

미래 인터넷을 위한 콘텐츠 기반 네트워킹 아키텍처

권태경
서울대학교

요약

본고에서는 미래 인터넷 구조를 제안하는 연구 중에서 콘텐츠 기반 네트워킹 구조에 관련된 연구들을 살펴본다. DONA는 모든 콘텐츠에 대해 각각 유일하고 flat한 이름을 부여하고 전체 콘텐츠들을 하나의 트리 구조로 라우팅하는 네트워킹 구조를 제시한다. CCN은 기존의 URL과 비슷한 계층적인 콘텐츠 이름을 사용하는 것을 제시하나 콘텐츠가 호스트와 무관하게 위치하기 때문에 역시 콘텐츠 이름 별로 flooding 기반의 라우팅을 하게 된다. 또한 본고에서는 기존의 호스트 기반 네트워킹과 콘텐츠 기반의 네트워킹의 장단점을 비교한다.

1. 서론

기존의 인터넷 구조는 호스트간의 자원을 공유하기 위한 목적으로 설계되고 사용되어 왔다. 이를 종단간 통신 모델이라고 흔히 얘기한다. 그러나 최근의 인터넷은 content delivery network (CDN), peer-to-peer (P2P) 등 사용자가 어느 호스트로 접속을 원하는 것이 중요한 응용보다 어떤 콘텐츠를 원하는 것이 중요한 응용들로 점점 초점이 옮겨지고 있다.

이러한 인터넷 흐름에 힘입어 새로이 콘텐츠 기반 네트워킹 관련 연구들이 최근에 제시되었다 [1,2,3]. 즉 현재의 인터넷은 어느 호스트로 연결을 할 것인가를 먼저 결정하고

그 뒤에 그 호스트에 있는 특정 자원이라 콘텐츠를 받아오는데 이와 대조적으로 호스트와 상관없이 찾고자 하는 콘텐츠를 명시하면 네트워크가 바로 콘텐츠를 찾아주는 것이다.

이러한 새로운 콘텐츠 기반 네트워킹 구조와 기존의 인터넷 구조에는 각각 route-by-name과 lookup-by-name이라는 두 가지 대조되는 원칙이 있는데, 이는 어떻게 원하는 콘텐츠를 인터넷에서 찾을 것인가 (resolution)에 대한 두 가지 방식이다. 본고에서는 먼저 route-by-name과 lookup-by-name 두 가지 메커니즘을 살펴본다.

콘텐츠 기반 네트워킹 구조를 제시한 연구들이 최근에 출판이 되었는데 그 중 가장 대표적인 연구는 data oriented network architecture (DONA)와 content centric networking (CCN)이다. 본고에서는 이 두 개의 아키텍처들을 중점적으로 설명한다.

콘텐츠 기반 네트워킹 구조는 효율적인 전송, 그리고 호스트에 비의존적인 좋은 가용성, 그리고 인증 (authentication) 보장이 쉽게 되는 장점들이 있지만, 결정적으로 확장성 (scalability)이라는 문제가 있다. 왜냐하면 모든 콘텐츠에 대해 이름 (혹은 식별자)을 붙여야 하는데 이 이름은 IP address 나 전화번호처럼 계층적으로 그룹화하기 힘들기 때문이다. 따라서 본고에서는 기존의 호스트 기반 네트워킹 (혹은 IP 기반 네트워킹) 구조와 콘텐츠 기반 네트워킹 구조를 결합하는 방안을 추가로 토론한다.

II. 콘텐츠 resolution

현재 네트워크에서 특정 콘텐츠를 찾는 방법에는 두 가지가 있다. route-by-name [4] 과 lookup-by-name [5] 이 그것들인데 서로 대조적인 접근 방식을 갖고 있다.

먼저 lookup-by-name은 현재 인터넷에서 쓰이는 방법인데 호스트 기반 통신 모델과 같이 쓰인다. 즉 인터넷에 연결된 호스트는 자신의 domain name을 갖게 되고 이 domain name으로부터 자신의 현재 위치 즉 IP address를 다른 사용자들이 lookup할 수 있게 하는 일종의 indirection system인 domain name system (DNS)를 필요로 한다. 즉 어떤 콘텐츠를 찾기 위해서는 그 콘텐츠가 있는 호스트의 domain name를 먼저 알고서 (예를들어 검색 엔진을 통하여) 그 뒤에 DNS를 통해 IP address를 알아내는 것이다. 이렇게 호스트의 위치를 알고 나면 호스트의 IP 주소와 호스트 내의 파일 디렉토리 path 정보를 통해 (uniform resource locator와 같이) 콘텐츠를 접근하게 된다.

Lookup-by-name은 인터넷 초기부터 사용되다가 점점 호스트 수가 많아지면서 1983년부터는 체계적인DNS가 도입되었다. DNS는 현재까지 인터넷이 폭발적으로 증가하였음에도 불구하고 domain name과 IP address 사이의 resolution을 서비스하여 왔다. 하지만 속도 면에서 보면 콘텐츠를 찾기 위해 DNS를 거치는 시간이 추가적으로 걸린다는 단점이 있고, 결정적으로 호스트에 의존적인 식별자 (예를 들어 URL)를 사용할 때에 만약 그 URL에 명시된 호스트가 다운되거나 그 호스트로 가는 링크에 문제가 생긴 경우 그 콘텐츠가 다른 호스트에 있더라도 다른 사람들이 콘텐츠를 찾을 수 없는 가용성 (availability) 문제가 있다.

그러나 현재까지 DNS는 에러에 민감한 트리 구조임에도 불구하고 캐싱 (caching) 및 anycast를 이용한 많은 root server의 설치를 통해 놀라운 가용성 (availability)를 사용자들에게 제공하여 왔다. 또한 DNS의 유연한 계층적 구조는 좋은 확장성 (extendibility)을 갖고 있어서 domain name에서 IP address로의 매핑뿐만 아니라 RFID tag ID 등 여러가지 식별자들을 통합할 수 있는 기반구조로서 대두되고 있다.

Route-by-name 원리는 IP 네트워킹의 lookup-by-name 원

리와는 대조적이다. Route-by-name방법에서는 모든 콘텐츠는 자신의 고유한 식별자가 있고 이것이 곧 routing entry로 동작한다. 즉 이름(식별자)가 곧 라우팅을 위한 key index로 동작하게 된다. 따라서 DNS같은 indirection system이 필요 없다. 그러므로 속도면에서 더 빠르게 콘텐츠를 찾을 수 있다. 그리고 호스트에 비의존적으로 동작하므로 원래 콘텐츠를 갖고 있었던 특정 호스트가 다운 되더라도 해당 콘텐츠가 어느 곳이라도 있으면 찾을 수 있다. 따라서 속도와 가용성 면에서 장점을 가진다.

하지만 결정적으로 모든 콘텐츠에 대해서 routing entry를 유지하는 것은 확장성 (scalability) 문제가 심각하다. IP address의 경우에는 network 식별자 (즉 network prefix)만 routing entry로 유지하기 때문에 콘텐츠 수에 비해 훨씬 적다. 네트워크 식별자의 총 숫자는 2010년 8월 현재 대략 33만 - 34만 개 정도이고 계속 늘어나고 있다. 이 정도에도 현재 IP networking의 routing scalability문제가 심각하다고 현재 인터넷 구조 연구자들은 얘기하고 있다. 이에 반해 콘텐츠의 수는 적어도 몇 십억 개 이상이 존재할 텐데 모든 콘텐츠에 대한 routing entry를 유지한다는 것은 거의 불가능하다. 전 세계의 모든 콘텐츠를 연결하는 route-by-name 기반 네트워킹은 가까운 미래에는 현실성이 없는 것으로 현재 생각되고 있다.

따라서 보다 현실성 있는 route-by-name 원칙을 설계 및 제안하는 연구가 콘텐츠 기반 네트워킹 구조를 위해서 필요한 실정이다.

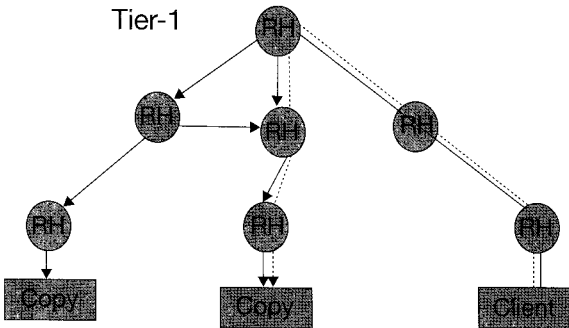
III. DONA

본 장에서는 Data Oriented Network Architecture (DONA)에 대해 소개하고 특히 서론에서 설명한 이슈들이 DONA에서 어떻게 다루어지는지 살펴본다.

DONA는 먼저 각 콘텐츠의 식별자로서 Persistence와 authenticity를 위해 flat, self-certifying name을 사용한다. 하나의 콘텐츠의 name은 그 콘텐츠를 만든 생산자 P와, P가 선택한 라벨 L의 조합으로 (즉 P:L)로 정해진다. P는 생산자의 공개키의 해쉬(hash) 값으로 정해지고, L은 P가 유일하게

만 선택하면 된다. 이렇게 name이 할당된 하나의 콘텐츠는 실제로 세 개 요소의 조합으로 이루어진다 (content data, public key, signature). 이렇게 콘텐츠 이름과 콘텐츠 내용을 구성하여 각각의 콘텐츠는 self-certifying 성질을 갖게 된다. 또한 특정 호스트나 위치에 구속되는 이름이 아니므로 인터넷 어느 entity에도 저장될 수 있는 persistence를 갖는다.

그러나 콘텐츠 이름에 어떤 계층적 성질이나 그룹화 될 수 있는 성질이 없기 때문에 이름과 콘텐츠의 위치를 매핑 (resolution)하기 위해서는 각 이름 별로 routing entry가 필요하다. DONA에서는 이렇게 resolution을 담당하는 entity를 resolution handler (RH)라고 명명하였다. 각 인터넷의 호스트는 자신과 가장 가까운 RH와 연결을 갖고 인터넷의 콘텐츠를 접근하게 된다.



(그림 1) DONA의 RH 라우팅 구조

DONA에서 RH들은 (그림 1)과 같이 트리 형태의 연결성을 취하게 된다. 각 RH는 자신이 속한 호스트들이 register한 콘텐츠들의 대한 routing entry를 트리를 따라 가장 상위의 root RH에게까지 전달하게 된다. 이런 방식을 통해 하나의 RH를 통해 register된 콘텐츠는 현재 자신이 속한 RH와 그 상위 RH 모두들에게 자신에 대한 routing entry를 만들게 한다. 따라서 어떤 노드가 특정 콘텐츠를 찾으려면 (find) 최악의 경우 root RH에게 까지 질의가 도달하면 어디에 콘텐츠가 있는지 알게 된다.

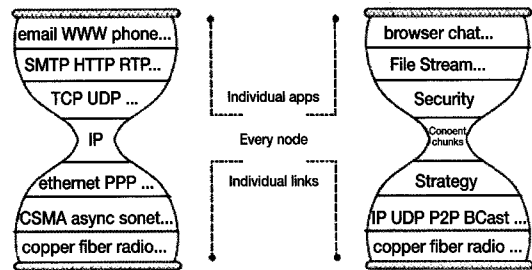
이러한 RH들의 관계는 작게는 현재 router와 비슷한 entity이나 크게 보면 현재 ISP와 비슷한 entity로 생각할 수 있다. RH간의 관계는 ISP들간의 customer/peer/provider 관계와

비슷한 연결성을 가질 수도 있다. 위와 같이 register나 find를 RH 간에 전달할 때는 RH 사이에 미리 보안 관계가 구축되어 있어 암호화되어서 전달되는 것으로 간주한다.

이러한 트리 구조의 라우팅 체계는 root RH가 모든 콘텐츠의 라우팅 정보를 다 가져야 하므로 확장성 (scalability) 문제점이 심각하다. 그러나 확장성 문제 외에는 mobility, multi-homing, multicast, authentication 등 많은 측면에서 장점이 있다. 따라서 향후 routing scalability 이슈가 큰 문제가 안 될 정도로 저장 및 프로세싱 기술이 성숙하게 된다면 앞으로 유망한 미래 인터넷의 네트워킹 구조의 후보 기술이라고 할 수 있다.

IV. CCN

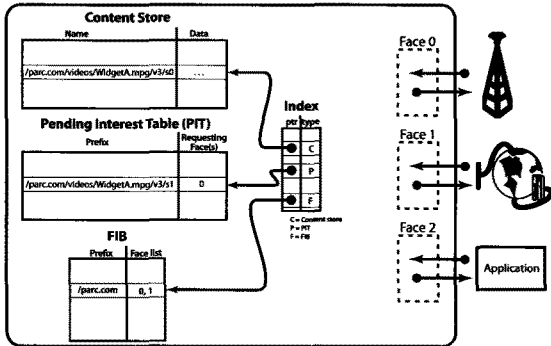
PARC 연구소에서 제안한 content-centric networking (CCN)은 기존의 호스트 기반의 인터넷 구조를 콘텐츠 기반의 새로운 네트워킹 구조로 바꾸기 위한 제안되었다. (그림 2)에서 보여 주는 바와 같이, 기존의 IP가 여러 상위의 응용 프로토콜과 다양한 링크 기술들을 하나로 연결했다면 CCN에서는 content 계층이 여러 응용 관련 기능들을 위로 지원하면서 아래로는 여러 통신 기법들을 이용할 수 있다는 장점을 내세운 것이다. 여기서 strategy 계층은 예를 들어 콘텐츠 전송이 여러 interface 혹은 여러 connectivity를 사용해서 전송할 수 있는 기능을 의미한다.



(그림 2) CCN 과 IP의 계층 구조

CCN에는 두가지 패킷 유형이 있다. Interest와 Data 패킷들인데 전자는 특정 콘텐츠를 찾기 위해 쓰이고, 후자는 실제

로 콘텐츠를 요청한 클라이언트에게 전달되는 콘텐츠 내용을 담은 데이터 패킷이다.



(그림 3) CCN 노드의 자료 구조들

CCN node는 IP 라우터와 비슷한 entity로서 (그림 3)과 같이 세 가지 자료구조를 갖고 있다. Forwarding information base (FIB)는 interest 패킷을 해당 콘텐츠 데이터가 있는 곳으로 라우팅하기 위한 테이블로서 거의 IP의 라우팅 테이블과 유사하다. 콘텐츠 이름을 보고 longest prefix matching을 통해 해당 interface로 전달하게 된다. 한가지 차이점은 interest 패킷을 라우팅할 때 한 곳이 아니라 여러 곳으로 전달할 수 있다는 것이다.

Content store는 콘텐츠 내용을 담은 data 패킷들을 저장하는 곳으로 향후 같은 콘텐츠를 다른 클라이언트가 또 요구할 경우 해당 콘텐츠를 저장하고 있는 CCN 노드가 바로 콘텐츠를 전달할 수 있다.

Pending interest table (PIT)는 interest가 어느 곳에서 왔는지를 기억하는 장소로 이것을 보고 나중에 data 패킷이 오면 PIT를 참조해서 해당 클라이언트에게 전달하게 된다. 하나

의 PIT entry는 soft state로 유지되어서 일정 시간 동안 data packet이 오지 않으면 없어지게 된다.

CCN에서의 하나의 콘텐츠는 여러 컴포넌트 (혹은 스트링) 들이 계층적으로 구성된다. (그림 4)를 예를 들어 설명하면 다음과 같다. 편의상 콘텐츠의 이름이 / 로 시작되고 각각의 스트링도 / 로 구분된다고 하자. 먼저 parc.com은 호스트 네임과 같은 의미이나 굳이 꼭 호스트 네임으로 시작할 필요는 없다. CCN에서 네이밍은 어떤 스트링이던지 계층적으로만 구성되면 된다. 예를 들어 /ThisRoom/Printer 혹은 /Local/Friends 등 어떤 스트링의 조합도 가능하다. 그림 4에서는 편의상 호스트 이름, 디렉토리 구조, 파일 이름으로 앞부분을 구성한 다음 그 뒤에 버전과 세그먼트 번호를 부쳐서 좀 더 유연한 콘텐츠 네이밍을 가능하게 하였다.

각 콘텐츠는 특정 호스트와 연관이 된 것이 아니기 때문에 각 콘텐츠는 flooding 기반으로 자기 위치를 알리고 나중에 클라이언트가 해당 콘텐츠를 찾을 때는 flooding 시에 형성된 FIB entry를 찾아서 interest 패킷이 포워드되게 된다.

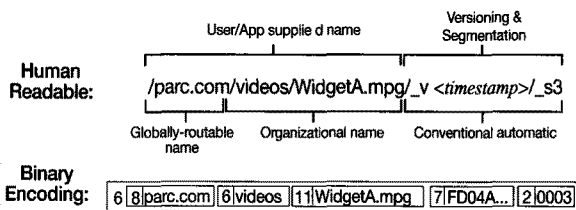
콘텐츠에 대한 인증은 디지털 서명을 통해서 이루어진다. 기존의 DONA 같이 self-certifying 구조를 가지기 보다는 CCN은 콘텐츠 이름과 콘텐츠 데이터 사이의 연관(binding)을 인증하는 데에 초점을 맞추었다. 이를 통해 콘텐츠 publisher는 어떤 이름이던 자유롭게 콘텐츠 데이터에 연관시킬 수 있다.

CCN은 콘텐츠 보안, 유연하고 효율적인 콘텐츠 전송 등의 장점을 가지나 역시 콘텐츠 수가 늘어남에 따라 확장성의 문제는 피할 수가 없다.

V. 호스트 기반 네트워킹과 콘텐츠 기반 네트워킹의 결합

이상의 DONA와 CCN에서 살펴보았듯이 콘텐츠 기반의 네트워킹은 콘텐츠의 문맥을 중간 노드들이 알게 되고 또한 콘텐츠를 저장할 수 있음으로서 multicast, mobility, multi-homing, multiple interface, flash crowd 등의 여러 복잡한 네트워킹 상황에서 효율적인 콘텐츠 전송을 할 수 있다.

반면에 IP 네트워킹은 중간 노드들이 콘텐츠의 문맥을 알



(그림 4) CCN 에서 콘텐츠 이름 예

수 없기 때문에 효율적인 전송은 어려우나 전 세계적으로 네트워크를 운영하기 위한 라우팅 확장성이 좋고 DNS를 통해 유연한 이름들을 지원할 수 있다.

따라서 작은 규모의 네트워크 환경에서 매우 인기 있는 콘텐츠들에 대한 콘텐츠 기반 네트워킹 솔루션을 도입하고 전 세계적으로는 IP기반 네트워킹 구조로 운영하는 조합이 현재 기술 수준으로는 적절한 것으로 생각된다.

VI. 결 론

본 고에서는 최근 부각되고 있는 콘텐츠 기반 네트워킹 관련하여 현재 IP networking에서 사용되는 lookup-by-name 과 콘텐츠 기반 네트워킹에서 사용될 route-by-name 두 가지 원칙을 설명한 후 가장 대표적인 두 가지 콘텐츠 기반 네트워킹 구조들인 DONA와 CCN 의 동작 원리, 콘텐츠 네이밍 기법, 특징 등을 살펴보았다. IP 네트워킹과 콘텐츠 기반 네트워킹이 모두 각자의 장단점을 갖고 있기 때문에 앞으로 네트워킹 요소 기술이나 인터넷 환경의 변화에 따라 다양한 두 구조의 조합 중에 어떤 것이 적절한 것인지가 향후 중요한 연구 주제로 생각된다.

참 고 문 헌

- [1] Teemu Koponen et al. "A Data-Oriented (and Beyond) Network Architecture," ACM SIGCOMM 2007, Aug. 2007.
- [2] Van Jacobson et al., "Networking named content," Proceedings of the 5th ACM CoNEXT 2009, Dec. 2009.
- [3] Kideok Cho et al. "Content-Oriented Networking as a Future Internet Infrastructure: Concepts, Strengths, and Application Scenarios," International Conference on Future Internet (CFI) 2008.
- [4] Mark Gritter and David R. Cheriton, "An Architecture for Content Routing Support in the Internet," In the

USENIX Symposium on Internet Technologies and Systems, March 2001.

- [5] Paul Mochapetris, "DOMAIN NAMES - CONCEPTS AND FACILITIES," IETF RFC 1034, 1987.

약 력



권 대 경

1993년 서울대학교 컴퓨터공학과 학사
 1995년 서울대학교 컴퓨터공학과 석사
 1998년 IBM Watson 연구소 방문 학생
 1999년 University of North Texas, 방문 연구원
 2000년 서울대학교 컴퓨터공학과 박사
 2001년 ~ 2002년 UCLA, 박사후 연구원
 2003년 ~ 2004년 City University New York, 박사후 연구원
 2004년 ~ 현재 서울대학교 컴퓨터공학부, 조교수, 부교수

