

Hierarchical Load Balancing for Distributed Content Delivery Network

TaeYeon KIM, HoYoung SONG

ETRI (Electronics and Telecommunications Research Institute), Korea

tykim@etri.re.kr, hsong@etri.re.kr

Abstract—Future network should focus more on what users want to get from the network rather than where users want to get through the network. Distributed content network equipped with storage and streaming server could provide effective and efficient content service to the users. We propose the hierarchical content management method to find the optimal content server in consideration of load balancing over the distributed content service network.

Keywords— Content Centric Network, Global Server Load Balancing, Traffic Optimization

I. INTRODUCTION

As the explosive growth of internet users continues, internet service providers must quickly adapt and scale their network infrastructures in order to provide high quality service to their users. According to recent predictions, global IP traffic will increase by 4 times from 2009 to 2014, reaching 64 exabytes per month in 2014.

However this fast-growing network traffic causing occasional problems in telephone service has been triggering some significant concerns that it might lower the competitiveness of the IT industry in the long run.

Even with the current network infrastructure needed to be extended and upgraded to cope with the upcoming traffic explosion, network providers are unwilling to invest on establishing more network infrastructure due to unsatisfactory returns for all their efforts and costs.

So, we are now facing a new phase to come up with new eco system to bring about a win-win situation for all providers in network, content and application as well as service users in need of enhanced user experiences from the network.

Responding to various needs from the ever-growing internet user, future network should offer alternative promising service more than end-to-end connection to the network end users. To guarantee

the fast connection for the end users when they are using the internet is not only role of network anymore. Future network should focus more on what users want to get from the network rather than where users want to make a connection through the network [1-3].

For the purpose of providing optimized high quality service, especially for the content service to network users directly, the service network is equipped with storages and servers inside the network.

The content service network consists of several content service nodes and each node is a combination of server, storage and network devices altogether. The network device of the node performs both interconnections amongst network nodes and combining server and storage into the network device to build up the content service node.

Another network element for the content service network is the content service network manager performing network monitoring, coordination and portal server of the content or application service. Authentication, consent and path provisioning functions are also embedded in the content service network manager [4-5].

The rest of this paper is organized as follows. Section II summarizes global server load balancing functional architecture for the distributed content service network. Then, Section III presents hierarchical load balancing operational procedures for the content service network. Finally, a brief conclusion is provided in Section IV.

II. GSLB FOR DISTRIBUTED CONTENT DELIVERY NETWORK

Global Server Load Balancing (GSLB) is originally created to minimize the damage from the natural disasters such as earthquakes or tornados.

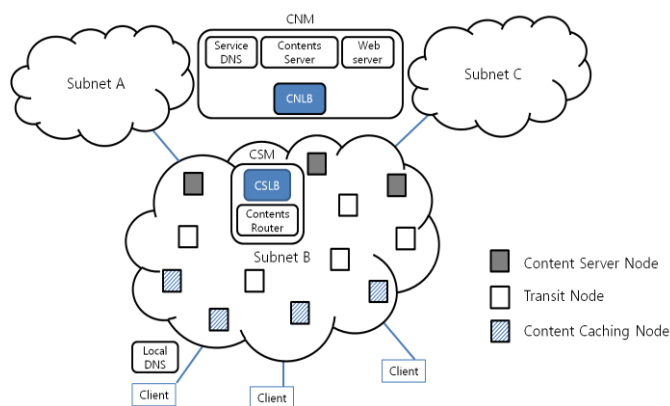


Figure 1. An example of Distributed Content Service Network

Its main role is maintaining consistency for the datum in data center safely using backup servers or remote recovery data center.

In distributed content delivery network shown in Figure 1 where a multitude of contents are distributed over the network a special method to manage all of the contents is required. To do this we subdivided the whole network into subnets connecting clients with content servers based on regional locality.

A. Content Network Management

Content Network Manager (CNM) performs service gateway for the clients who want to get content service from the network by informing content server address to the clients. CNM has abstract knowledge of subnets in the content network and selects optimal subnet to provide the content requested.

Content Network Manager taking care of whole content network cooperates with CNLB (Contents Network Load Balancing) to find the optimal subnet and Content Server for connection between the client subnet which the client sits in and server subnet which holds the requested content in it. It is also needed for service DNS function to translate the content name into IP address for the user side in cooperation with the web server which provide the users web service interface.

Figure 2 shows structure of Content Network Load Balancing. The gathering function of the subnet abstract information on each subnet resides in the Subnet Information Gateway (SIG) joint with the CSLB in each subnet.

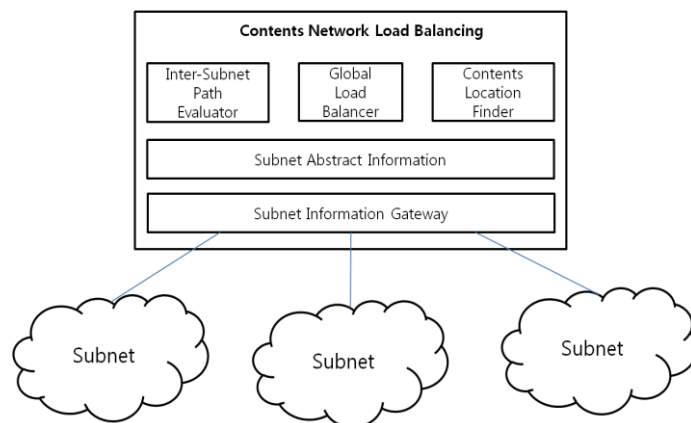


Figure 2. Content Network Load Balancing System Structure

The Subnet Abstract Information includes condensed information about contents directory and status for each subnet to identify which subnet contains the requested content without unnecessary information of the subnet.

Inter Subnet Path Evaluator estimates the costs between the server subnets to determine the optimized subnet when multiple server subnets were founded. Also, searching for the content and finding the subnet including the content are the role of Content Location Finder. To sum up these functions Global Load Balancer returns the optimal subnet in response to the content service request.

B. Content Subnet Management

In each subnet, there are a group of service nodes composing the subnet and Contents Subnet Manager (CSM) for the purpose of collecting and manage the information of network and contents distributed over the subnet.

CSM has the CSLB (Contents Subnet Load Balancing) function that finds the optimal content service node in the subnet. Content Router function takes responsibility of connection between the client service node and content service node.

The structure of Contents Subnet Load Balancing (CSLB) element in CSM illustrates in Figure 3.

Subnet Monitoring Element (SME) gathers information of the subnet by monitoring the network nodes, interfaces and links in the subnet. It generates network traffic load status information as well as content server load status. And also SME produce content routing information base by monitoring on content location over the network.

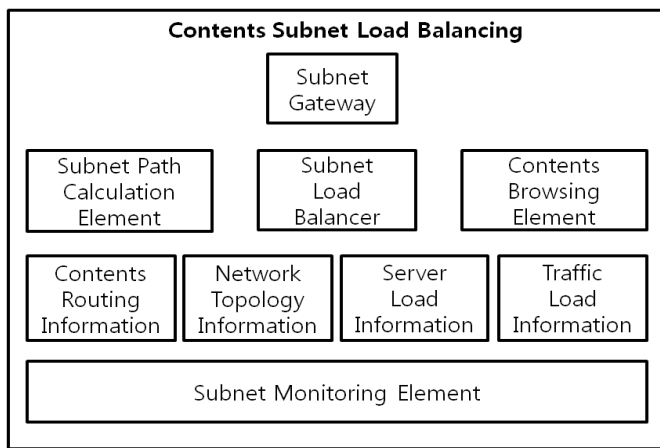


Figure 3. Content Subnet Load Balancing System Structure

Confirmation process to decide whether the requested content belongs to its subnet is performing at the Content Browsing Element. Subnet Path Calculation Element computes optimal path from the content service node to the user node using collected information: Network Topology Information, Server Load Information and Traffic Load Information.

Subnet Load Balancer controls CSLB procedures after receiving the content service request through Subnet Gateway. Firstly it checks whether its subnet contains the requested content or not with Content Browsing Element and if so, it computes the path using Subnet Path Calculation Element.

Finally Subnet Load Balancer returns the optimal content service node to the Subnet Gateway. Preparing for the case that the requested content was not in the corresponding subnet user resides Subnet Gateway notify Subnet Information Gateway in the CNLB of the abstract information of the subnet.

III. HIERARCHICAL LOAD BALANCING

A. Optimal Subnet Selection

The optimal subnet selection procedure shows in Figure 4. First of all CNLB confirms if the subnet that a user resides has the content that user wants to get. If exists there is no need to access other subnet deliberately. Because the content network is composed of several subnets duplicate data could be spread over the entire network.

After the CNLB selects the subnet it sends the identification number of the subnet to the Content Server in the CNM.

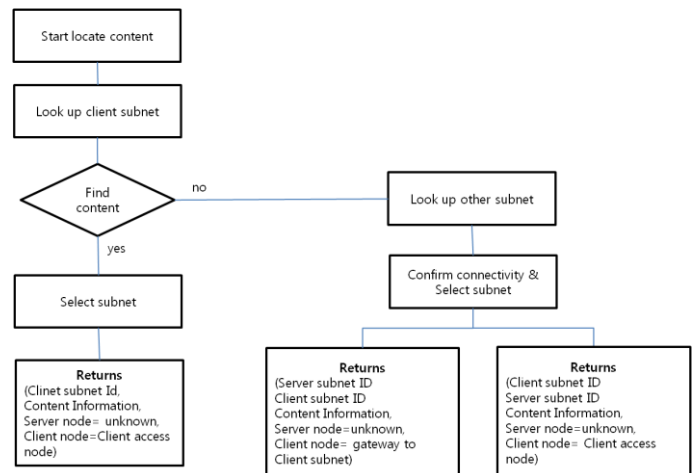


Figure 4. Optimal Subnet Selection Procedure

Then Content Router in the corresponding subnet starts to find the content service node according to the request of the Content Server with the content information and IP address of the client.

B. Optimal Content Node Selection

As a response to the request Content Router inquires the CSLB to find the content service node available to offer the service to the client. CSLB uses node information table and real-time streaming flow table keeping content routing information and network topology maps. Figure 5 illustrates the procedure to select optimal content service node.

With the information provided candidate nodes providing the requested content will be extracted. An optimal node will be chosen according to the costs for connecting the server nodes with the client node when multiple content service nodes are qualified. Content server node as well as cache node could be given selection as a content service node in the subnet.

For the general content service such as massive file transfer or VoD services content service node will be served using the network caching capability. But, in case of real-time streaming more information should be required to serve streaming service to the client.

Real-time streaming service has a feature that it requires to convey the real-time content flow to the destination with minimum delay without saving or caching through the network.

Real-time content request from the client node look up the node information table to distinguish real-time streaming server node in the subnet.

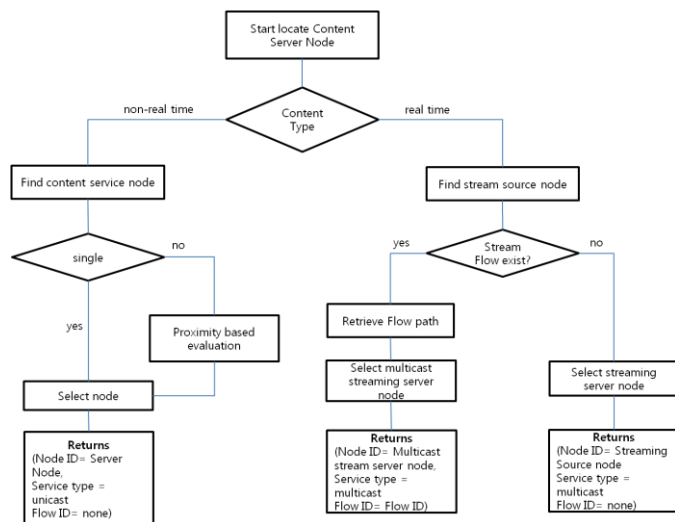


Figure 5. Optimal Content Service Node Selection Procedure

The available content service node among the corresponding real-time streaming flow that capable of streaming with a higher proximity to the client will be selected for real-time streaming service.

With the address of the content service node selected finally, client could make a reliable connection to receive the content from that network.

However, the requested content was not found within the subnet which the requested client connected with CNLB will check whether other subnets own the requested content or not.

The next step to enable inter-subnet connection for the content is to confirm the connectivity between the client subnet where the client sits and the server subnet where the content belongs to.

Upon holding a reliable connection between the subnets in terms of traffic overload and network topology Content Server requests the path provisioning over the server subnet from the content service node in that subnet to the gateway node interconnecting the server subnet with the client subnet.

In the client subnet, a path from the gateway node connected to the server subnet to the client node should be established to provide the content from another subnet to the client.

For the effective content network provisioning client network manager have to determine whether the client subnet stores the corresponding content as its own content while the content transit cache node or server node in the client subnet. And also the optimal cache node should be on the path that the

content should drop by in terms of operator's policy and network optimization.

IV. CONCLUSION

A rapid growth of internet traffic is mainly attributed the various forms of content especially video (TV, VoD, Internet Video and P2P) that are expected to exceed 91 percent of global consumer traffic by 2014. Therefore the foremost service that expected to be introduced to the real world with smart service network might be a content networking service in the near future [6].

Up until now the increasing demand for massively distribution and replication of large amounts of resources has led to two main developments: P2P and CDNs. But there are still some issues such as P2P peer selection that often cost network provider of inter-provider traffic as currently addressed by the ALTO working group and traffic overhead for both P2P and CDN scenarios currently being dealt with by the DECADE WG [7-8].

We introduced distributed network equipped with storage and streaming servers within the network and proposed its hierarchical load balancing structure for the smart content service.

If we could take advantage of both the network knowledge and the information of servers and storage distributed over the network, the service network will be the more competent for the upcoming content explosion era.

REFERENCES

- [1] Diana Smetters and Van Jacobson, "Securing Network Content", October, 2009, PARC Technical Report.
- [2] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard. Networking Named Content. In CoNEXT'09, Rome, Italy, Dec. 2009.
- [3] Ahlgren, B.; Aranda, P.A.; Chemouil, P.; Oueslati, S.; Correia, L.M.; Karl, H.; Söllner, M.; Welin, A.; Content, connectivity, and cloud: ingredients for the network of the future, IEEE Communications Magazine, Volume: 49, July 2011
- [4] J. Ash, Ed., J. L. Roux, and Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements," IETF, RFC 4657, Sep. 2006.
- [5] E. Oki, J.-L. L. Roux, and A. Farrel, "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering," draft-ietf-pce-interlayer-frwk-05, work in progress, IETF, Sep. 2007.
- [6] Gu, Y., Zong, N., Zhang, H., Zhang, Y., Lei, J., Camarillo, G., and Y. Liu, "Survey of P2P Streaming Applications", draft-gu-ppsp-survey-02 (work in progress), October 2010.
- [7] Alimi, R., Penno, R., and Y. Yang, "ALTO Protocol", draft-ietf-alto-protocol-06 (work in progress), October 2010.
- [8] D. Kutscher, M. Stiernerling and J Seedorf, "Design Considerations for a DECADE SDT", draft-kutscher-decade-protocol-00(work in progress), October 2011.