Content-Oriented Networking as a Future Internet Infrastructure: Concepts, Strengths, and Application Scenarios

Kideok Cho, Jaeyoung Choi, Dong-il Diko Ko, Taekyoung Kwon, Yanghee Choi
School of Computer Science and Engineering
Seoul National University, Seoul, Korea
{kdcho, jychoi, diko}@mmlab.snu.ac.kr, {tkkwon, yhchoi}@snu.ac.kr

Introduction

While the current Internet architecture is designed based on the host-to-host communication paradigm, the real Internet usage shows that content-oriented traffic is dominant. Good examples of this paradigm shifting are Content Delivery Networks (CDNs) and Peer-to-Peer (P2P) Networks. A CDN is a system of cooperating networked computers and it provides content delivery services to end users transparently hiding its network architecture. P2P is a cooperative network constructed by users' participations for file sharing. In these networks, users do care about not from whom they retrieve data but the data itself.

A recent report analyzes that most of the Internet traffic is dominated by P2P traffic which does not conform to the host-to-host paradigm [1]. Due to the inconsistency between the Internet design and the real usage, there is an unnecessary indirection overhead when users try to retrieve desired data [2]. In consequence, it is needed to redesign the Future Internet based on the content-centric paradigm to provide data/contents to the users efficiently. In this paper, we define the concepts of content-oriented (networking) paradigm. Also, strengths and application scenarios of content-oriented networks (CONs) will be investigated.

This paper is organized as follows. First, we describe the concepts of CONs. And then we discuss strengths and application scenarios of CONs. We also introduce some current research activities on the CONs.

Concepts of Content-Oriented Networks

In this section, we first define what a CON is by comparing a CON with the Internet. A CON is a network based on the content-oriented paradigm. The content-oriented paradigm is different from the host-to-host paradigm in that it focuses not on the communication party but on the data itself. What users have to do in the CON is just specifying which data they need. Then the CON finds the desired data and forwards it to the users efficiently. So, unlike the traditional Internet in which the communication party is important, it does not matter from whom users get the desired data in the CON. Rather, the content (or data) itself is more important.

Because users do not care from whom they retrieve the desired data, the CON should provide efficient ways to retrieve the data from one of the candidate sources. So, we assume that CON routers[1] have storages/caches for the contents to provide fast and efficient retrieval of contents. Also, the CON itself should provide ways to validate data which they provide to users; for example, it is possible to validate the content ID (CID) of data through MD5 hashing.

There are many terms referring a network based on the content-oriented paradigm: *content-oriented, content-centric, content-based, data-oriented, or data-centric network*. We insist that these concepts be equivalent in that they focus on not the communication party but the content or data itself. So, in this paper, we use the term content-oriented network (CON) to refer a network based on the content-oriented paradigm.

Strengths of CONs

In this section, we investigate the major strengths of CONs. Compared with the Internet, CONs can provide 1) pervasive experience, 2) performance improvement, 3) flexible (or extensible) dissemination. We discuss these issues as follows.

Pervasive Experience

CONs can provide a pervasive experience in terms of content access. In the CON environment, the right for the content is decoupled with the content itself. So, we assume that users obtain the right for the copyrighted file by any other means. For example, users can acquire rights for the

[1] We refer to the network elements in the CON as CON routers for the sake of convenience.

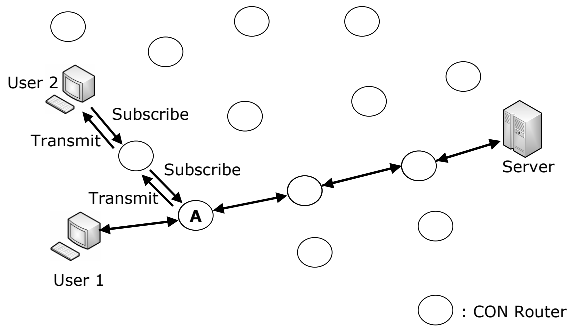

Figure 1. Multicasting in a CON

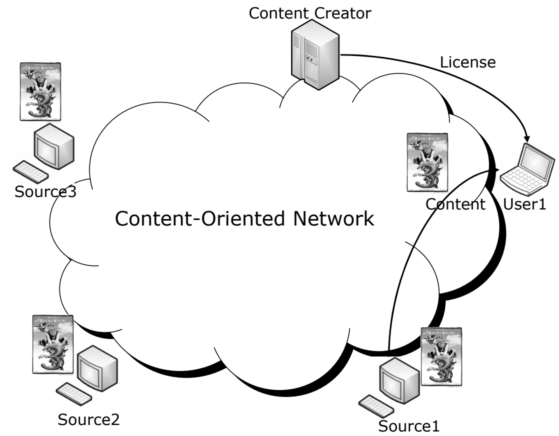

Figure 2. Application Scenario in a CON

contents through the Digital Right Management (DRM) [3] or in the offline shop. Then CONs can be considered as a huge storage for the contents. Users can access the contents they have purchased through the CON even though the contents are not stored in a local system. In consequence, in the CON environment, users can access contents anytime, anywhere and with any device.

Performance Improvement

In the CON environment, network efficiency will be substantially improved by retrieving the contents from the closest place. In CONs, the more popular a content is, the more probably it will be cached/stored in the local CON router. Most of the popular contents will be served by the CON routers which is located near the users. Moreover, it is expected that the overall congestion and latency of the CON will be decreased as contents will be uniformly distributed in the network. CONs will provide performance improvement at the expense of storage overhead at the service providers. However, we expect that the storage cost will be inexpensive in the future.

Flexible (or Extensible) Dissemination

CONs can provide a more flexible (or extensible) dissemination infrastructure than the conventional Internet. For example, we can easily leverage multicasting [4] in a CON. The conventional Internet only supports the end-to-end data transmission. Figure 1 shows how multicasting can be employed in the CON. When *User 1* subscribes to a stream which is published by a *Server*, the stream can be stored in any CON router (*CON router A* in Figure 1) which is located between *User 1* and the *Server*. If *User 2* wants to subscribe to the same stream, he/she can retrieve that from *CON router A* not from the *Server*. This is the typical scenario of multicasting: we can save the redundant transmissions for *User 2* between *CON router A* and *Server*. As you can see, it is much simpler to support multicast in CONs than in the conventional Internet: we do not need any tree construction, group management, or additional signaling in CONs. Besides, CONs can be easily extended to support many other kinds of functionalities such as Digital Right Management (DRM), QoS provision, or indirection infrastructure since a CON router takes a role of a middlebox in the network. With a middlebox functionality upon a network layer, a CON router can leverage the application context such as what is the required content/service.

Application Scenarios of CONs

Recently, retrieving contents is one of the most popular usage in the Internet. When users are going to download contents through the network, CONs in the Future Internet would be beneficial to users in terms of two perspectives: the contents accessibility and the digital right versatility.

Because the Internet was created for the purpose of the host-to-host communication, users suffer from a cumbersome procedure while downloading contents through the Internet. The usual procedure is as follows. First a user searches for the content with some search engines and eventually finds the URL of the content source. When the user connects to the content source, he/she has to figure out the way how he/she can download the content. The downloading methods are different at every source; e.g. a source might request signup or extra payment, another might request for installing its own application and so on. In CONs, users search for contents with search engines and get the content ID (CID) of the content itself which is closest to the user. Without bothering to learn about sources, users can just download the content in a consistent manner: just click and download it. This will greatly improve the contents accessibility in the Future Internet.

It is very inconvenient to get contents via content providers due to Digital Rights Management (DRM). DRM refers to access control technologies used by publishers and copyright holders to limit the usage of digital media or devices [3]. The problem is that every content provider adopts a different DRM technology and the DRM is integrated into the data file itself. If a user downloads a music file or a movie file from a content provider, he/she would play the content only by a specific media player or a device; it is a huge obstacle for users to enjoy online contents nowadays. In the CON environment, the DRM is not embedded in the content. The network will separately provide the content and its license, which is issued by the content creator, not by every content provider. This scenario is illustrated in Figure 2. User1 acquires the license for the contents from the content creator/publisher. User1 can download contents wherever they can access CONs and the network will guarantee the validity of the content: CON will provide the requested contents from closest source, Source1. Once having got the valid licenses, users only need to download the contents because licenses are provided in a consistent way. It will make users easily enjoy the online content services and encourage the online content business.

Current Research Activities on CONs

In this section, we introduce several representative activities on the CON research including our approaches.

DONA

Data-Oriented Network Architecture (DONA) is a content-oriented Internet architecture proposed by Koponen et al. [2, 5]. They realize a content-oriented network by two operational primitives: FIND and REGISTER. Users can request named data from the network by using the FIND primitive while content providers can publish a data object which will be served to the users by using the REGISTER primitive. To support these two primitives in CONs, DONA introduces resolution handlers (RH) which forward contents to the users in an overlay manner. With DONA, a user does not need to know the name or IP address of the host which contains the content he or she wants since DONA supports the route-by-name approach that couples the name resolution and routing.

PARC Content-Centric Network

Palo Alto Research Center (PARC) launches a content-centric networking research program named assurable global networks (AGNs) [6]. They insist that the data should be a first-order object in the network, so the data will be requested by the name. They also focus on the point-to-multiparty or multiparty-to-multiparty information dissemination rather than traditional point-to-point conversations. The main feature of content-centric AGNs is that the security will reside in the data itself not in the network channel as in today's Internet. The network only concerns how to distribute the data and the publishers control the security of the data. In consequence, the content-centric network will be a huge storage of authenticated data.

Open CDN Project

Open Content Delivery Network (OpenCDN) by the University of Roma is an application-level content delivery network for live and recorded multimedia contents [7]. Live streaming contents cannot be served efficiently to a very large number of clients in the client-server architecture due to bandwidth and processing power limitations. To handle this problem, OpenCDN constructs application-level tree via relay nodes which distribute the multimedia contents. To coordinate relay nodes, the request routing and distribution management (RRDM) is introduced as a control entity. RRDM collects client-related information from relay nodes and decides the best relay node for a newly joining client. OpenCDN is implemented for linux-based machines and OpenCDN 0.7.7 source code is now available in the website.

Oscar Network

Oscar is a content-oriented structured overlay network for heterogeneous environments [8]. Many measurement studies reveal that there arises heterogeneity in P2P systems due to differences in storage capacity, bandwidth, content repository among peers. Also, the key distribution of content over the key space, queries for content, and data access pattern show non-uniform distribution. To adapt to arbitrary distribution patterns and heterogeneity, Oscar collects sampling information of key distribution during the overlay construction and uses that information to choose routes based on small-world graphs. Since peers in Oscar have flexibility in choosing routes, they can decide on the amount of resources to be used for Oscar such as bandwidth and storage locally and autonomously which makes Oscar adequate to heterogeneous environments.

Siena

Siena (Scalable Internet Event Notification Architectures) is a content-oriented network featuring a generic scalable publish/subscribe event-notification service [9]. Clients advertise the information about events which they generate through Siena. Or they subscribe the service for notifications of interest. To support this publish/subscribe service in wide area networks, it is needed to maximize expressiveness for event description while not sacrificing scalability for event delivery. To this end, they formulate a general model of addressing and routing for content-based routing to maximize both expressiveness and scalability. Prototype implementation of Siena is also available on the website.

DACON & CODI

Our approach for the content-oriented network is two-fold: Decentralized and Autonomous Content-centric Overlay Networking (DACON) and Content Delivery Infrastructure based on DHT (CODI). DACON focuses on pervasive content access services without additional Internet access cost. To this end, DACON pays attention to user-deployable devices such as IEEE 802.11 access points and cellular home base stations. These devices construct a peer-to-peer overlay structure in an autonomous and decentralized manner; DACON node will participate the overlay network based on its own policy. Also we propose an incentive system to motivate users to participate DACON.

CODI adopts a publish/subscribe model to decouple senders and receivers in transmitting contents, and a network caching to retrieve contents efficiently. In order to optimize the performance, we investigate a DHT-based routing infrastructure and a cache management scheme based on popularity.

Conclusions

The inconsistency between the Internet design and the real usage motivates us to propose Content-Oriented Networks (CONs). A CON is a new network based on the content-oriented paradigm; users care only about the data and network will provide the requested data to the users in an efficient way. Compared with the Internet, CONs can provide 1) pervasive experience, 2) performance improvement, 3) flexible (or extensible) dissemination. We are currently working on new network architecture: DACON and CODI.

Acknowledgments

This work was supported by the IT R&D program of MKE/IITA [2007-F-038-02, Fundamental Technologies for the Future Internet]. The ICT at Seoul National University provides research facilities for this study.

References

[1] An iPoque's Press Release. http://www.ipoque.com/news_&_events/internet_studies/internet_study_2007.

[2] Teemu Koponen, Mohit Chawla, Byung-Gon Chun, Andrey Ermolinskiy, Kye Hyun Kim, Scott Shenker, and Ion Stoica. A data-oriented (and beyond) network architecture. In *SIGCOMM '07: Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 181–192, New York, NY, USA, 2007. ACM.

[3] Wikipedia. *Digital Right Management*. http://en.wikipedia.org/wiki/Digital_rights_management.

[4] K. C. Almeroth. The evolution of multicast: From the mbone to interdomain multicast to internet2 deployment. *IEEE Network*, 14(1):10–20, Jan.Feb. 2000.

[5] DONA project page (Berkeley RADLab) http://radlab.cs.berkeley.edu/wiki/DONA

[6] PARC content-centric network http://www.parc.com/research/projects/networking/contentcentric/

[7] Open CDN project (Content Delivery Networks) http://labtel.ing.uniroma1.it/opencdn/

[8] Oscar Network - Data-oriented Structured Overlay for Heterogeneous Environment http://www.ntu.edu.sg/home/anwitaman/ResearchOscar.html

[9] Siena (Content-based Network) http://www-serl.cs.colorado.edu/ carzanig/siena/