J. Basic. Appl. Sci. Res., 4(1)160-165, 2014
© 2014, TextRoad Publication

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Content Searching Scheme with Distributed Data Processing Service in Content Centric Networking

Muhammad Zafrullah*, Muhammad Roman and Muhammad Khawar Islam*****

*Professor of Electrical Engineering, UET Taxila – 47050, Pakistan
** MSc Student, Department of Electrical Engineering, UET Taxila – 47050, Pakistan
*** Director, Faculty of Engineering, University of Gujrat, Gujrat, Pakistan

Received: September 29 2013
Accepted: November 22 2013

ABSTRACT

Content Centric Networking is a recently proposed internet communication paradigm which is based on content instead of the physical location which holds the contents. People are interested in data only. The physical location or IP address of this data is not of their concern. But internet does not know anything about data. It only knows addresses. This new content based architecture promises solutions to all existing internet problems. In this work we study content searching in content centric networking, combining it with a Distributed Data Processing Service. We argue that data can be processed by network instead of servers or clients to improve the overall experience of data retrieval. We implement a service in network which extracts an audio file from a video file on 'audio file request'. The main focus of this scheme is to reduce unnecessary data traffic to improve download efficiency. Implementation results of this scheme show that content searching and retrieval can be made more efficient with Distributed Data Processing Service.

KEYWORDS— Quality of Service, Distributed Data Processing, Content Store, Pending Interest Table, Forwarding Information Base, Interest Packet and Data Packet

I. INTRODUCTION

Networking started with the birth of Telephone. Later internet system was developed as an overlay on telephone architecture. The idea was to share the expansive machines like printers and tape drives. So communication model was designed on location address or host abstraction. In this location based network the request from one machine accesses the resources and other machines provide access. Since then networking purpose has shifted to media distribution but model is still based on physical locations. In the current scenario machines are cheaper and disposable commodities, while ultimate goal of internet is content retrieval. There are various techniques proposed by research community like "*Data Oriented Network Architecture* (DONA)" [1], "*Publish/Subscribe Internet Routing Paradigm* (PSIRP)" [2], and "*Content Centric Networking* (CCN)" [3], which are discussed as future internet models. These proposed models have different Naming and Security mechanisms, Routing methodologies and Transport and Caching schemes but all these models are based on named data instead of named locations. CCN proposed by "*Palo Alto Research Centre* (PARC)" [4] is a widely accepted next generation networking model. It can be implemented over the existing physical location based TCP/IP networking model. Two main features of CCN model are '*named data*' and '*distributed data storage*' over the entire network [3]. Explicit and self-explanatory naming is used for routing instead of IP addresses [5]. However this new named data based address space requires some content management and routing issues [6, 7].

According to IBM studies, 90% share of total contents in the world has been produced in 2011 and 2012 [10]. A research work carried out by Cisco indicates that more than 12 Exabytes/month data is moved by internet [11]. Although users are interested in contents, this TCP/IP model is still based on physical locations. This mismatch of technology and its application results into various issues like complex configuration, poor security, no user based control and quality of service (QoS) problems. In the application layer users express their interest in contents but in network layer it has to be mapped with locations. This translation needs a junk of middleware. In CCN, both application and network layers work on contents so it makes configuration simpler [12]. Internet knows nothing about contents. It only knows containers. So instead of securing contents it tries to secure containers and wires. On the other hands, CCN knows about contents so it can secure contents by signing it [13]. CCN enables user based control by providing two way local communications. It also solves QoS problems of current internet. CCN is also more flexible in terms of data storage as data can be stored on all nodes. Whenever a node forwards data it can keep a copy of that data to facilitate future requests for the same data. Here routing is done on content names which are hierarchical and of variable lengths. There are three fundamental components in CCN forwarding. Content Store (CS) is a local storage of each node. Pending Interest Table (PIT) is a table which keeps record of all requests sent upstream for data. Forwarding Information Base (FIB) is similar to IP Routing table but unlike IP it can have multiple faces for a single data request.

In the basic scheme of content searching a content request is generated and broadcasted over the available interfaces [3]. Nodes having requested contents respond with data. In this process every intermediate node keeps a copy of data for future use. This content based routing with distributed storage reduces latency and delay [8]. In this work we combine content searching with a data processing service (DPS) which can be distributed over various nodes in the network. In this scheme content is processed according to user's demand before its retrieval. As an example we implement a service which converts an *avi video file* to an *mp3 audio file*. We implement it on a linear topology using CCNx prototype [9].

***Corresponding Author:** Muhammad Zafrullah, Professor of Electrical Engineering, UET Taxila – 47050, Pakistan
Ph: 0092 51 99047545, Fax: 0092 51 99047545, dr.zafrullah@uettaxila.edu.pk

Implementation results prove that content searching is more efficient when combined with Data Processing Service.
Section II describes problem statement and motivation of this research work. The basic content searching scheme is discussed in **section III** with its merits and demerits. Detailed design of proposed scheme and its experimental setup are explained in **section IV**. CCNx prototype is also illustrated in this section. Implementation steps are defined and test examples are performed to prove the concept. Results obtained from experiments are presented and discussed in **section V**. Finally **section VI** concludes the paper.

II. Problem Statement and Motivation

Content Centric Networking (CCN) is considered as a replacement of existing TCP/IP networking model. In various applications, basic content searching scheme of CCN produces unnecessary traffic and overhead which reduces download efficiency. In this work we combine content searching scheme with distributed data processing service to improve overall download experience by reducing unnecessary data traffic.

As CCN is a newly proposed architecture, it badly needs efficient routing and content searching techniques to achieve desired results. Our proposed scheme can be deployed in the practical implementation of CCN architecture as it improves the efficiency of content searching.

III. Content Searching in CCN

In Basic Content Searching scheme the request is generated and forwarded to other nodes. When a node receives a request it looks into its local store for data. If data is found it is retrieved otherwise it looks into PIT. If it matches with some PIT entry it means another node has already requested same content and it has been forwarded to upstream nodes. So this request is also added to PIT. If nothing found in PIT it looks into FIB. If some interface with requested content information is found, request is routed to that node otherwise search is failed or request is flooded. The same is shown in Fig. 1.

This basic scheme reduces traffic but increases overhead in terms of content management and storage maintenance. Therefore a keyword based searching scheme is proposed by constructing an overlay of content providers over network. This scheme reduces size of table but a considerable time is consumed in overlay network deployment. There is also no way defined to control traffic towards content providers [14]. A ranking model is also proposed to isolate malicious items of data to increase the reliability of content searching [15].

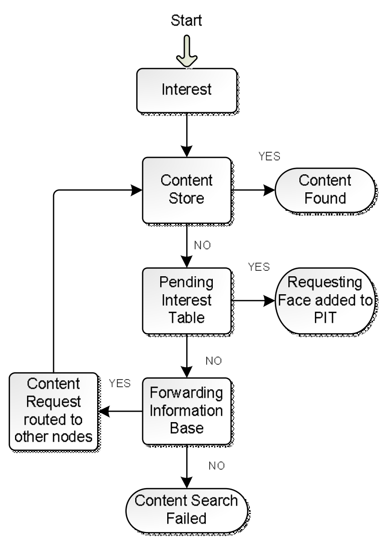

Fig. 1: *Content Searching in CCN*

IV. Content Searching with Distributed Data Processing Service

Content Searching can be combined with Data Processing Service. This DPS can be implemented on servers or on intermediate nodes. To implement it on servers we will have to modify all servers, which is not feasible. So this is implemented on intermediate nodes and distributed over network.

A. Experimental Setup

- A “Content Centric Network” consisting of a linear topology of three systems, content provider, middle node and client
- Linux based ubuntu 12.04 LTS running as OS on each node [16]
- CCNx-0.7.0 installed and running on Linux machine
- DPS implemented in CCNx prototype and this modified CCNx running on middle node

B. CCNx Prototype

CCNx is a prototype implementation to test the concepts of CCN. This prototype is a transport protocol which runs on TCP/IP and uses UDP for transport and its default port is 9695. Some of the related terms related to CCNx are *Interest Packet* (carries data request), *Data Packet* (retrieves data), *ccnd* (manages all connectivity and forwarding) and *Repository* (local storage of CCNx in which data is cached).

C. Test Examples

In these test examples we implemented a service which converts an '*avi video file*' to an '*mp3 audio file*' on client's request. File is retrieved from content provider, processed at middle node and forwarded to client.

Case 1: Retrieving Content without Data Processing Service

Block diagram of this scheme is shown in Fig. 2. Following are the steps involved.

- Client requests an 'avi file' from content provider
- An interest message is created and forwarded to 'ccnd' of middle node
- 'ccnd' of middle node checks its repository
- When file is not found it forwards the request to 'ccnd' of content provider
- 'ccnd' of content provider gets 'avi file' and sends it to 'ccnd' of middle node
- Middle node keeps a copy of file in its content store and sends original file to 'ccnd' of client
- 'ccnd' of client delivers it to client

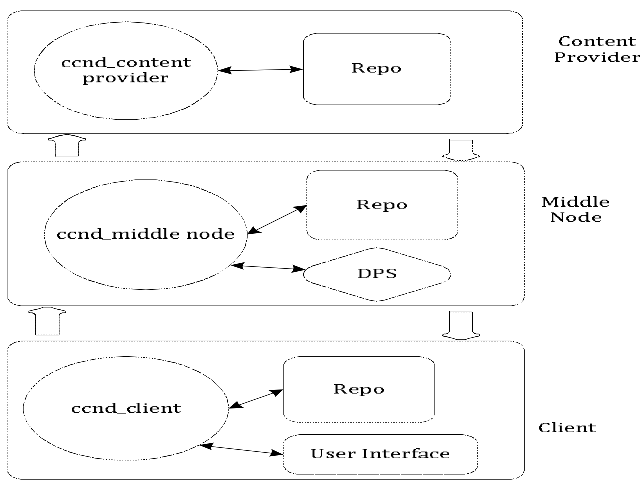

Fig. 2: *Content Searching Combined with Data Processing Service*

Case 2: Retrieving Content with Data Processing Service

Block diagram of this scheme is shown in Fig. 3. Following are the steps involved.

- Now client requests DPS for the retrieval of mp3 format of same file
- 'ccnd' of client sends DPS request to ccnd of middle node
- 'ccnd' of middle node checks its repository for file
- As file was cached in content store it is found
- 'ccnd' invokes DPS to provide requested format
- DPS converts the file from avi to mp3 and uploads it to repository
- 'ccnd' gets requested file from repository and forwards it to client

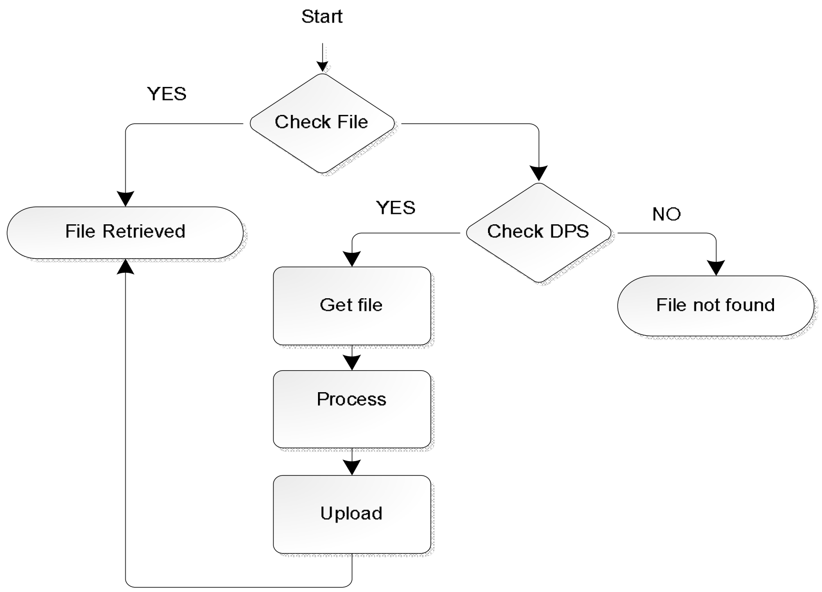

Fig. 3: *CCN Repository Process with DPS*

Case 3: Retrieving file when it is already processed and cached in middle node
This is another problem situation, regarding the data format. If, during some previous retrieval, the file is already processed, it is available in the mp3 format. Now DPS does not have to convert it as converted file was also cached. So it is retrieved without any processing.

V. RESULTS AND DISCUSSIONS

Two tests of content retrieval were performed. In first test, an 18 MB avi video file was processed and downloaded. Second test was performed with a 12 MB avi file. Download time and file size were measured through *Ubuntu* terminal. Decrease in download time proves the efficiency of this scheme. Table 1 summarize the results of processing 18 MB video file and download time for processing 12 MB video file are shown in Table 2. The same information is presented in Fig. 4 and Fig. 5 respectively. Comparison in download time in all the three cases shows the advantage of the proposed scheme with data processing service.

TABLE I: *Processing 18 MB video file*

Case	File size	Download Time
1	18 MB	90 seconds
2	6 MB	30 seconds
3	6 MB	5 seconds

TABLE II: *Processing 12 MB video file*

Case	File size	Download Time
1	12 MB	65 seconds
2	4 MB	20 seconds
3	4 MB	4 seconds

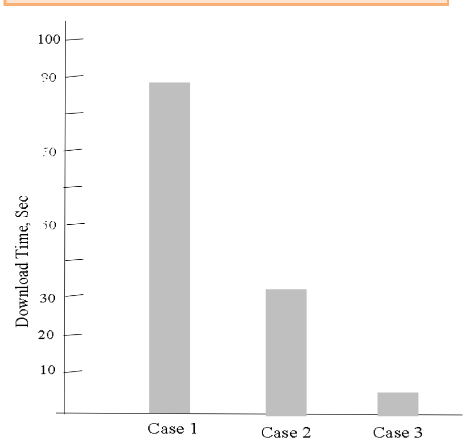

Fig. 4: *Download Time of different test cases given in Table 1*

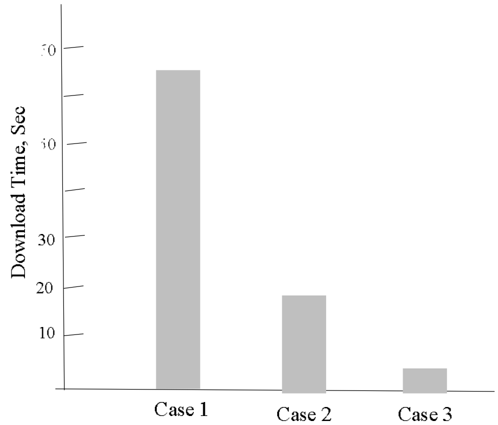

Fig. 5: *Download Time of different test cases given in Table 2*

In the 1st test case of example 1, a client requests an 18 MB avi video file. Middle node does not have requested file so it retrieves file from upper node which is content provider. As per cache policy of CCN, middle node keeps a copy of that file in its local storage. As file is retrieved from top node it is downloaded in 90 seconds at client.

In the 2nd test case, client requests mp3 format of same file. As original file was cached at middle node, it does not have to retrieve file from content provider. In this case middle node converts 18 MB avi video file to 6 MB mp3 audio file and sends mp3 file to client. Middle node also keeps a copy of this processed file in its local cache. This file is retrieved in 30 seconds.

In the 3rd test case, another client requests same mp3 file. Now middle node does not have to retrieve original file from content provider or to process it as processed file was also stored at middle node. It just delivers requested file to client. File is downloaded in 5 seconds. We can see file size is 6 MB in 2nd and 3rd test cases but download time is 30 seconds in 2nd case and 5 seconds in 3rd case. Because in 2nd test case, file is processed prior to its retrieval and time is consumed in processing. But in 3rd case, requested file is just retrieved without any processing.

These test cases explain how content retrieval process is improved when combined with Data Processing Service. Download time is reduced by reducing unnecessary network traffic and latency. Network is more intelligent as it can provide multiple formats of a file in the locality of a user.

VI. CONCLUSION

By deploying Data Processing Service, network traffic can be reduced by a significant amount. Reduction in traffic reduces delay which in turn makes data delivery efficient in CCN. If data processing service is not distributed over the network and client has to process it after content retrieval, the original large size file will be transferred to each client. Instead if DPS is deployed at various nodes, clients will retrieve an already processed file. Thus to improve the efficiency of content searching, it can be combined with a distributed Data Processing Service.

DPS is not about converting file formats or to extract audio file from a video file. The example (video to audio conversion) we discussed is just a test case to prove the concept of Data Processing Service. We can benefit from DPS concept in various ways. It makes the network flexible in terms of availability of different file formats in the vicinity of a user. If multiple formats of a file are available in the vicinity of a user, content requests are not routed to upper nodes in the network. Instead they are facilitated from local nodes. This local communication reduces delay, congestion and improves delivery efficiency and Quality of Service. DPS concept can be applied to audio and video streaming. High definition video streaming makes a major portion of today's network traffic. If local users can provide HD streaming, it can reduce network traffic significantly. DPS concept can also be extended to advanced services like online banking, online shopping and social networking.

VII. Future Recommendations

As described earlier these tests are performed to prove the concept of content searching combined with data processing service. This concept can be extended to advanced services to improve the overall experience of data retrieval.

Bandwidth, Latency and other network parameters are not considered in these examples as tests are performed on a linear topology consisting of three nodes. Experiments can be performed on a large network considering the effect of these network parameters.

Acknowledgment

The authors declare that they have no conflicts of interest in this research.

REFERENCES

[1] T. Koponen, M. Chawla, B.G. Chun, A. Ermolinskiy, K.H. Kim, S. Shenker and I. Stoica, “A data-oriented (and beyond) network architecture,” *Proc. ACM SIGCOMM*, Aug. 27-31, 2007, Kyoto, Japan.

[2] N. Fotiou, G..C Polyzos and D. Trossen, “Illustrating a publish-subscribe internet architecture,” 2nd Euro-NF Workshop on Future Internet Architectures and New Trends in Network and Services Architectures, 2009.

[3] V. Jacobson, D.K. Smetters, J.D. Thornton, M.F. Plass, N.H. Briggs and R. L. Braynard, “Networking named content,” *CoNEXT ’09: Proceedings of the 5th international conference on Emergingnetworking experiments and technologies*. New York, NY, USA: ACM, 2009, pp. 1–12.

[4] http://www.parc.com/

[5] Y. Barouni, M. Bicudo and P. Spathis, "Content Centric Routing for Future Generation Networks," *IEEE INFOCOM Workshops 2009*, pp.1-2, 19-25 April 2009.

[6] D. Perino and M. Varvello, “A reality check for content centric networking,” *ACM Sigcomm Workshop on Information-Centric Networking* (ICN) 2011.

[7] M. Tortelli, I. Cianci, L.A. Grieco, G. Boggia and P. Camarda, "A fairness analysis of content centric networks," *International Conference on the Network of the Future (NOF)*, 2011, pp.117-121, 28-30 Nov. 2011.

[8] A. Carzaniga, M.J. Rutherford and A.L. Wolf, "A routing scheme for content-based networking," *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, pp. 918- 928 vol.2, 7-11 March 2004.

[9] Project CCNx. Available online at http://www.ccnx.org/

[10] www.ibm.com/software/data/bigdata

[11] Cisco Visual Networking Index: Forecast and Methodology, http://www.cisco.com

[12] A. Detti, and N. Blefari-Melazzi, "Network layer solutions for a content-centric internet," *Trustworthy Internet*, Springer Milan, Jan. 2011.

[13] D. Smetters and V. Jacobson, "Securing network content," RelatórioTécnico TR-2009-1, Xerox Palo Alto Research Center-PARC (2009).

[14] R. Haw, C.S. Hong, B. Kim and H.Y. Song, "Content searching scheme using interesting keyword based overlay network," *13th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 2011 , 21-23 Sept. 2011

[15] N. Fotiiou, G.F. Marias and G.C. Polyzos, "Information ranking in content-centric networks." *IEEE Future Network and Mobile Summit*, 2010.

[16] http://www.ubuntu.com/