

Applying an Effective Model for VNPT CDN

Nguyen Quy Sy

*Electronic Faculty, Posts and Telecommunications Institute of Technology (PTIT), Vietnam
synq@ptit.edu.vn*

Abstract - Most operations of Content Distribution Network (CDN) have measured to evaluate the ability to serve users with content or services they want. Activity measurement process provides the ability to predict, monitor and ensure activities throughout the CDN. Five parameters or regular measurement units are often used by content providers to evaluate the operation of CDN including Cache hit ratio, reserved bandwidth, latency, surrogate server utilization and reliability. There are many ways to measure CDN activities, one which use simulation tools. The simulation CDN is implemented using software tools that have value for research and development, internal testing and diagnostic CDN performance, because of accessing real CDN traces and logs is not easy due to the proprietary nature of commercial CDN. In this article, we will apply a CDN simulation model (based on [5]) for design a CDN based on the network infrastructure of Vietnam Posts and Telecommunications Group (VNPT Network)

Keywords - Content Distribution Networks, CDNsims, VNPT network, Cooperative Environment .

I. INTRODUCTION

The CDN was developed to overcome the limitations of the inherent Internet aspects of quality of service (QoS) that customers feel when you access Web pages. A CDN reproduce content from an original server to the surrogate server scattered on the world, in order to transport reliably and timely the content to the end user from surrogates near the users. Distribute content over the Internet has received much research interests. The research combined the development of technologies of computing with infrastructure network quality and technical management of distributed replication. Here is some concepts often used in CDN network:

- Surrogates: is one or group computer located near the client (they have usually weaker configuration than original servers) instead for the server implementing requirements of the client. Instead of transporting the client request to original server, it transferred these request to nearest surrogate. Through technical backup (replication) and caching, surrogate allows to taking data from nearly client that reduces the load on accessing core network and reduces latency from client. By surrogate contains only a partial data at the original server so that sometimes it must be redirected the requirements of clients to the original server.

- Replication: to create and maintain a copy of including a part or all data on original server in the surrogate servers.

There are 2 basic technical backups, such as: backup the entire site content and backup partial site content. For backup the entire site content, all data on the original server will be transferred to the surrogate. Then all requests from the clients are met in the nearest surrogate. But this approach is difficult, due to equipping storage for surrogates. This isn't ability, by no storage which can store all the data form the distribution of content, as well as difficulties and complexes in updating data management when a change of the site in the original server. The second, to backup partial site content overcome the difficulties of first method, it only backups the embedded objects from appropriate CDN. Therefore the base HTML pages will be taken from the original server, and embedded objects from CDN cache (the surrogate). Moreover, the embedded objects rarely change, so the approach to backup partial site content showing better enforcement.

- Caching: It is a technique to select the backup object in the surrogates for surrogates can meet the most optimal requirements form the clients based on storage capacity that is not redirected these requirements to the original server. There are many caching techniques depending on the CDN network policy. Some caching method is LRU (least recently use), FIFO (first in first out), LFU (least frequently use), FIFO (first in first out) and replace randomly.

In simulation models will be presented below, we use LRU caching method as the basis for research.

II. CDN SIMULATOR

A. Developing policy network

CDN networks can be built in cooperative environment or non – cooperative environment. The selection cooperative environment or non-cooperative environment will identify surrogates have neither collaborated or not with each other when cache miss occurs. “Collaboration” that happens when not found in surrogate, the requested contents will be transferred from the nearest surrogates. “Non-collaboration” that means contents which haven't in a surrogate, will be transferred form original servers.

When clients request objects form original server, it will be redirected to the nearest surrogate following distance (dijkstra algorithms). Then the surrogate will then check in its cache, if cache hit, it will directly serve client instead of the original server. For a cache miss it has 2 processing option:

- Cooperative environment: Surrogate will try to retrieve objects from other surrogates containing objects is requested.

Then the objects will be stored in this surrogate and are responded for client. If the requested objects aren't on any surrogates, the this surrogate will redirect these objects to the original server, and will store the objects in its cache and service to the client.

- Non-Cooperative environment: Surrogate will directly get the objects from the original server, rather than from the other surrogates. Then also stored in these objects in its cache and serve clients instead of original server.

In CDN simulator model that uses 4 CDN network policy as follows [5]:

- Cooperative Environment (closest surrogate). The client is redirected to the closest surrogate server in terms of network hops. Upon a cache miss, the surrogate server retrieves the object from the closest alternative surrogate server that contains the requested object. The object is stored in the cache and then it is served to the client. If the object is not outsourced at all in any surrogate server then the surrogate server retrieves the object from the closest origin server. Consecutive cooperation may occur if in the meantime the alternative surrogate server removed the object from the cache causing another cache miss.

- Non - Cooperative Environment (closest origin). The client is redirected to the closest surrogate server in terms of network hops. Upon a cache miss, the surrogate server retrieves the object from the closest origin server, stores it in cache and serves it to the client.

- Cooperative Environment (random surrogate). The client is redirected to a random surrogate server. Upon a cache miss, the surrogate server retrieves the object from a random alternative surrogate server that contains the requested object. The object is stored in the cache and then it is served to the client. If the object is not outsourced at all in any surrogate server then the surrogate server retrieves the object from a random origin server. Consecutive cooperation may occur if in the meantime the alternative surrogate server removed the object from the cache causing another cache miss.

- Cooperative Environment (surrogate load balance surrogate). The client is redirected to the closest surrogate server in terms of network hops. If the surrogate server is loaded ~95% then the client is redirected to the least loaded surrogate server. Upon a cache miss, the surrogate server retrieves the object from the closest alternative surrogate server that contains the requested object. Again if the load is ~95% the surrogate server is redirected to the least loaded surrogate server that contains the

object. The object is stored in the cache and then it is served to the client. If the object is not outsourced at all in any surrogate server then the surrogate server retrieves the object from the closest origin server. If the origin server load is ~95% then the surrogate is redirected to the least loaded origin server. Consecutive cooperation may occur if in the meantime the alternative surrogate server removed the object from the cache causing another cache miss.

B. Configuring the base network (network infrastructure)

CDN network is overlapping network which is building up the basic network including the routers. Depending on the configuration of each network, routers are connected differently and this base network is predetermined CDN network.

In this simulation models, configurations of the network can change as you wish, to create backbone network topology. The routers will do the connection between the network nodes and the responsibility of routing packets to correct destination on the network. The purpose of simulation model will be built CDN network, set the original server and surrogates connecting with that routers so that quality of service is best.

Simulation program will calculate so that the mount is the most reasonable. Setting up network configuration is done using a file data that each line describes a connection to router each other, "int int \n" which corresponds with first router, second router and "\n" is new line character. Each line will correspond with one connection; each connection is the default connection which is a 2-dimensional.

Currently, the model simplifies 2 parameters that are bandwidth and distance between two routers. So it allows to choice a constant speed between 2 routers. Also assume that distance between 2 routers is equal.

C. Elements of overlapping CDN network

- Client: It is a computer of users that request services. The larger clients grow, the huger the requested objects are. Number of the clients and the objects which the clients request

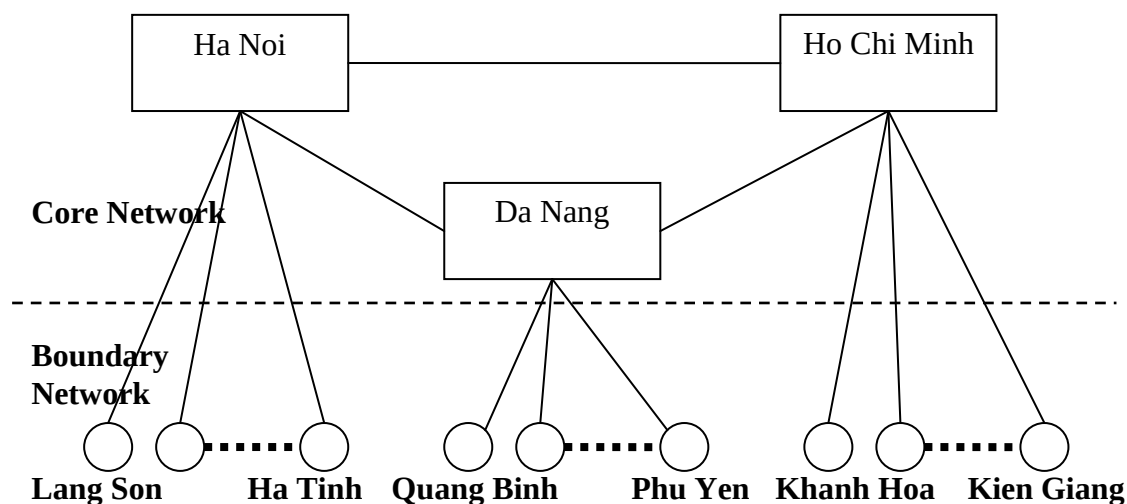


Figure 1. Topology of basic network

are necessary parameters of simulation model.

- Original server: Maybe, there are multiple original servers existing, but for simplicity, in this simulation model, we only used one original server.

- Surrogate server: Selecting the number of surrogates takes important role deciding to service quality. For this theory, parameter will be reverse rate with the average of responsible time. By servers have been limited capacities, so parameters of number of input or output are important with this server.

- Database: including the website and traffic that describe requests of clients. The data is built into the structure data file for each line of "double int int \n". "Double" is time stamp that client make its request, first "int" is order number of client, second "int" is requested object. At one time stamp, maybe, there are multiple client that requests synchronously. File data is changed by a program created outside.

- Caching technical: collaborating with re-create technical of data network to create the data backup effectively in CDN network. Caching technical is a parameter can't be lack of simulation model.

D. Results

Simulation results will indicate the location of original server and surrogate in the CDN network for each case along with the quality of service is average response time and cache hit ratio.

III. CDN SIMULATION BASED ON THE BASIC CONFIGURATION OF VNPT NETWORK.

A. Purpose of simulation

We construct the input parameters for the simulator to meet 3 goals:

- Compare response time of different policies on the CDN network that has the same clients and increases surrogates.

- Compare response time of different policies on the CDN network that has the same surrogates and increases clients.

- Evaluating the impact of size of cache memory on the surrogate for response time of CDN network.

B. The input parameters of simulator

Having 4 policies in the simulator to be compared include: Cooperative environment (closest surrogate), Non-cooperative environment (closest origin), cooperative environment (random surrogate), cooperative environment (surrogate load balance). Also we use the first 300 web object samples taken from 16000 pre-create object of the simulator. Objects are numbered from 0 to 299 with its random size. Each client has requested individual QoS. The data traffic that consists of objects that clients require and time to give require is created the random sample by external programs (written in Java).

Speed connections between the routers on the network are kept fixed is 100 Mb/s.

Base on simulator, we have applied simulation to CDN based on basic network of VNPT. Basic network model is

hierarchical model such as: core network and boundary network.

Core network has 3 routers allocated in 3 cities: Ha Noi, Da Nang, Ho Chi Minh. And this each router will be connected with boundary routers that are located in provinces around. Core routers are connected to each other in pairs, and each boundary router is connected with one core router. Assuming the clients are connected directly to the boundary router and servers are connected directly to boundary routers and core routers.

Topology network is built as Figure. 1.

We evaluated 4 above policies based on 2 different networks that have different size of caches on the surrogates (they have completely same topology):

- A network with the surrogates which have their cache that can contain the entire 300 web objects.

- A network with the surrogates which have their cache that can contain the first 29 web objects (maximum) in above 300 web object sample.

To make model the same reality, we pre-load the cache for predetermined objects. Size of cache can contain the entire 300 objects, the load is first 100 objects and Size of cache can contains above 29 objects, we loaded all 29 objects to cache at the time initialize of the network.

For each network configuration, we perform to the change in the number of clients or surrogates to evaluate its capability:

- Hold the number of surrogates are 5, increase the clients: 100, 200, ..., 1000.

- Hold the number of clients are 1000, increase the surrogates: 1, 3, 5, ..., 19.

Number of client requests the object are fixed as 3000 times, to ensure accuracy we reduce sphere of access time (to increase the number of access at a time).

In this simulation, each client can have more than one connection simultaneously to request object, the maximum connection parameter (simultaneously) is 10. Moreover, during operation of the network, when congestion occurs due to overloading surrogate than its ability or path is too long on the network that make the requirements of client not been responded after pre-determined time (we set in the simulator is 2), this client will send again its request, the number of re-sending is limited (is 5 in the simulation).

The surrogates in the network can process 50 input connections and 50 output connections simultaneously. The number of surrogates is a dynamic parameter to evaluating QoS of network. Origin server has maximum 200 input connections.

At a surrogate, caching technique is applied (LRU).

C. Simulation results

The results of the simulation model are presented through 4 charts, to give a visual assessment for research:

Increasing clients:

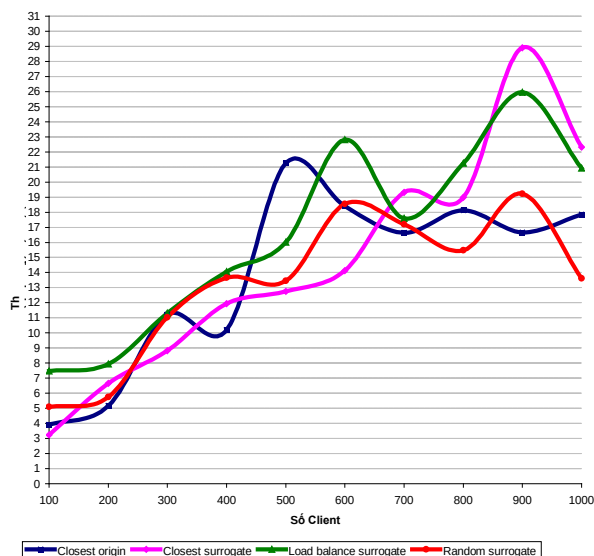


Figure 2. A network with the surrogates which have their cache that can contain the first 29 web objects

Increasing surrogates:

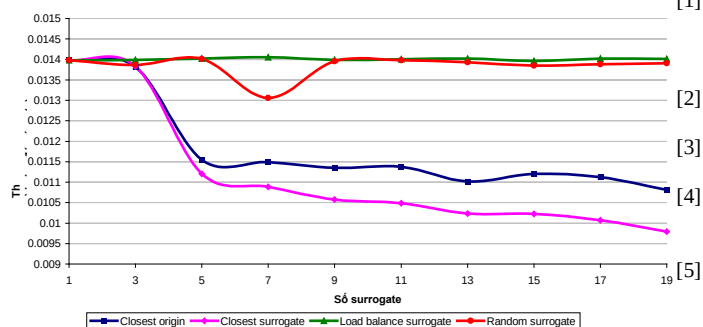


Figure 3. A network with the surrogates which have their cache that can contain the first 29 web objects

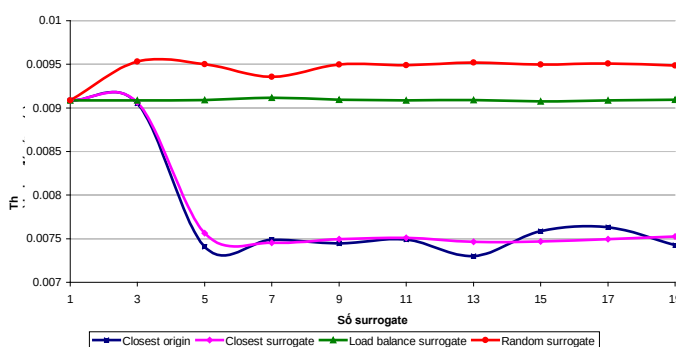


Figure 4. A network with the surrogates which have their cache that can contain the entire 300 web objects

Evaluating:

- Increasing surrogates has one critical point, if it exceeds this point, increasing surrogates almost doesn't reduce response time of CDN network.
- The greater size of surrogate cache is, the smaller response time is.
- With current network configuration in simulator, when increasing number of surrogate, closest surrogate and closest origin bring QoS clearly better. Moreover, closest surrogate policy is more stable than others.

IV. CONCLUSION

In this article we have presented a simulation model for CDN network to evaluating QoS. It works well based on basic network existing in Vietnam. However to continue to improve, this model can add some other options such as modifying neither bandwidth or distance between two routers to evaluate more accurate correcting and more actual deployment.

REFERENCES

- [1] T. Loukopoulos, I. Ahmad, D. Papadias, An overview of data replication on the Internet, in: Proceedings of IEEE International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN), 2002, pp. 27–32.
- [2] M. Rabinovich, O. Spatschek, Web Caching and Replication, Addison-Wesley, New York, 2002.
- [3] P. Mirchandani, R. Francis, Discrete Location Theory, Wiley, New York, 1990.
- [4] B. Li, M.J. Golin, G.F. Italiano, X. Deng, K. Sohaby, On the optimal placement of web proxies in the Internet, in: Proceedings of IEEE INFOCOM, 1999, pp. 1282–1290.
- [5] CDNSim, “A Content Distribution Network Simulator,” 2007. <http://oswinds.csd.auth.gr/~cdnsim/>

Author's profiles

Nguyen Quy Sy received his B.Sc in 1995 from Hanoi University of Technology, Vietnam.

He received his Ph.D. degree of Telecommunication engineering from Post and Telecommunication Institute of Technology in 2003.

From 6/1995 to 2004, He was as a researcher of Switching Department of Research Institute of Posts and Telecom (RIPT). From 2004 to present, He is a lecture of Post and Telecommunication Institute of Technology

His research interests are in the fields of switching system, access networks, data communication, advanced computer architectures.

