DBpia

논문 17-42-04-10 The Journal of Korean Institute of Communications and Information Sciences '17-04 Vol.42 No.04
https://doi.org/10.7840/kics.2017.42.4.780

Federated Named Data Networking Testbed for Climate Science

Alexander Ni◆, Huhnkuk Lim°

ABSTRACT

Data discovery and distribution application that is utilized by climate, high energy physics, and other scientific communities are experiencing performance and large scale data managing problems, that are rooted from the shortcomings of IP architecture. To solve this problem, newly developed data managing applications based on NDN architecture were introduced. In this letter, we present the federated NDN testbed with an NDN-based climate science application and the set of experiments that reflect the performance of NDN based climate application in general with determined and applied optimization.

Key Words : NDN, named data, networking, climate applications, data-intensive science

Ⅰ. Introduction

Applications for data-intensive science are dealing with large volumes of scientific data such as climate data and high energy physics (HEP) data and both of them have existing systems for data management and distribution based on current internet architecture. Using the systems based on IP networking technology, they have a bottle neck in their performances such as user latency and corrupted ratio. For the instance, high energy physics (HEP) and climate science applications are already facing the issues, resulting from ESGF and Xrootd system that are currently used for data management in current IP networks. By being based on the host to host transport control protocol for data fetching, both applications have inefficiencies like high latency and corrupted ratio. The continually increasing scale of high volumes of scientific data in big science is driving the idea that new innovative data management solution should be found.

Named Data Networking (NDN)[5] is one of the promising future internet architectures that contains key features as hierarchical naming; in-network security and caching, and name based symmetrical forwarding of Interest and Data packets. That focus network on data name, not the data location that is used for current IP networking. All the packets are signed with certain network key, before being sent, that prevents unauthorized use of those packets. Packets can be used only by the user which is authorized by the network.

In this letter we establish the federated NDN testbed for climate science and test its performance, using an NDN based climate science application. The established testbed was connected between Korean' NDN testbed (KISTI domain)[1] and US' NDN testbed (CSU domain) for climate science. By using the federated NDN testbed for climate science application we make several performance tests to identify optimal configuration with good performance.

Ⅱ. Related Work

Colorado State University (CSU) is developing their own science application (ndn-atmos)[3] for data discovery and data transfer based on NDN. For now,

※ This research was supported by the Research and Development of NDN SW Platform for Data-intensive Science Program, 2015 funded by the Korea Institute of Science and Technology Information (KISTI) (K15-L05-C02-S01).
◆ First Author : University of Science and Technology (UST), alexni1992@gmail.com, 학생회원
° Corresponding Author : Korean Institute of Science and Technology Information (KISTI), hklim@kisti.re.kr, 정회원
논문번호 : KICS2017-03-062, Received March 3, 2017; Revised March 22, 2017; Accepted April 13, 2017

www.dbpia.co.kr

their application contains CMIP5 and HEP sub modules, and they are actively upgrading this sub-modules with a support of science communities. They also have a testbed which was built on ESnet nodes. CSU testbed is used to test their application and for its future serving inside of large-scale scientific networks. Several experiments were done by CSU, in order to represent performance of NDN architecture and ndn-atmos application. Other scientific approaches are going in NDN research community. One is at the Imperial College of London and the other one in California Technological Institute. Their main interest is focusing on HEP science. However, none of these represent a federated NDN testbed for data-intensive science.

Ⅲ. Federated NDN Testbed for Climate Science

By using existing NDN platform (NFD, ndn-cxx), NLSR and NDN based climate science application. Several domains currently operating in the US and linked by 10G connection with each other. The NDN testbed network consists of 7 nodes that are located in different sites in US and together it is a group of several catalogs and data repositories. One NDN node in our domain has been connected with the NDN testbed network for climate science in the US by the 1G connection. Climate science data files between continents can be searched and fetched by consumers based on NDN names.

3.1 NDN based climate science application

Ndn-atmos[2] application consists of Catalog, User Interface, and Data Storage. Catalog is service point that supports data discovery inside of the NDN network. To prevent faults and increase a reliability of the system, several catalogs are combined in a group, where all catalogs can exchange their information concerning to changes inside of the data lists like newly published or deleted data. Synchronization is supported by ChronoSync application and with the use of the sync messages, it propagates update information to other catalogs. Catalogs synchronization help to achieve information similarity among all catalogs to provide an ability to see same data set disregard of the location to access the NDN network. Furthermore, by using various forwarding strategies network can decide to send queries to nearest catalog or to the least congested one[4].

UI was created to help users to discover and to retrieve a published data. Web interface supported with various filtering tools for data discovery which makes it easier to be found by users. UI also support additional searching ways. First, is searching by a prefix name and at the same time system will automatically propose for user possible prefix segments that can be typed. The second one is a well-known tree search that is used in most operation systems and was presented here as a searching way.

Data storage that is used in the ndn-atmos application is a persistent NDN based data storage. Service consist of several data storages with a different location, such structure will improve the speed of data delivery and exclude the problem of single point of failure.

3.2 Establishment of a federated NDN testbed between the US and Korea

To create KISTI climate application domain we have configured ndn-atmos catalog and data repository on one of our nodes. Then it was connected to the existing catalog group in the US that also has its own data storages. It was established with the use of the NDN link-state routing (NLSR)[6] protocol. That protocol together with ChronoSync application support consistency of content of domestic catalog and catalogs located in the US. The schema presented in Figure 1 explain the performance test procedure from event of data publishing and to its actual retrieval. In order to add a new content publisher request a publication from the catalog. After that catalog store a new data and update dataset names between all other catalogs. From that point potential consumer will be able to search for the data by asking it from one the several synchronizes catalogs and then retrieving the list of published data. The user selects a certain dataset and

www.dbpia.co.kr

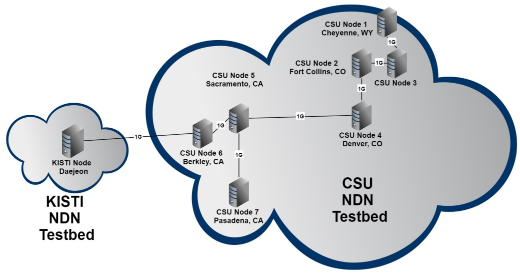

Fig. 1. CSU and KISTI federated testbed

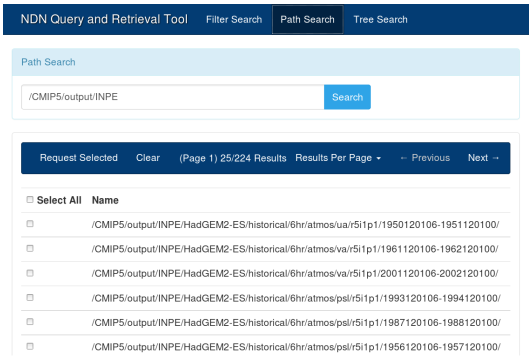

Fig. 2. Atmos application UI, installed on KISTI node, with result list of published data.

retrieves it from one of the data storages. If a connection to data storage was lost in a time of retrieval download will continue from other storage that is also containing this data.

To support intra-connection between sites of the testbed we are using NLSR protocol. Which will propagate published prefixes to the neighboring nodes of the specific node, when a ChronoSync application will keep them up to date by sending periodic synchronization messages between all nodes. At the same time, NLSR provides hierarchical data signing function and hierarchical key exchange. Hierarchical key signing enhances the trust to the interest and data packets inside the network and will prevent malicious data appearance.

Ⅳ. Performance Result

Performance test result was done by making four delivery tests between four consumers and one remote data repositories. Supposed user by utilizing the UI of ndn-atmos application transfer some certain dataset from the given remote repository and to one on his side. But even if the first delivery test is done from the actual producer the next three were delivered from the intermediate router where data cached after a first delivery test. Furthermore, throughput measurements of data transferred between one consumer and producer are presented depending on the configuration of NDN platform.

Producer node of KISTI domain serve the data with use of UI which is shown in Figure 2. In performed delivery tests we have published more than multiple NDN based climate dataset names with actual data. Several data sets can be chosen in one time and fetched by the user to be delivered to one of the existing data storages.

Figure 3 is a representation of performance test between one CSU producer and one local consumer. At the same time, after subdividing result from Figure 3 it is possible to see two different outputs as remote fetching from the producer and from the intermediate router that shown in Figure 4 and 5 respectively. Both local and remote figures were received from the division of result from Figure 3. All three graphs based on 4 different pipeline sizes and 10 packet sizes, where a pipeline size is a number of interest packets that will be sent without receiving back data packet and packets size is a maximum transferring unit (MTU) of NDN. In each figure, we can see throughput growing with the increase of packet size and pipeline size values. That point testifies the ability to control and optimize the data fetching throughput. By considering results from three figures we have estimated that 2MB packets size with pipeline size equal 4 was a most suitable

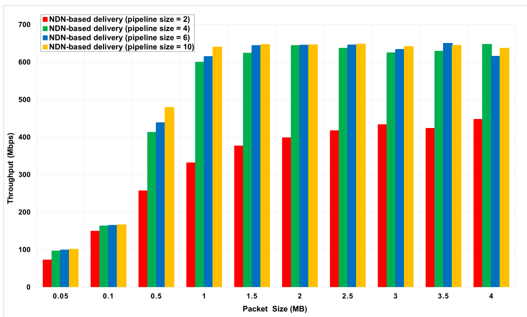

Fig. 3. Average throughput of local and remote fetching

www.dbpia.co.kr

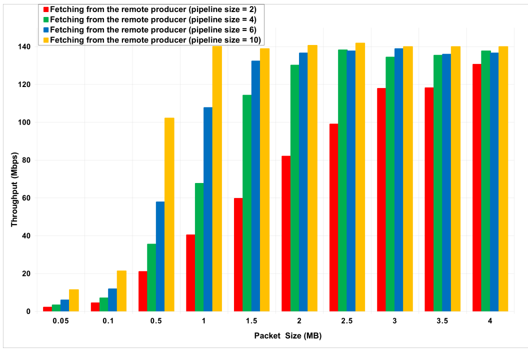

Fig. 4. Data fetching throughput from a remote host

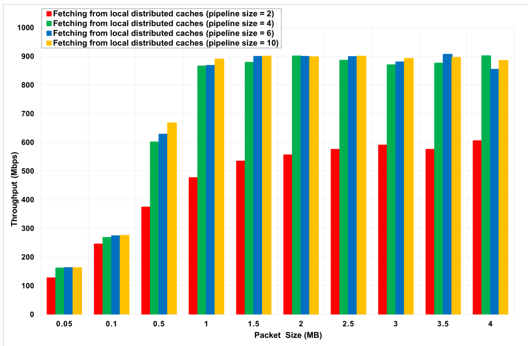

Fig. 5. Data fetching throughput from a middle node cache

for getting the highest throughput from US node. Optimal packet and pipeline sizes helped to achieve maximum throughput efficiency. At the same time, multiple delivery tests that were performed between server and clients create the caching effect that can be seen in Figure 5 and increased throughput that is nearly six times exceeds the throughput from the remote host (Figure 4) and utilize almost full 1G link speed.

V. Conclusion

The federated NDN testbed is a representation of established connection of our domestic and foreign NDN testbeds for data-intensive science. The NDN testbed consists of our 11 nodes and CSU 7 sites linked by 1G connections. And both of this testbeds are utilizing common NDN based climate application to create a new perspective for its performance testing between two highly distant locations. And the results presented in this letter show 6-time increase in actual throughput between producer and consumer nodes that were archived due to the caching effect of NDN architecture on data delivery performance. Moreover, the performance test helps to identify the most suitable packet size and pipeline size for the data delivery in a federated testbed scale. Such optimization helped us to significantly increase the throughput for remote and local data fetchings. By performing the test with several different packet and pipeline sizes we were able to identify the most efficient packet and pipeline size.

References

[1] A. Ni and H. Lim, “A named data networking testbed with global NDN connection,” *J. KICS*, vol. 40, no. 12, pp. 2419-2426, 2015.

[2] C. Olschanowsky, S. Shannigrahi, and C. Papadopoulos, “Supporting climate research using named data networking,” in *IEEE LANMAN*, pp. 1-6, Reno, NV, USA, May 2014.

[3] S. Shannigrahi, et al., “Named data networking in climate research and HEP applications,” *J. Physics*: *Conf. Series*, vol. 664, no. 5, p. 052033, 2015.

[4] C. Fan, et al., “Managing scientific data with named data networking,” in *Proc. Int. Wksp. Network-Aware Data Management*, p. 1, ACM, Nov. 2015.

[5] A. Afanasyev, et al., “Named data networking,” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 66-73, 2014.

[6] V. Lehman, A. M. Hoque, Y. Yu, L. Wang, B. Zhang, and L. Zhang, “A Secure Link State Routing Protocol for NDN,” NDN, Technical Report NDN-0037, 2016.

www.dbpia.co.kr

Alexander Ni

May 2014 year Graduated from International Information Technology University (IITU), Almaty, Kazakhstan

From March 2015~present Master degree in University of Science and Techonology (UST), Daejeon, Korea

<Interests> Named Data Networks (NDN)

임 헌 국 (Huhnkuk Lim)

1999년 2월 : 항공대학교 전자공학과 졸업

2001년 2월 : 광주과학기술원 정보통신공학과 석사

2006년 2월 : 광주과학기술원 정보통신공학과 박사

2006년 3월~현재 : 한국과학기술정보연구원 선임연구원

<관심분야> 미래인터넷 구조, 정보중심 네트워킹 (ICN), Named data Networking (NDN), 네트워크 자원제어, 네트워크 가상화

www.dbpia.co.kr