캐시 효용성을 높이기 위한   
캐시 콘텐츠 및 자원 최적화 기술 연구

Sangmin Lee

Kyonggi university Department of Computer Science

Suwon, Republic of Korea

[e-mail: d9249@kyonggi.ac.kr]

\*Corresponding author: Sangmin Lee

Abstract

현재 인터넷 환경에서 기존 호스트 기반 IP 네트워크는 구조적 한계에 직면하고 있다. 이러한 문제를 극복하기 위해 기존의 네트워크 구조 전체를 바꿀 수 없기에 IP 기반 호스트 중심 네트워크에 P2P, CDN, CPN 등을 Overlay로 설치해 옥상옥 구조 (patch over patches)로 운영을 통한 임시방편적 한계를 탈피해, 콘텐츠 중심 서비스를 구조적으로 제공하기 위한 ICN (Information Centric Network)을 차세대 모델 사용을 고려하고 있으며, 기존 IP 네트워크를 통해 ICN 시스템의 콘텐츠에 접근할 때 발생할 수 있는 문제점을 파악하고, 이를 근본적으로 해결할 수 있는 새로운 IP 지원 ICN 시스템 구조를 설계하여 ICN 시스템 내 CN (Cache Node)에 저장된 캐시 정보를 공유하고 조정하는 CNC (Cache Node Coordinator)를 구조적으로 제안한다. CN에 콘텐츠 일부만 저장되거나 불필요한 콘텐츠 캐시가 남아 새로운 콘텐츠를 저장되지 못하는 캐시 파편화 현상을 방지하기 위 함이며, 서로 다른 CN에 저장된 콘텐츠 캐시 내용을 CNC를 통해 능동적으로 CN끼리 정보를 교환하여 경로 상 불필요한 콘텐츠 캐시를 제거하여 협력적 관리를 통해 캐시 효율을 전역적으로 최적화하기 위함이다.

이러한 해결방안을 설계하기 위해서 과거 사용자 요구 콘텐츠 정보와 네트워크 토폴로지 정보를 바탕으로 미래 요구 콘텐츠가 무엇인지를 GNN (Graph Neural Network)을 통해 예측하는 딥러닝 모델을 연구하며, 이러한 연구를 통해 IP 지원 ICN 시스템을 구축하게 되면 네트워크 효율성을 획기적으로 증가시켜 멀티미디어 콘텐츠 전송이 트래픽의 대부분을 차지하는 현재 인터넷 서비스의 품질이 크게 개선될 것이다.

Keywords: Information Centric Network, Graph Neural Network, Cache Node, Cache Node Coordinator

1. 연구배경

음성 중심의 전화망을 개선하여 데이터 중심의 서비스를 제공하기 위해 구축된 호스트 기반 IP 네트워크 구조는 이메일, 웹 서핑, 파일 전송 등 기본적인 인터넷 서비스에서 널리 활용되고 있다. 그러나 스마트폰, 태블릿, IPTV, 스마트 센서, 스마트 장치 등을 통해 VoD, Edge, Cloud, IoT, IoV(Internet of Vehicle), V2V(Vehicle-To-Vehicle), M2M(Machine-To-Machine), Smart metering, Smart home 등의 다양한 서비스를 제공해야 하는 현재 인터넷 환경에서 기존 호스트 기반 IP 네트워크는 구조적 한계에 직면하고 있다. 특히 트래픽의 80% 이상이 멀티미디어 콘텐츠인 현재 인터넷은 정보를 어디에서 획득할 것인가 보다 어떤 정보를 어떻게 획득할 것인가에 더 관심을 두고 있기에, 기존 IP 기반 호스트 중심 네트워크에 P2P, CDN, CPN 등을 Overlay로 설치해 옥상옥 구조 (patch over patches)로 운영하는 임시방편적 한계를 탈피해, 콘텐츠 중심 서비스를 구조적으로 제공하기 위한 ICN (Information Centric Network)을 차세대 모델로 고려하고 있다.

ICN 구조에서는 네트워크가 콘텐츠의 이름(named content)을 인식하고 이를 통해 라우팅 정보를 결정한다. 사용자가 콘텐츠 이름을 이용해 네트워크에 데이터를 요청하면 네트워크는 사용자와 가까운 위치에 있는 캐시 노드를 통해 콘텐츠를 전달하여 네트워크 효율성을 높이고 사용자 및 콘텐츠의 이동성을 보장한다. 그 결과 다수가 공유하는 멀티미디어 콘텐츠를 보다 효율적으로 전송할 수 있으며, 콘텐츠 중심의 보안 정책과 접근 제어 모델을 자연스럽게 구현할 수 있다. 이러한 ICN 구조를 구체화하기 위해 지난 10년간 미국과 유럽 중심으로 DONA, PURSUIT, SAIL, CONET, CONVERGENCE, MobilityFirst, DOCTOR, POINT, RIFE, ICE-AR 등 다수의 프로젝트가 수행되었고 지금도 진행 중이다.

기존 IP 구조가 적합한 인터넷 서비스들도 많이 존재하기에 네트워크 구조를 모두 ICN 시스템으로 바꿀 필요도 없기에 IP 기반 구조를 한꺼번에 콘텐츠 기반 ICN 구조로 바꿀 수는 없다. 따라서 IP 기반 네트워크와 ICN 기반 네트워크 간에 서로 공존하여 효율성을 높일 수 있는 기술 연구가 필요하다. 최근 Hybrid-ICN, NDN-LAN, ICNOverP4 등 IP 네트워크와 ICN 네트워크가 공존하는 방법에 관한 연구가 진행되고 있다. 그러나 대부분 두 네트워크를 단순히 연결하거나 각각 병렬적으로 운용하는 기술만 제시하고 있고, IP 네트워크를 통해 ICN 시스템에 접속하여 콘텐츠를 효율적으로 획득하는 방법에 관한 연구는 거의 없다. 또 사용자 단말에 ICN 시스템 접속을 위한 별도의 프로토콜이나 프로그램 설치를 요구하거나, HTTP와 같은 특정 프로토콜에 한정된 해결책만을 제시하고 있다. 따라서 본 연구에서는 IP 기반 네트워크가 대부분인 현 상황에서 사용자 단말에 특정 프로토콜이나 프로그램의 추가 설치 없이 범용적으로 접속 가능하고, 실현 가능한 ICN 시스템 구현에 관한 연구를 수행한다.

해당 연구의 최종 목표는 ‘IP 기반 네트워크 사용자를 투명(transparent)하게 지원하며 네트워크 통합 관리를 통해 콘텐츠를 효율적으로 전송할 수 있는 실현 가능한 ICN 시스템 구조와 이를 위한 핵심 시스템 기술 개발’ 이다. 특히 최근 네트워크 하부 구조로 널리 구축되고 있는 SDN (Software-Defined Networking) 기술을 바탕으로 빠른 콘텐츠 탐색, 효율적 콘텐츠 캐싱, 동적 컴퓨팅 자원 할당, 최적 경로 설정 등을 통합적으로 제공하며 현실 세계에서 실질적으로 구현될 수 있는 ICN 시스템 구조 및 핵심 기술에 대해 구체적으로 연구한다.

연구의 첫 단계로 IP기반 네트워크를 지원하는 관점에서 기존 ICN 모델을 먼저 분석한다. 단말의 변경 없이 기존 IP 네트워크를 통해 ICN 시스템의 콘텐츠에 접근할 때 발생할 수 있는 문제점을 파악하고, 이를 근본적으로 해결할 수 있는 새로운 IP 지원 ICN 시스템 구조를 설계한다. 제안하는 구조는 이미 널리 구축되고 있는 SDN 기술을 활용하여 네트워크 환경을 전역적인 관점에서 통합적으로 관리한다. 또 사용자가 IP 네트워크를 통해 요청한 콘텐츠를 ICN 시스템에 전달하는 SG (Service Gateway), 다수의 사용자가 요구한 콘텐츠를 공동으로 처리하여 네트워크 비용을 최소화하며 SG까지 전달하는 최적 경로를 계산하는 RPC (Routing Path Coordinator), 사용자 요구에 맞게 콘텐츠를 정제하는 SFC (Service Function Chaining) 기술을 구현하기 위해 컴퓨팅 자원을 동적으로 할당하는 CRA (Computing Resource Allocator), ICN 시스템 내 CN (Cache Node)에 저장된 캐시 정보를 공유하고 조정하는 CNC (Cache Node Coordinator)를 구조적으로 제안한다.

현재 인터넷의 80% 이상을 차지하는 멀티미디어 콘텐츠를 ICN 시스템에서 효율적으로 전송하기 위해서는 하나의 OCS (Original Content Server)로부터 다수의 CN (Cache Node)을 거쳐 사용자들이 연결된 SG까지 이어지는 콘텐츠 전송 트리를 SDN 제어기를 통해 효율적으로 형성할 수 있어야 한다. 사용자가 콘텐츠를 가장 빨리 획득하기 위해서는 해당 콘텐츠를 가지고 있는 CN 중 SG와 가장 가까운 노드로 연결하면 되는데, 이 경우 전체 네트워크 전송 비용이 최소화되지 못하고 CN과 OCS 사이 간격이 크면 콘텐츠 전송 지연 시간이 길어진다. 실시간 콘텐츠처럼 전송 지연 시간에 민감한 콘텐츠들은 OCS부터 CN을 거쳐 SG까지 최단 거리로 연결하는 트리를 형성하여 전송하면 되는데, 이때에도 네트워크 전송 비용은 최소화되지 않는다. 따라서 콘텐츠 특성에 따라 콘텐츠 전송 트리 형성 및 수정 방법이 달라져야 하는데, 이를 위해 라우팅 경로 정보를 SDN 제어기를 통해 수집하여 전역적 관점에서 최적 경로를 설정하는 RPC (Routing Path Coordinator)를 도입하고, 이를 기반으로 사용자 서비스 요구 수준에 따라 QoE (Quality of Experience)를 보장하는 효율적인 콘텐츠 전송 트리 구성 방법을 설계하며, 실제 인터넷에서 사용되는 콘텐츠 전송 트리는 한 시점에 다수의 사용자를 대상으로 생성되지 않고, 점진적으로 사용자가 늘거나 줄면서 동적으로 변화한다. 또 네트워크 비용을 최소화하기 위해 기존 트리를 변화시키면 기존 사용자의 QoE가 달라지는데, 이점까지 고려하여 점진적 사용자 변화와 기존 사용자 QoE 변화까지 반영한 동적 콘텐츠 전송 트리 형성 및 유지 방안을 연구한다.

ICN은 네트워크 자원 효율성과 빠른 데이터 확산을 위해 CN (Cache Node)을 통한 INC (In- Network Caching)을 채용한다. INC는 이동성이 높은 사용자를 지원하는 IoV, 5G/6G 서비스에서 네트워크 효율성을 증대하는 에지 컴퓨팅 기술 구현을 위한 거의 유일한 방안으로 주목받고 있다. INC의 효능은 CN에 저장되는 콘텐츠 내용과 CN의 위치 및 용량에 좌우되는데, 이를 극대화하기 위해 다양한 콘텐츠 캐싱 방법이 연구되었다. 그러나 기존 방법들은 대부분 독립적이고 분산된 네트워크 환경에서 이웃 CN끼리만 지역적으로 정보를 공유하여 그 성능이 제한적이다. 반면, 본 연구에서 제안하는 ICN 시스템은 SDN을 기반으로 RPC (Routing Path Coordinator)와 CRA (Computing Resource Allocator)를 통해 콘텐츠 이동 경로를 전지적으로 파악하고, 서로 다른 CN에 저장된 콘텐츠 캐시 내용을 CNC (Cache Node Coordinator)를 통해 협력적으로 관리하여 캐시 효율을 전역적으로 최적화한다. 또 CN에 콘텐츠 일부만 저장되거나 불필요한 콘텐츠 캐시가 남아 새로운 콘텐츠를 저장되지 못하는 캐시 파편화 현상을 방지하기 위해, CNC를 통해 능동적으로 CN끼리 정보를 교환하여 경로 상 불필요한 콘텐츠 캐시를 제거하는 방안도 제시한다. 그리고 CNC에서 사용자 요구 콘텐츠를 선제적으로 파악하고 이를 적합한 CN에게 할당하는 방법을 제시한다.

2. 연구 추진 방안

ICN 시스템에서 네트워크 효율성을 높이기 위해서는 다수의 사용자가 요구하는 콘텐츠를 어떠한 CN을 통해 제공하느냐가 중요하다. 이는 OCS에서 여러 CN을 거쳐 다수의 사용자와 연결된 SG들까지 이어지는 콘텐츠 전송 트리를 최소 비용으로 형성하는 스타이너(Steiner) 트리 문제로 귀결된다. 스타이너 트리 문제는 NP-hard의 복잡도를 가지기에 이를 지역적으로만 최적화하는 휴리스틱 알고리즘들이 많이 제안되었는데, 대부분 네트워크 규모가 커지면 계산량이 많아져 실시간으로 변화하는 사용자의 변화를 따라가며 비용이 작은 적절한 트리를 생성하지 못한다. 따라서 본 연구에서는 최근 주목받고 있는 딥러닝 기술을 바탕으로 네트워크 환경을 다양하게 학습한 에이전트를 통해 사용자 QoE를 만족하는 최적 트리를 실시간으로 계산하는 방안을 전략적으로 연구한다.

딥러닝은 학습 과정에서 많은 시간을 요구하지만 학습이 끝난 모델을 통해 입력 네트워크 환경에 대한 최적 트리를 계산할 때는 그 결과를 빠르게 획득할 수 있다.

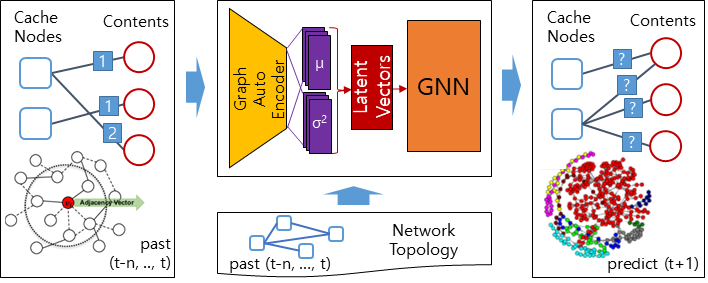
ICN 시스템에서 CN (Cache Node)은 사용자가 요구할 콘텐츠를 미리 예측하여 선제적으로 저장할 수 있으면 좋은데, 이를 위해 과거 사용자 요구 콘텐츠 정보와 네트워크 토폴로지 정보를 바탕으로 미래 요구 콘텐츠가 무엇인지를 GNN (Graph Neural Network)을 통해 예측하는 알고리즘을 전략적으로 연구한다. GNN은 그래프를 적용한 신경망으로 노드, 간선, 그래프 등을 예측할 때 사용되는데, 최근 구글이 제안한 attention기반 transformer [21]를 적용한 GAT (Graph Attention Network) [22] 모델이 시계열 데이터를 분석하고 예측하는데 매우 좋은 성능을 보여주고 있다. attention 모듈은 고전적인 RNN 모델의 한계를 벗어나 방대한 학습 네트워크 안에서 학습 에이전트가 중요하게 학습해야 할 부분이 무엇인지를 지정할 수 있게 도와준다. 본 연구에서는 CN과 캐시 콘텐츠를 정점으로 두고 둘 간의 상호 연관성을 간선로 표현한 뒤, GAT 모델 기반 학습을 통해 새로운 간선을 예측함으로써 CN에서 선제적으로 저장해야 할 콘텐츠의 우선순위를 정한다. 이때 네트워크 토폴로지 변화 정보와 학습 모델의 출력인 캐시 콘텐츠 예측 결과를 재귀적으로 모델에 입력하여, 네트워크 상태를 반영한 시계열 예측이 가능하도록 설계한다. 그리고 LINE, Node2vec, GraphSAGE, DeepLinker 등 최근 GNN 학습 모델도 도입하여 그 성능을 분석한다. 마지막으로 캐시 노드 수에 비해 매우 많은 콘텐츠 수는 그 비대칭성으로 인해 학습량이 너무 많아 모델 성능을 저하시킬 수 있다. 이를 방지하기 위해 콘텐츠 정보를 K-means clustering 등으로 유사도에 따라 묶는 기계학습 모델을 접목하여 학습량을 경량화하는 방안과 AAE 기반 멀티 태스크 러닝을 통해 두 모델을 통합하여 시너지를 높이는 방안을 연구한다. 연구에 필요한 학습 데이터는 Neflix, Disney+ 등 세계적인 OTT 서비스의 콘텐츠 랭킹 정보 등 인터넷 공개 데이터를 통해 확보할 예정이다.

Figure . 캐시 노드와 콘텐츠 상관관계 예측 학습 모델 예

3. 활용방안

네트워크를 기능별로 통합 관리하여 딥러닝 기술을 네트워크 시스템에 밀 접목할 수 있는 합리적인 구조를 제안한 본 연구의 결과물은 네트워크 효용성을 극대화하고 네트워크 구축 비용을 크게 절감해 미래 인터넷 산업 분야에서 널리 활용될 수 있다. 또 새로운 인터넷 플랫폼 구축을 가속화하여 지금까지 생각하지 못한 인터넷 서비스 모델을 발굴할 수 있다.

4. 기대효과

본 연구의 연구 결과를 바탕으로 IP 지원 ICN 시스템을 구축하게 되면 네트워크 효율성이 획기적으로 증가하여 멀티미디어 콘텐츠 전송이 트래픽의 대부분을 차지하는 현재 인터넷 서비스의 품질을 크게 개선할 수 있다. 또 새로운 형태의 인터넷 플랫폼에 의해 다양한 부가가치 서비스가 생겨 나고, 새로운 인터넷 비즈니스 모델이 발굴되어 인터넷 서비스가 경제적, 산업적으로 더욱더 확산하는 계기가 될 것이다.

5. 결론

전통적으로 구성된 IP 기반 구조를 한꺼번에 콘텐츠 기반 ICN 구조로 변경할 수 없기에 IP 네트워크에 접속하여 ICN 네트워크를 통한 콘텐츠 전송 기법에 관한 효율성을 재고하기 위해서 CNC (Cache Node Coordinator)를 연구한다. 사용자가 요구할 콘텐츠를 미리 예측하여 CN에 저장을 하기 위해서는 많은 정보가 필요하며, 이러한 많은 정보를 고려하여 사용자가 원하는 콘텐츠 전송을 더욱 효율적으로 전송하여 네트워크의 부하를 줄이기 위해 CNC에 GAT와 같은 Graph Neural Network 모델을 활용하여 CN을 효율적으로 관리하기 위함이다.

References

1. Sunitha Safavat, Naveen Naik Sapavath, Danda B. Rawat, “Recent advances in mobile edge computing and content caching”, Digital Communications and Networks, Sep 4, 2020
2. Junaid Shuja, Kashif Bilal, Waleed Alasmary, Hassan Sinky, Eisa Alanazi, “Applying Machine Learning Techniques for Caching in Next-Generation Edge Networks: A Comprehensive Survey”, JNCA, Nov 4, 2020