International Innovative Research Journal of Engineering and Technology
ISSN: 2456-1983 Vol: 3 No: 3 March 2018

Bandwidth Optimized Resource Allocation Process in Querying Distributed Storage

Ezhumalai.P[1], Ezhilvanji.S[2]
[1]Assistant Professor, Department of Computer Science & Engineering, Christ Institute of Technology, Pondicherry University, Puducherry, India
[2]Scholar, M.Tech, Computer Science & Engineering, Christ Institute of Technology, Pondicherry University, Puducherry, India
[1]ezhusuccess16@gmail.com, [2]ezhilganga15@gmail.com

Abstract: Distributed storage system is an advanced form of storing the data on multiple storage devices with high reliability. However, the challenging factor in the distributed storage system is the utilization of bandwidth which is more costly. In this paper, proposed Bandwidth Optimization Administrator (BOA)for querying distributed storage system and minimizing the bandwidth utilization. The data can be accessed and retrieved by the query language in the distributed storage system. The optimized bandwidth framework reduces the bandwidth utilization and greatly improves the query process.

Keywords: Bandwidth optimizations, distributed storage system, query process.

1. INTRODUCTION

The volume of data is increase exponentially; the legitimate solution to handle the huge amount of data is distributed storage system. In distributed storage systems, data are stored remotely in a number of storage nodes without redundancy [1]. The main goal of distributed storage system is to provide availability, fault tolerance, and scalability. Applications involve storage in large data centers and peer-to-peer storage systems such as OceanStore, Total Recall, DHash++ [2] that utilizes the node across the Internet for distributed file storage. It should be capable to adopt the rapidly growing storage demands, avoid loss of data in case of hardware issues, allows efficient distribution of the stored content[3][4].

Query lies as the fundamental process for accessing the distributed storage system. Queries contain the following operations such as join, union, difference, projection, and restriction [5]. Each and every operator has a group of specified tasks (scan, build, probe, etc.) which are waiting to be scheduled. The tasks are scheduled based on the available resources which extremely influence the runtime of the queries. Resource allocation in the distributed system corresponds to different factors such as CPU, memory, storage unit or network. Other factors include data resources, metadata and its contents such as a database, services, programmed to accomplish a specific task [6]. On the other hand, a node corresponds to a computer in the distributed system which contains some of these resources with a set of characteristics. Resources allocation for the query processing domain can be described as non-injective, non-subjective, multi-valued function from the set of tasks to the set of candidate resources, where tasks corresponds to the operations, which compose operators in a query (union, join, difference, etc.).

2. RELATED WORKS

Li-Der Chou et al., [7] Dynamic Power-saving Resource Allocation (DPRA) mechanism based on a particle swarm optimization algorithm in cloud is proposed. The DPRA mechanism not only considers the energy consumption of physical machine (PM) and virtual machine (VM) but also handles the energy efficiency ratio of air conditioner. Additionally, the least squares regression method is utilized to forecast PM's resource utilization for allocating VM and eliminating VM migrations. Experimentally DPRA mechanism is compared with three familiar allocation schemes and one previous solution.

Ju Yong Lee et al., [8] different classes of traffic are estimated using the fluid simulation techniques that can reduce the simulation complexity, compared to packet level

International Innovative Research Journal of Engineering and Technology
ISSN: 2456-1983 Vol: 3 No: 3 March 2018

simulation. Bandwidth optimization process is proposed between the guaranteed bandwidth vector and the corresponding performance by exterior penalty function. The main objective of this model is to minimize the total bandwidth. Fluid simulation is performed with Generalized Processor Sharing scheduling using the bandwidth allocation algorithm. Packet-level simulation is also performed with weighted fair queuing (WFQ), which approximates packet-level GPS.

Bart Theeten and Nico Janssens [9], designed Streaming analytics systems to perform on large datacenters. However this took unlimited bandwidth between data center nodes. Continuous Hive (CHive), a streaming analytics platform tailored for distributed telecommunication clouds. This optimizes query plans to minimize their overall bandwidth consumption when deployed in a distributed telecommunication cloud. Additionally, these optimized query plans have a high degree of parallelism built-in, benefiting speed of execution. Chive query language (CHiveQL) is strongly inspired by Esper's event processing language (EPL).

Li Liu and Qi Fan [10], a two-stage optimization strategy is proposed to provide a resource allocation in cloudlet environment. Initially, a cloudlet selection process based on mixed integer linear programming (MILP) is proposed to obtain the cloudlet for mobile users by optimizing latency and mean reward. Later, a resource allocation model based on MILP is presented to allocate resources in the selected cloudlet by optimizing reward and mean resource usage. A comparison of resource allocation is analyzed with a cloudlet selection model based on MILP and an existing cloudlet selection strategy in the multi-cloudlet environment.

Dinil Mon Divakaran and Mohan Gurusamy [11] an approach to use bandwidth allocation is based on bandwidth-guarantee (BG) or time-guarantee (TG), but not both simultaneously. Hence the solutions are tailored for one type of requests. A BG request demands guarantee on bandwidth. It define a new model that allows users to not only submit both kinds of requests, but also specify flexible demands: two-phase, adaptive and flexible bandwidth allocator (A-FBA) that, in one phase admits and allocates minimal bandwidth to dynamically arriving user requests, and in another phase, allocates additional bandwidth for accepted requests maximizing revenue. The problem formulated in first phase is NP-hard, while the second phase can be solved in polynomial time.

3. PROPOSED SYSTEM

The priority of the data in the network are divided into four levels based on real-time requirements, each level is identified with the first two bits of the identifier, for example, "00" stands for emergency information; "01" represents a strong real-time information; "10" represents weak real-time information; "11" represents the non-real-time information.

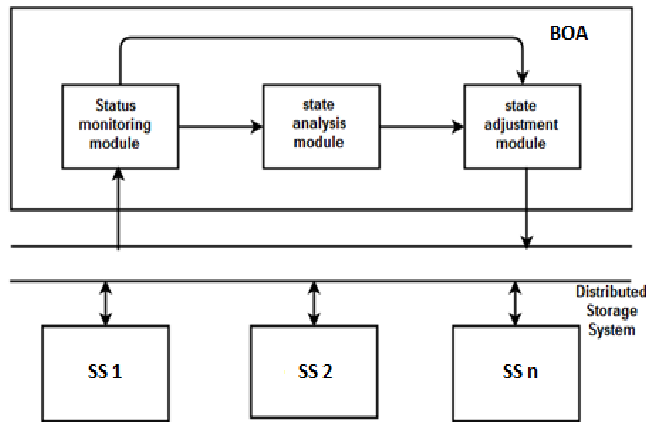

Figure 1. Bandwidth Optimization Framework

Distributed storage system consists of three modules, there are network status monitoring module, network status analysis module and network status adjustment module. The module functions as follows: (1) network status monitoring module is used to get the network running status,and send request of collecting network status in every cycle, besides, it receive feedback data packets of running status from the sub-system; (2) network state analysis module is used to calculate and find the distributed storage system with best operating performance in the whole distributed storage system; (3) network state adjustment module selects control sub-systems in DS_n with excellent overall performance, the selected sub-systems have better overall performance than the average value of the sub network DS_n. Bandwidth Optimizing Administrator will delay the first competition time of its waiting data frame for bus, namely decrease the priority of the data frame.

4. PERFORMANCE EVALUATION

We assume a Distributed storage system (DSS) with 16 sub-systems named nodes, the monitoring cycle of DSS network is: hm=1s, and transfer rate is: rate = 600kbps.The sub-networks named SS-1, SS-2 and SS-n have four sub-control systems, respectively. We used random query of different time intervals for each sub-control system, and set up six kinds of random query cycle T, the value of T can be 10s' 20s' 30s' 40s' 50s and 60s.In order to evaluate the overall performance of our proposed method BOA framework, and compared the two methods of BOA and FBA (Fixed bandwidth allocation) in terms of control quality and requirement degree of bandwidth. The distributed storage system runs for 20 minutes under two kinds of bandwidth allocation strategy of BOA and FBA, respectively. And for each interference state added up data, the cumulative sum of the absolute error of 16 control subsystems is run for 20 minutes.

The bandwidth allocation strategy of FBA and BOA, from it can know that the Data accessed from DSS using BOA is smaller than the Data accessed from DSS using FBA. As a result, our proposed method improves the overall control performance of DSS.

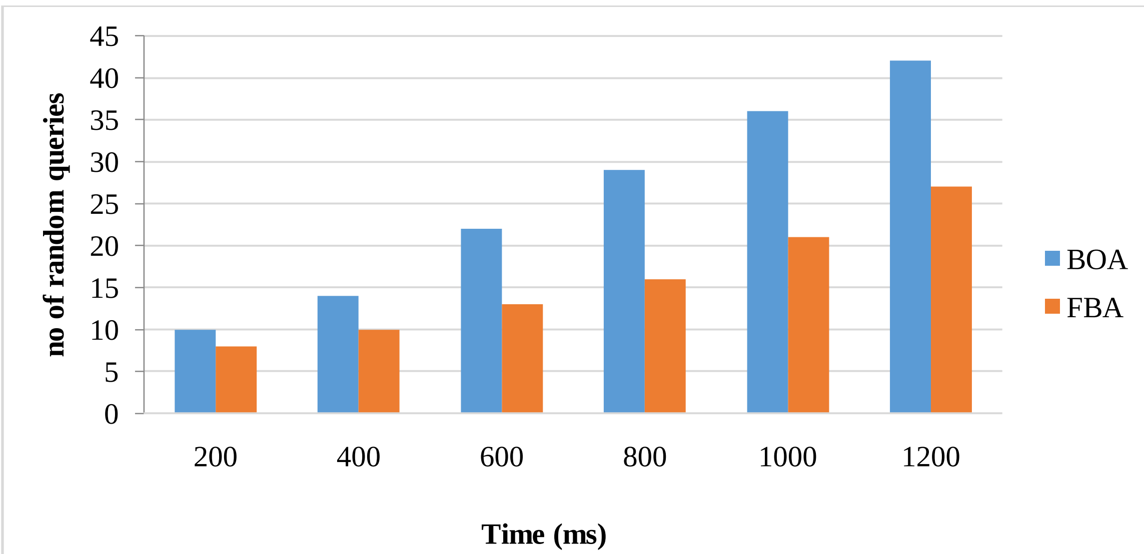

Figure 2. Represents the comparison between the BOA and FBA for random query with time

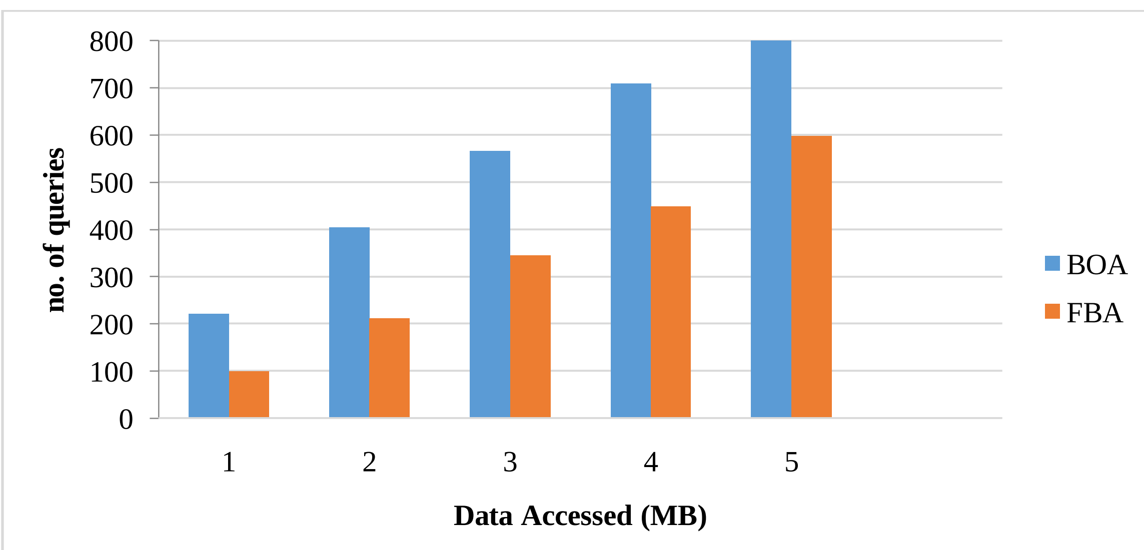

Figure 2. Represents the comparison between the BOA and FBA for number of query with data accessed

In addition, we counted the number of occurrences of six different random queries and the related system operation data during the whole network operation, respectively. It calculates the control performance and transmission performance of four sub-network systems with different priority levels based respectively on BOA and FBA strategy.

5. CONCLUSION

Distributed storage systems provide reliable access to data through redundancy spread over nodes across the Internet. A key goal is to minimize the amount of bandwidth used to maintain that redundancy. This paper proposes a bandwidth optimization administrator in distributed storage system. The BOA framework allows the system to access the data at minimum bandwidth. The BOA can optimize network resource allocation and management based on the current state of network operation, so it allows more data to access and retrieve.

REFERENCES

[1] Xiaohu Tang, Bin Yang, Jie Li, Henk D. L. Hollmann, “A New Repair Strategy for the Hadamard Minimum Storage Regenerating Codes for Distributed Storage Systems”, IEEE Transactions on Information Theory, pp. 5271-5279, Vol. 61, Issue: 10, Oct. 2015.

[2] Alexandros G. Dimakis, P. Brighten Godfrey, Yunnan Wu, Martin J. Wainwright, Kannan Ramchandran, “Network Coding for Distributed Storage Systems”, IEEE Transactions on Information Theory, Vol. 56, No. 9, Sep. 2010.

[3] Juntao Fang, Shenggang Wan, Ping Huang, Changsheng Xie, Xubin He, “Early Identification of Critical Blocks: Making Replicated Distributed Storage Systems Reliable Against Node Failures”, IEEE Transactions on Parallel and Distributed Systems, Vol. 29, Issue 11, 2018.

[4] A.Mei, L.V. Mancini, S. Jajodia, “Secure dynamic fragment and replica allocation in large-scale distributed file systems”, IEEE Transactions on Parallel and Distributed Systems, Vol. 14, Issue: 9, 2003.

[5] Roman Kolcun, David E. Boyle, Julie A. McCann, “Efficient Distributed Query Processing”, IEEE Transactions on Automation Science and Engineering, Vol. 13, Issue: 3, 2016.

[6] Yunfeng Lin, Ben Liang, Baochun Li, “Priority Random Linear Codes in Distributed Storage Systems”, IEEE Transactions on Parallel and Distributed Systems, Vol. 20, Issue: 11, 2009.

[7] Li-Der Chou, Hui-Fan Chen, Fan-Hsun Tseng, Han-Chieh Chao, Yao-Jen Chang, “DPRA: Dynamic Power-Saving Resource Allocation for Cloud Data Center Using Particle Swarm Optimization”, In IEEE Systems Journal, Vol. 12, Issue 2, 2018.

[8] Ju Yong Lee, Sunggon Kim, Deok seong Kim, Dan Keun Sung, “Bandwidth Optimization for Internet Traffic in Generalized Processor Sharing Servers”, IEEE Transactions on Parallel and Distributed Systems, Vol. 16, No. 4, Apr. 2005.

[9] Bart Theeten and Nico Janssens, “CHive: Bandwidth Optimized Continuous Querying in Distributed Clouds”, IEEE Transactions on Cloud Computing, IEEE Transactions on Cloud Computing, Vol. 3, Issue 2, Apr-Jun 2015.

[10] Li Liu and Qi Fan, “Resource Allocation Optimization Based on Mixed Integer Linear Programming in the Multi-Cloudlet Environment”, In IEEE Access, Vol. 6, pp. 24533-24542, 2018.

[11] Dinil Mon Divakaran, Mohan Gurusamy, “Towards Flexible Guarantees in Clouds: Adaptive Bandwidth Allocation and Pricing”, In IEEE Transactions on Parallel and Distributed Systems, Vol. 26, Issue 6, pp. 1754 – 1764, 2015.

