

CONTENT DISTRIBUTION TECHNIQUE WITH IN VIRTUAL ORGANIZATION(vo) BASED PEERING CONTENT DELIVERY NETWORK

Tanjil Hossain , Jamil Ahmed Khan , Syed Tanvir Fayeze

Department of Computer Science and Information Technology, IUT, Gazipur, Bangladesh

tanj.hossain@gmail.com, munna.iut04@yahoo.com, tanvircit@gmail.com

Abstract-Content Delivery Network (CDN) is a system where computers networked together across the Internet. Contents (especially large media content) are replicated from the origin server to cache servers scattered over the world in order to deliver content to end users in a reliable and timely manner. Cooperation and coordination among them can solve many problems of a particular CDN and is achieved by Virtual Organization (VO). We tried to define an efficient content distribution technique for VO based CDN which is Similar Content Distribution Technique. Implementing this solution within VO will be able to increase performance by providing higher data transfer rate.

Keywords-Content Delivery Network, VO based peering, Content Sharing, Similar Content Distribution Technique

I. Introduction

Day by day Internet is facing the problem of unmanageable amount of traffic flow which results many requests are being lost. Content Delivery Network replicates the same content or services over several mirrored Web servers strategically placed at various locations to improve performance and scalability. The user's request is redirected to the nearest server and this approach helps to reduce network impact on the response time of the user requests. [1, 2, 3, 4]

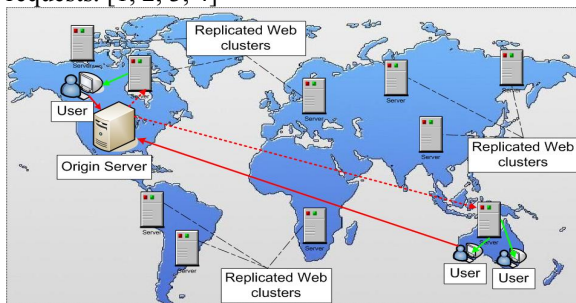


Figure 1- A simple model of CDN

In Figure 1, a user from Australia is requesting for contents from the origin server in USA. As the distance between these two continents is huge it takes a greater response time for the user. In this situation CDN provides a surrogate server (replica of origin server) in Australia to serve the end-user.

A CDN focuses on building its network infrastructure to provide the following services and functionalities: storage and management of content, distribution of content among surrogates, cache management, delivery of static, dynamic and streaming content, backup and disaster recovery solutions, and monitoring, performance measurement and reporting.

A. Problems of existing CDN

We have found some problems in existing CDN (especially commercial CDN) based on our observation. Existing commercial CDNs are proprietary in nature. Each of them has created its own closed delivery network, which is expensive to setup and maintain. Running a global CDN is even more costly, requiring an enormous amount of capital and labor. Content providers typically subscribe to one CDN and thus cannot use the resources of multiple CDNs at the same time, which may cause Service Level Agreements (SLA, An SLA is a contract between the service provider and the customer to describe provider's commitment and to specify penalties if those commitments are not met [5]) violation in many cases. To solve these problems Virtual Organization (VO) based peering CDN may be a possible solution.

B. VO based Peering

The main purpose of CDN is to provide necessary distributed computing and network infrastructure so that SLAs are met with customers. Better SLAs can be met by groups of cooperating CDN. Cooperation among the peering CDNs is achieved through a Virtual Organization (VO). Virtual organizations [6] are composed of a number of semi-independent autonomous entities (including different individuals, departments, and organizations) that come together to share resources and to collaborate on shared goal(s) [7].

C. Proposed Similar Content distribution technique

According to similar content distribution technique similar types (e.g. video, audio, images etc) of contents will be grouped together and cached in same surrogate in a virtual organization based CDN. Here the similarity will depend on the weight of the contents. In a VO the place from where most requests are coming will be measured first and will be considered as a cluster. Then distance of the surrogates from the cluster will be calculated. It may be done by measuring number of routers (Hop count) in case of same capacity link. Depending on the distance different types of grouped content will be placed in surrogates. The contents having highest weight will be placed in the closest surrogate and lighter contents will be placed in the distant surrogates. As contents having more weights needs more time to transfer, placing them in closer surrogates will reduce server load. In this technique surrogates can provide faster transfer of content and bandwidth consumption will be less.

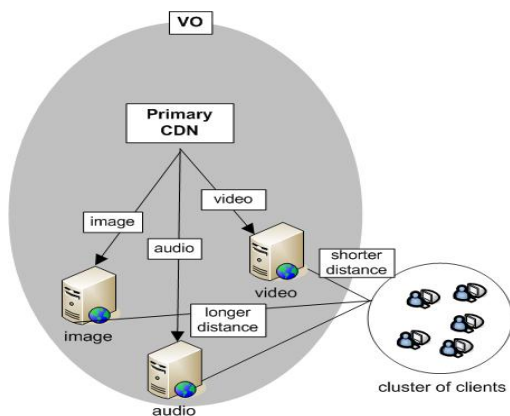


Figure 2 - Grouped Content Distribution

II. Performance evaluation

Experiment Setup

We took two approaches in our experiment. In the first approach different weighted contents are placed in each surrogate (Random Content Distribution Technique). In the second approach we placed almost similar weighted contents in a surrogate. In both case we placed surrogates in random distance from the user but for similar content distribution surrogates having heavier contents are placed closer to the user.

A. Performance criteria

Here we briefly present the performance criteria used in the experiments, namely

- ✓ mean response time
- ✓ response time CDF
- ✓ hit ratio and
- ✓ byte hit ratio

These criteria have been used since they are the most indicative ones for performance evaluation.

– **Mean response time.** This is the expected time for a request to be satisfied. It is the summation of all request times divided by their quantity. Low values denote that content is close to the end-user.

– **Response time CDF.** The Cumulative Distribution Function (CDF) in our experiments denotes the probability of having a response times lower or equal to a given response time. The goal of a CDN is to increase the probability of having response times around the lower bound of response times.

– **Hit ratio.** It is defined as the fraction of cache hits to the total number of requests. A high hit ratio indicates an effective cache replacement policy and defines an increased user servicing, reducing the average latency.

– **Byte hit ratio.** It is the hit ratio expressed in bytes. It is defined as the fraction of the total number of bytes that were requested and existed in cache to the number of bytes that were requested. A high byte hit ratio improves the network performance (i.e. bandwidth savings, low congestion etc.).

B. Performance Measurement

For performance measurement we have compared our proposed similar content distribution technique with general random content distribution, where content is distributed in random basis and unstructured way .

From our experiment we found that when client number increases in a network the mean response time always increases. But the mean response time of similar content distribution is always less than the mean response time of random content distribution. In Figure 3 for particular number of clients mean response time for similar content distribution is always less.

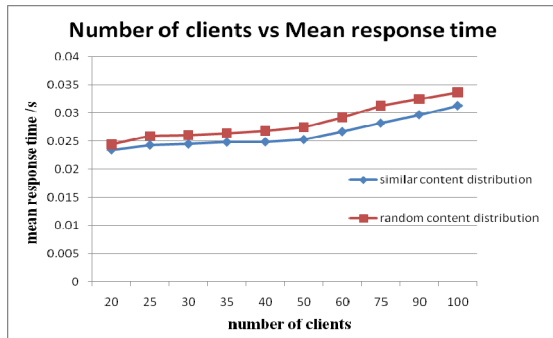


Figure 3 – Number of clients vs Mean response time

Another finding is, when client numbers are fixed in a network and the number of requests increases (Figure 4) then the mean response time of similar content distribution is always less than random content distribution.

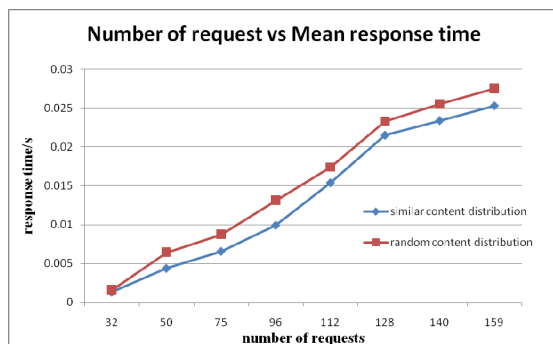


Figure 4 – Number of requests vs Mean response time

The reasons that make mean response time of similar content distribution less:

M/G/1 queue: M/G/1 queue can be used to show model of a general CDN (Shown in Figure 5). An M/G/1 queue consists of a FIFO buffer with requests arriving randomly according to a Poisson process at rate λ and a processor, called a server, which retrieves requests from the queue for servicing. We

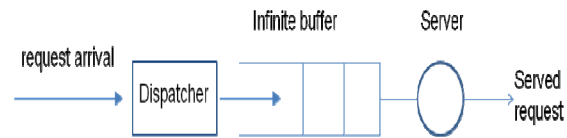


Figure 5- M/G/1 queue

assume that the total processing of the Web servers of a CDN is accumulated through the server and the service time is a general distribution. User requests are serviced at a first-come-first-serve (FCFS) order [8].

When multiple request comes at the same time then those request are inserted in this queue of that surrogate and requests are served as first come first serve basis. In this way in random content distribution technique surrogates may have long queue of requests. For this surrogate needs more time to serve multiple requests in random content distribution technique.

But in similar content distribution there will be less number of requests in each surrogate due to request distribution between different surrogates. So queue size is reduced. That's why it takes less response time.

Parallelism: In general clients request for different types of contents. In random content distribution different types of content can be placed in same surrogate. So it needs more time to serve because it is a sequential process. But in similar content distribution different content are placed in different surrogates based on similarity. So those surrogates can serve the client in parallel basis.

Generally surrogates serve contents from its cache. Hit ratio percentage is the ratio between the number of contents a surrogate is serving and the number of content request it is receiving. A high hit ratio indicates an effective cache management policy. It improves network performance and bandwidth saving. From Figure 6 we can see that for particular number of request, hit ratio percentage of similar content distribution is always higher than hit ratio percentage of random content distribution.

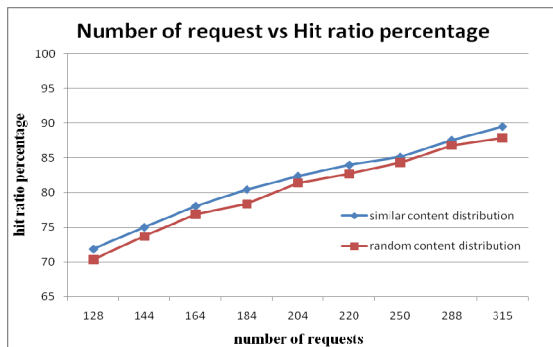


Figure 6 – Number of requests vs Hit ratio percentage

Less number of redirection: In similar content distribution most of the time the surrogates are able to serve the request as load is almost equally balanced, so redirection probability is less. But in random content distribution redirection this probability is higher, and in worst case it may happen that there is no requested content in a surrogate. So the surrogate redirects the requests to other surrogates that have those contents.

Figure 7 shows that when server load is in increase the expected waiting time also increases. But in similar content distribution technique expected time is less in compared with random content distribution technique.

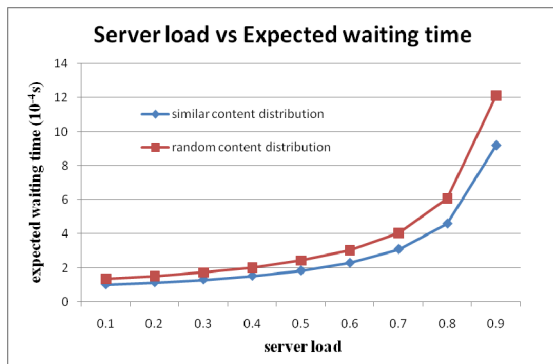


Figure 7 – Server load vs Expected waiting time

P-K formula for calculating expected waiting time:
 $E(W) = \lambda E(x) / 2(1-\rho)$

$E(W)$	Expected waiting time
$E(x)$	Mean service time
λ	Mean arrival rate
μ	Mean service rate
ρ	Load (λ / μ)

According to P-K formula when server load increases the mean service time and mean arrival rate also increases proportionally. In the parallel content distribution mean service time is less in compared with random content distribution due to M/G/1 queue.

III. Conclusions and Future work

With the proliferation of the Internet it is clear that CDN is playing a verge role, but due to huge cost it is still far away from third world countries. In this regards VO based CDN could be a desired solution considering its better QoS in service delivery to end-users and also cost cutting efficiency. In this paper we tried to define a core challenging issue, designing and building an efficient content distribution technique within a VO based peering CDN.

New forms of Internet content and services, such video-on-demand, which requires intensive bandwidth and predictable data transmission delay, are emerging. Here it is very much important to design VO based peering such that data transfer should be much faster to make the Internet more responsive. From our experiment done through CDNsim[9] simulator we found that our proposed similar content distribution technique will be much faster with in a VO in respect of response time. But proposed system still has a big challenge incase of finding an effective server level agreement between service providers. This fact opens several future research topics like finding policy management between service provider and an efficient clustering technique etc. Research on VO based peering are still undergoing and we believe that the successful solutions of different problems will make next-generation CDNs fly.

References:

- [1] G. Pallis, and A. Vakali, "Insight and Perspectives for Content Delivery Networks," Communications of the ACM, 49(1), ACM Press, NY, USA, pp. 101-106, January 2006.
- [2] A. Vakali, and G. Pallis, "Content Delivery Networks: Status and Trends," IEEE Internet Computing, 7(6), IEEE Computer Society, pp. 68-74, November-December 2003.
- [3] G. Peng, "CDN: Content Distribution Network," Technical Report TR-125, Experimental Computer Systems Lab, Department of Computer Science, State University of New York, Stony Brook, NY, 2003. <http://citeseer.ist.psu.edu/peng03cdn.html>
- [4] A. M. K. Pathan and R. Buyya, "A Taxonomy and Survey of Content Delivery Networks,"
- [5] J. Bouman, J. Trienekens, and M. van der Zwan, "Specification of Service Level Agreements, Clarifying Concepts on the Basis of Practical Research," In Proc. of the CoSTEP'99, pp. 169, 1999.
- [6] T. J. Norman, A. Preece, S. Chalmers, N. R. Jennings, M. Luck, V. D. Dang, T. D. Nguyen, V. Deora, J. Shao, W. A. Gray, and N. J. Fiddian, "Agent-based formation of virtual organizations," Knowledge-Based Systems, 17 (2-4), pp. 103-111, 2004.
- [7] A. M. K. Pathan's "Coordinate management and peering of content delivery network".
- [8] M. Arlitt and T. Jin, "A Workload Characterization Study of 1998 World Cup Web Site," IEEE Network, 14(3), pp. 30-37, May/June 2000.
- [9] CDNsim Content Distribution Network Simulator <http://oswinds.csd.auth.gr/~cdnsim>