutes a significant portion of existing cost structure in any business firm. By providing an efficient data management with lower cost cloud can give a way to higher profit. Any personal or business data can be stored in cloud in a consolidated manner and made accessible to the end users on demand. As a result the complexities of supervising resources, accessing data from physical locations are reduced to a great extent. Also the cost of maintaining hardware, software and management staffs for handling resources are replaced by a single cost of payment to the cloud authorities for accessing those data. Not only storing and accessing data, in cloud technology attention has also given to the safety or security aspect of data outsourced by the users.

Consistent flow of data through the used server is one of the key features of cloud. The application of cloud first started in e-commerce business with gradually increasing its applicability for the use of other business enterprises, individual as well as various other organizations. With increasing scope of uses the obvious issue that becomes relevant is the protection of the cloud storage system along with proper management of enormous data. Tempo, reliability, eases of access, sensitivity etc. are some of the challenges that coming into forth. Since the amount of data has increased more rapidly than the development of cloud technology, difficulties are faced in front of combating all the problems in storing data efficiently.

Gossip protocols are designed to exchange information between the parties involved. The nodes engaged in each step exchange only a small amount of iterative information. The nodes to be used for the task are chosen randomly and the

system tends to satisfy some global property with some probability. Several others protocol prior to formal introduction of gossip used gossip like architecture. For example, internet routing protocol. Here the process of data interaction is same as like that in gossip structure. A routed network can be formed through the system of point to point communication between the nodes in order to pass the information smoothly through the gossip layers.

High scalability and regularity are the key features of gossip protocol. In addition it runs in a symmetric and decentralized mode which make it advantageous for large scale use and particularly in the wireless networking system. In gossip as various nodes are connected with each other maintaining a stable or consistent connection among the nodes is a primary necessity. Though only a small amount of information is exchanged through each node but eventually all the information is spread over all the nodes connected in the system. Thus a global networking is formed among the nodes from the point to point partial communication system.

Nowadays, gossip is becoming a well-known communication technique in the large scale distributed architecture. Its uses show a rising trend due to its fairly stable nature even with varying environment component. Apart from partial exchange of data between any pairs of nodes, all the information must pass through the entire system of nodes making those comparable to gossip users.

In the rest of the paper we discuss on the following aspects of gossiping. In the section 2 we consider some of the important working using gossip protocol. Section 3 focuses on the methodology to be used for proposed architecture using algorithm. Section 4 deals with the execution of the discussed methodology and discusses the results obtained from it. Finally in section 4 the summary conclusion is presented with mentioning the further scope of research in this field.

RELATED WORKS

Several interesting research work is already done using gossip protocol. Here we present a brief literature survey considering some of the methodology of the previously done papers that are relevant for our current purpose. There is a possibility of losing data or information in case of interruption in network connectivity or unstable power supply. Once this happens it becomes very difficult to retrieve the information again unless a secured backup has already taken.

Analyzing the threat of data loss in Cloud Environment

Cloud system storage provides regaining of data in case of data lost for some reason or other in an efficient manner. In (F. Machida et al.2011) a framework of the working model is presented where the data are accessible from the data center where the data are disposed in the cloud service. The flexibility of data usages are major concern for the authors in (B. Addis et al.2010). A systematic program based administration of cloud resources are the main objective of the concerned research work. A proposal to extend the reference model in cloud computing is given in (A. Widjajarto et al.2010). The expansion of the reference model is expected to optimize the resource use by eliminating the idle ones in cloud hierarchy. Continuous pressure on a system results in inefficiency of the system and also lagging down the devices connected with it. Cloud computing is no exception in this regard. So, balancing load is an important aspect for the system to be persistent and retaining efficiency for a longer time horizon. This issue is taken care of in (Sarddar Debabrataet al.2015). The authors, Debabrata et al. suggests the combination of services leads to an efficient routing system by offloading the pressure and controls the folding of data. Paper (Demers et al.1987) explains the rationale behind using gossip protocol in computer network following gossip used in communication networks. The use of gossip in large scale wireless network makes the system efficient that would otherwise be. This is either because of the system is inconveniently large or simply because the gossip is most compatible given the architecture of the system. The optimization of storage space and lesser energy consumption is the focus of the paper (Rajesh Bose et al.2015). Here the client can store their information in the desired location for the ease of accessibility. Less power is consumed since only a part of the entire storage is searched to fetch any formation rather than scanning the entire system. The authors of paper (Sudipta Sahana et al.2016) suggest a new technique for storage optimization. The new technique is one that is based on the categorization according to the file type. In this categorization inclusion of RAID type files helps to construct a new model that can proved to be robust and able to combat a situation of storage failures. Also in the underlying model the security is further ensured by a system of proper authentication at times of storing or accessing data. A new model named weight based optimized routing is introduced in paper (Debabrata Sarddar et al.2015). Due to multiple accesses at the same time sometimes the network become congested interrupting the smooth exchange of information among the nodes connected via cloud. The mitigation of network congestion accelerates the speed of data exchange also in the mobile network moving towards a green cloud computing environment. The technique of using gossip protocol in computer communication is discussed in papers (M. Jelasity et al.2005, F. Wuhib et al.2010). Most of the gossip techniques used till date deals with static input and generation of single value output. The papers also consider the advantages and disadvantages of a dynamic input system over the static one. Whenever the input changes the system restarts automatically and new output needs a global synchronization. The biggest advantage is that it avoids the complication world- wide synchronization and any change in the input will

be adopted in the system. On the other hand the disadvantage of a dynamic input system is that a protocol that is subject to successive execution of a series of input is complex to investigate. In addition a system that gives quick responses to a change in the inputs make the system more costly than one where execution of input is taken place after each synchronized runs. Finally, the networking system using gossip rely on the dynamic input system where inputs are executed on a continuous basis as the advantages of such a system overstates all its shortcomings so far discussed (F. Wuhib et al.2009), (F. Wuhib et al.2010).

METHODOLOGY

This paper looks for a method that not only makes the data available whenever needed but also focuses on the issues of storage optimization. In the existing technology the data are well accessible to the users but the problem is to make those data available the service provider relied on replicating multiple copies of the same data. As a result a significant storage space is required which again subject to huge power or energy consumption. The proposed method attempts to solve this issue and make cloud more efficient through the use of gossiping protocol. The storage service is provided in a fashion that not only requires smaller storage space but also saves power as a related consequence. The storage method is designed as follows. Firstly, regional data centers (RDCs) are constructed depending on the availability of the data in the system and their geographical location. Then the RDCs are connected following the technique of gossiping. As like gossip protocol the information are exchanged between the connected pair of nodes here the connected RDCs shares their content table with each other. As a result any data can be accessed any time from the system storage. Finally a central data center (CDC) is formed for backing up all the data from the respective RDCs under a single roof. All the RDCs are connected with CDC and the users nodes are connected with RDCs.

Accumulation of New Information at RDC:

Not only existing data but also the newly generated data or information is stored in the RDC. With every addition of new information the RDC's content table is updated by default. Any data is replicated once for keeping a backup in the CDC. In case any resources are missing from the RDCs, the same can be retrieved from the CDC. The existing or updated index table is shared among the RDCs by the means of gossip technique. Any new information is also updated in the CDC apart from RDCs.

Accumulation Algorithm

Step 1: New information has been received by the RDC.

Step 2: RDC stores the entire information in its local storage.

Step 3: The replication of the new information has been stored in CDC

Step 4: RDC then creates an index in the indexed table.

Step 5: The updated indexed table has been shared among all RDC along with CDC.

Process of Information Retrieval:

In the traditional method whenever a user needs to access any information, the request is first sent to the local data center to check the availability. Then again the information is searched in the available RDCs. Though the users get those data but the request passes through a cumbersome and time consuming technique. But in the proposed system since the content table of respective RDCs are available to all the RDCs searching is done only once. If the required data is not available in the current RDC then the request processing required one more step that is redirection through shortest path routing otherwise the data is physically transferred to the client. Even though the information may not be found in the current RDC, storage space and searching time is optimized because of the efficient routing path between the RDCs.

Information Retrieval Algorithm

Step 1: Client sends an information retrieval request to the RDC it connected with.

Step 2: RDC scans the content table of its own and other RDC as well

Step 3: After completion of scanning, RDC find out the availability of the content

Step 4: If the RDC carry the content itself then it is sent directly to the client

Step 5: Else the request would be redirected to the destined RDC using shortest path.

Step 6: The destined RDC receives the redirected request.

Step 7: As it is a redirected request so information has been sent directly without scanning again.

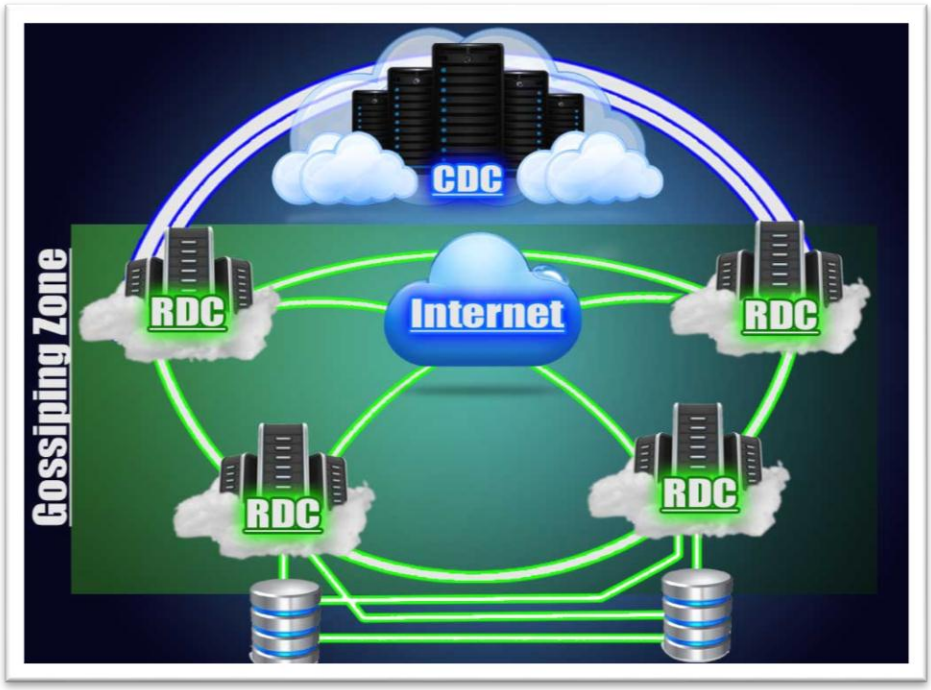
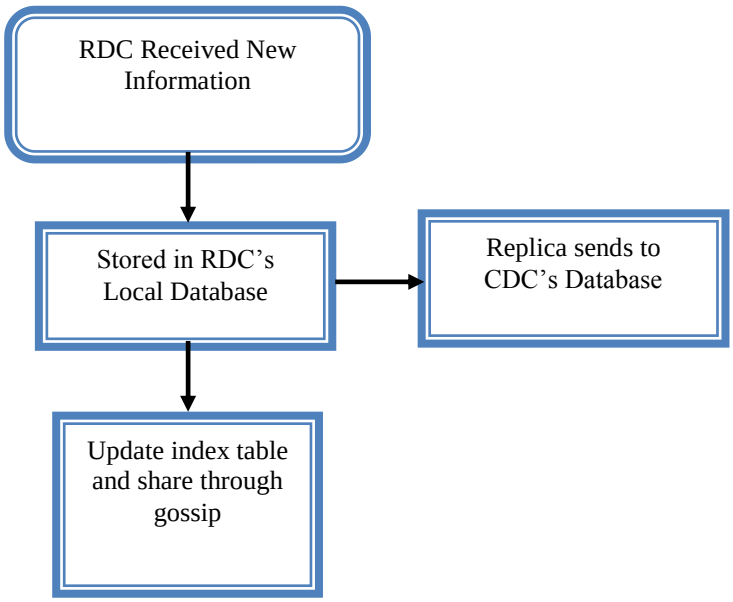


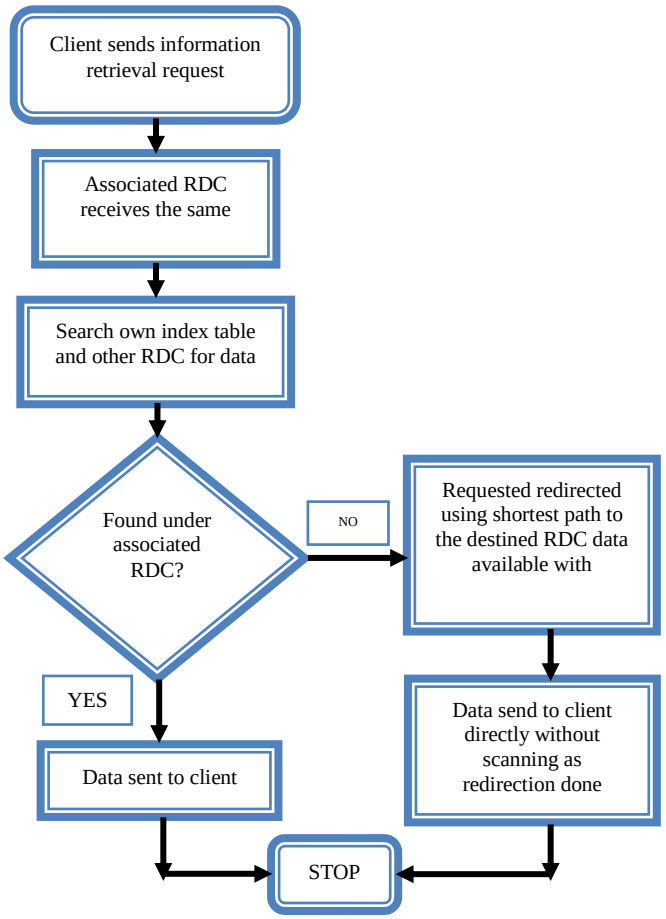
Figure 1: Proposed Architectural Diagram

Flowchart

Flow Chart for Accumulation



Flow Chart for Information Retrieval



RESULTS AND DISCUSSION

The objective of the paper is to optimize searching technique in the cloud. For this we use a networking system similar to that of a gossip protocol. Here the RDCs are connected in a way similar to that of the pair of nodes in gossip. Each RDCs exchanges information and also knows the efficient and shortest routing path. As a result any information request is processed in the least possible time. The possibility of RDC's failure is also not ignored. So, to prevent any data loss the idea of a central data center (CDC) is considered. In CDC all the information of all the RDCs are replicated. Since the CDC is not directly connected for clients use the information is kept in a safe secure manner. The following two figures are used to present our analysis graphically. Figure 2 clearly shows the average time required in the proposed technique is significantly less as compared to the existing one. Another important aspect is routing time. The average routing time for various processes is represented in figure 3.

As is evident from the figure below, search times are more in case of hierarchical searching technique. However, our approach makes use of the advantage offered by implementing gossip protocol. Parallel search requests can be sent out simultaneously among other RDCs.

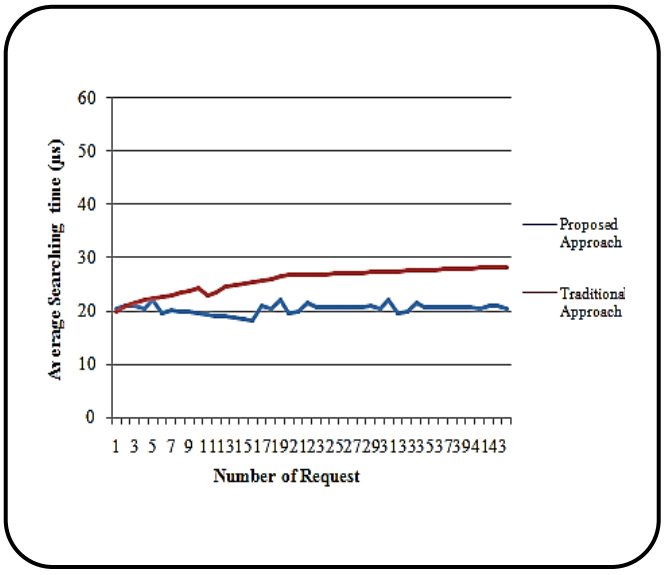


Figure 2: Average searching time vs. Number of Request

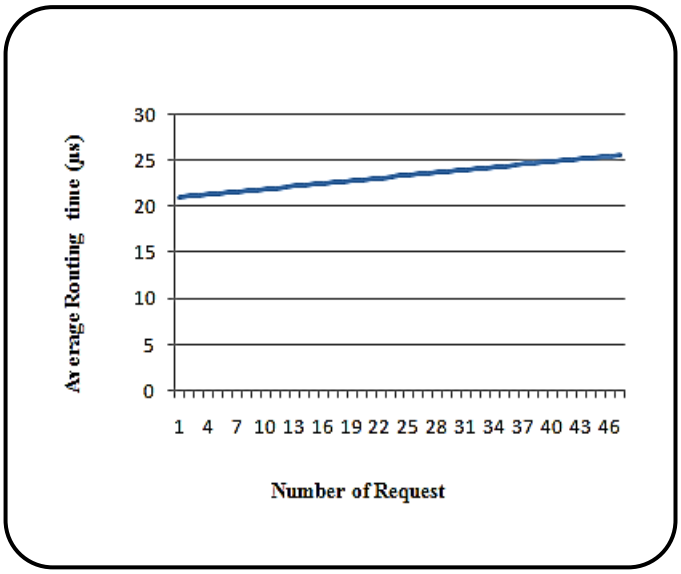


Figure 3: Average Routing time vs. Number of Request

CONCLUSIONS

Cloud computing not only helps to store the data coming from all the spheres (personal, business, IT etc.) but also opens up various scope of research in this field. In the present paper we focus on one of such thing namely on the storage and search optimization. Searching time is minimized since any requested information is searched only once. This is because of the smooth accessibility of information index tables among all the RDCs involved. The storage space is also optimized as all the data are copied once and stored in the central data center instead of storing data by the process of segmentation in the traditional approach. If the data storage requires larger space then energy consumption increases. As a result heats are generated at a faster rate causing damage to the hardware and shortening hardware life. The security and reliability issues are taken care of by building CDC and keeping backup of the data here. The present paper mostly concentrated on the secure storage in cloud. The detailed accessibility of information is not discussed here which is subject to further scope of research.

ABBREVIATIONS

RDC: regional data center; CDC : central data center; RAID : redundant array of inexpensive disk

ACKNOWLEDGEMENT

Authors gratefully acknowledge to CSE Department of University of Kalyani and JIS College of Engineering, for providing lab and related facilities for do the research.

REFERENCES

[1] F. Machida, E. Andrade, D. S. Kim, and K. S. trivedi, "Candy: Component-based Availability Modeling Framework for Cloud Service Management Using SysML," Proc. 30th IEEE Symp. on Reliable Distributed Systems (SRDS 11), IEEE Press, Oct. 2011, pp. 209-218.

[2] B. Addis, D. Ardagna, B. Panicucci, and L. Zhang, "Autonomic Management of Cloud Service Centers with Availability Guarantees," Proc. 3rd International Conference on Cloud Computing (CLOUD 10), IEEE Press, July 2010, pp. 220-227.

[3] A. Widjajarto, S. H Supangkat, Y. S Gondokaryono, and A. S Prihatmanto, "Cloud computing reference model: The modelling of service availability based on application profile and resource allocation," Proc. International Conference on Cloud Computing and Social Networking (ICCCSN 12), IEEE Press, April 2012 pp. 1-4.

[4] Sarddar Debabrata, Rajesh Bose, and SudiptaSahana. "An Enhanced Cloud Network Load Balancing Approach Using Hierarchical Search Optimization Technique." International Journal of Hybrid Information Technology 8.3 (2015).

[5] Demers, Alan; Greene, Dan; Hauser, Carl; Irish, Wes; Larson, John; Shenker, Scott; Sturgis, Howard; Swinehart, Dan; Terry, Doug (1987-01-01). "Epidemic Algorithms for Replicated Database Maintenance". Proceedings of the Sixth Annual ACM Symposium on Principles of Distributed Computing. PODC '87. New York, NY, USA: ACM: 1–12. doi:10.1145/41840.41841. ISBN 089791239X.

[6] Rajesh Bose, Sudipta Sahana, Debabrata Sarddar "An Enhanced Storage Management Scheme with Search Optimization for Cloud Data Center" in International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 12 (2015) pp. 32141-32150

[7] Sudipta Sahana, Rajesh Bose, Debabrata Sarddar "Harnessing RAID mechanism for enhancement of Data Storage and Security on Cloud " in Brazilian Journal of Science and Technology 2016 3:12, DOI: 10.1186/s40552-016-0024-5

[8] Debabrata Sarddar, Rajesh Bose, Sudipta Sahana "A Novel Approach on Weight Based Optimized Routing for Mobile Cloud Computing" in Brazilian Journal of Science and Technology (2015) 2: 3. doi:10.1186/s40552-015-0008-x

[9] M. Jelasity, A. Montresor, and O. Babaoglu, "Gossip-based aggregation in large dynamic

networks," ACM Transactions on Computer Systems, vol. 23, no. 3, pp. 219–252, 2005.

[10] F. Wuhib, R. Stadler, and M. Spreitzer, "Gossip-based resource management for cloud environments," in International Conference on Network and Service Management, Niagara Falls, Canada, October 2010.

[11] F. Wuhib, M. Dam, R. Stadler, and A. Clem, "Robust monitoring of network-wide no. 2, pp. 95 –109, June 2009.

[12] F. Wuhib, M. Dam, and R. Stadler, "A gossiping protocol for detecting global threshold crossings," IEEE Transactions on Network and Service Management, vol. 7, no. 1, pp. 42 –57, March 20