Computers and Electrical Engineering 000 (2017) 1–13

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Towards an improved heuristic genetic algorithm for static content delivery in cloud storage☆

Zhigao Zheng[a,1], Zunxin Zheng[b,2,*]

[a] Luoyu Road 1037, Wuhan, Hubei, 430074, China
[b] Nanhai Ave 3688, Shenzhen, Guangdong, 518060, China

ARTICLE INFO

Article history:
Received 5 January 2017
Revised 8 June 2017
Accepted 11 June 2017
Available online xxx

MSC:
00-01
99-00

Keywords:
Cloud storage
Static content delivery
Resource management model
Cost model
Genetic algorithm

ABSTRACT

A key challenge in computer networking is how to organize network topology effectively among a large number of servers in the cloud storage system. In a cloud environment, the topology, which is different from the underlying topology, may be established in any form at any potential edge peers. The cloud content delivery network (CDN) always faces problems of complex distributed path creation, cache update, load balancing, etc. To address the problem as a static content delivery, we propose an Improved Heuristic Genetic Algorithm for Static Content Delivery in Cloud Storage (IHGA-SCDCS) based on a resource management model and cost model. The static content delivery in cloud storage is abstracted into mathematical model for set solving problem, which is then solved by an improved Genetic Algorithm (GA). Finally, the optimal solution is reduced to an optimal content delivery program. The simulation experiment, based on CloudSim, shows that IHGA-SCDCS can effectively obtain optimal solution while reducing delivery cost.

1. Introduction

Reasonable network topology and resource management model can not only improve network performance but also guarantee effectiveness and load balancing for resource allocation, thus improving the performance of cloud storage services. Cloud storage is an Internet service itself, which emphasizes cloud data center providing resource while weakening hardware and software capabilities of terminals [1]. With cloud storage service, all content needed is kept in distance data centers and users access it via network. Thus, the cloud storage service may also use CDN in an accelerated manner, like other Internet services, to achieve higher access efficiency and better user experience [2]. The cloud CDN reduces its own cost using competitive price provided by different cloud. Combined with on-demand service of cloud, the cloud CDN can easily adjust its own storage and bandwidth usage according to the requirement. It may also reduce cost by reducing the quality of service.

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. R. Varatharajan.

* Corresponding author.
E-mail addresses: zhengzhigao@hust.edu.cn (Z. Zheng), zxzheng@szu.edu.cn (Z. Zheng).

[1] Zhigao Zheng is with Services Computing Technology and System Lab/Cluster and Grid Computing Lab/Big Data Technology and System Lab, School of Computer Science and Technology, Huazhong University of Science and Technology.

[2] Zunxin Zheng is with College of Economics, Shenzhen University. Dr. Zunxin Zheng is a Professor in Finance and he is the Vice Dean of College of Economics

http://dx.doi.org/10.1016/j.compeleceng.2017.06.011
0045-7906/

Many scholars have made preliminary studies on CDN programs based on cloud storage. The most representative one is MetaCDN proposed by Broberg et al.[3], which is a low-cost CDN using a storage cloud resource. The system provides a mechanism to place content on network provided by a different storage cloud service supplier and regularly replies appropriate copy answering requests from users. However, the system does not incorporate a new caching strategy and load balancing algorithm. The F_cache content acceleration cache technique developed by FastWeb just caches and finds optimal match on web acceleration content according to strategy. Within a period of time, it does not access file entity from the source website for repeated access but rather it copies content directly from cache to users, thus effectively improving response and saving bandwidth [4]. The cost of cloud CDN includes bandwidth and storage cost. Nevertheless, existing content delivery methods have not arrived at a reasonable solution to a pricing mechanism of cloud CDN. Meanwhile, the network topology of cloud storage is quite different from that of traditional CDN [5]. To operate cloud CDN, it is necessary to study an efficient content delivery program and reasonable load balancing strategy in cloud storage. The CDN is a whole system made up of four elements: content delivery, load balancing, content management and distributed storage after strategic analysis and implementation. The content delivery strategy is one of the key factors in CDN network planning, whose design directly determines whether the core idea as nearest service of CDN can be realized. In accordance with nearest service principle and edge server load balancing strategy, CDN ensures providing service for resource requests from users with an extremely efficient way.

The cost model in existing content delivery technologies has been developed to include one or more types of costs as download, storage and upload. As to reducing cost in content retrieval, research has shown that replica placement in general network topology is an NP-complete problem [6,7] and the optimal solution for tree topology has also been determined. Some heuristic algorithms were evaluated in [8] to find a greedy algorithm providing optimal performance. A heuristic content delivery algorithm based on fan-out was proposed by Radoslav et al. [9] and Jamin et al. [10]. Except for retrieval cost, the upload cost was further added [11]. The cost for storage was also supplemented [12,13]. In addition, the fee for retrieval, upload and storage were comprehensively considered in [14], which also provided a solution for tree topology. Other research shifted focus to adding service quality requiring all user requests to arrive at the edge server within certain network distance. The algorithm for optimizing overall storage and update cost was brought out in [15], where it is assumed that the request starts from any peer and the retrieval cost is ignored. In [16], the limitation on server capacity was added while simultaneously optimizing storage and retrieval cost.

The essence of the above solutions is static delivery. Some methods assume requests initiated by all peers evenly. Some algorithms use past request modes to customize delivery strategies. In [17], the delivery strategy was modeled as a Markov decision process and a centralized heuristic algorithm was proposed. In [18], the distributed heuristic algorithm was investigated further. The content delivery and traffic redirection was optimized in [19] to achieve request load balancing in the content delivery process while the problem of transferring a group of cached copy was solved in [20].

The traditional content delivery technologies in CDN have been widely studied. However, existing results cannot directly apply to cloud storage CDN in that many researchers in the past assumed network topology is provided as tree stored in the source server. In the present cloud environment, it is possible to establish any topology among all potential edge cloud storage peers, which may be different from the underlying network topology. Therefore, content delivery has complex problems of distributed path building, cache update and load balancing in cloud CDN. Additionally, the edge is usually undirected in traditional CDN. However, the cost for upload and download in cloud storage is different and needs a directional edge. This means that it is not enough to consider a pricing mechanism in a single direction, which asks for more reasonable pricing strategies.

Load balancing is an integral part of content delivery technology. Any content delivery algorithm should determine an appropriate load balancing strategy. Once the content delivery is completed, the user requests are redirected to a corresponding edge server using random methods in CDN like Uniform Resource Locator (URL) rewriting or transparent interception of requests to implement reasonable uniform delivery of users' requests.

To address static content delivery problem in CDN, the paper provides a resource management model based on a union tree and delivery cost model. An improved heuristic genetic algorithm for static content delivery in cloud storage is proposed. The rest of the paper is organized as follows: Section 2 designs the resource management model and cost model of CDN. In Section 3, the improved static content delivery algorithm is put forward. In Section 4, we carry out simulation experiments and compare the performance of the proposed algorithm with CloudSim simulator. We conclude this paper in Section 5.

2. Resource management model and content delivery cost model

2.1. Tree resource management model

Traditional cloud storage system manages data and resource with central indexing. The centralized management has a simple structure and is easy to design and management. Combining existing proxy mechanisms developed in [21], we build a tree resource management model based on P2P.

The tree structural model of cloud storage is a hierarchical structure that divides peers into several regions corresponding to the physical distance among peers, which are managed by region center peers [22]. The whole network is made up of center peers and normal peers. To improve the efficiency of cloud storage services, the CDN technology can be used among

Please cite this article as: Z. Zheng, Z. Zheng, Towards an improved heuristic genetic algorithm for static content delivery in cloud storage, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.06.011