데이터센터 스위치의 트래픽 분배 기법 비교 및 분석

이준석, 유연호, 양경식, 유혁
고려대학교 정보대학 컴퓨터학과

jslee@os.korea.ac.kr, yhyoo@os.korea.ac.kr, ksyang@os.korea.ac.kr, chuckyoo@os.korea.ac.kr

An Analysis of Traffic Splitting Methods of Datacenter Switches

Junseok Lee, Yeonho Yoo, Gyeongsik Yang, Chuck Yoo
Department of Computer Science and Engineering, Korea University

요 약

클라우드 데이터센터는 인터넷 서비스의 발전으로 방대해지는 고객들의 수요량을 충족하기 위해 호스트 서버들 간 여러 개의 경로가 존재하는 네트워크 토폴로지를 사용한다. 이는 다양한 서비스들이 필요로 하는 네트워크 트래픽 부하를 여러 경로 및 스위치에 분산시켜, 고객들의 요구량을 만족시키는 동시에 네트워크 자원 활용률을 높이기 위함이다. 해당 기법을 트래픽 로드 밸런싱이라고 하며, 지금까지 다양한 기술들이 연구 및 발전되고 있다. 본 논문에서는 데이터센터에서 관리자가 의도한 로드 밸런싱을 정확하게 달성하기 위한 선결 조건인, 스위치 수준에서의 트래픽 분배 기법들을 비교 및 분석한다. 본 논문은 대표적인 트래픽 분배 기법인 해싱과 라운드로빈 기법들을 실제 데이터센터에 가장 흔하게 사용되는 가상 스위치인 Open VSwitch 에 구현한다. 이를 통해, 실제 데이터센터 트래픽을 대상으로 트래픽 분배 정확도와 오버헤드, 워크로드의 평균 전송 완료 시간을 측정하여 비교 및 분석한다.

Ⅰ. 서 론

최근 대부분의 인터넷 서비스 기업들은 그들의 애플리케이션을 클라우드에서 제공한다. 클라우드 환경을 통해 비디오 스트리밍, 웹 서치, 온라인 게임, 머신 러닝 기반 서비스 등 수준 높은 서비스들을 높은 네트워크 대역폭과 낮은 지연 시간으로 고객들에게 공급할 수 있다. 다양한 서비스들이 등장함에 따라 클라우드 서비스의 사용량은 크게 증가하고 있고, 클라우드의 데이터센터는 많은 수의 서버와 거대한 네트워크 트래픽을 수용할 수 있어야 한다.

이를 위해, 데이터센터는 fat tree, VL2 와 같이, 데이터센터 내의 호스트 서버들 간 하나의 경로가 아닌, 여러 경로가 존재하는 네트워크 토폴로지를 사용한다[1]. 데이터센터 관리자는 호스트들이 송수신하는 많은 양의 네트워크 트래픽을 여러 경로로 적절하게 분배하여, 특정 서비스가 네트워크 경로를 독점하는 것을 막을 수 있다. 각각의 경로는 여러 개의 스위치들과 스위치들을 연결하는 링크로 구성되어 있으며, 링크의 상태(e.g., 링크 대역폭 및 링크 장애)를 기반으로 어느 경로에 트래픽을 전송할지 결정한다. 즉, 현대 데이터센터에서는 여러 경로들의 링크들을 활용하여 트래픽을 효율적으로 분배 및 병렬로 전송하게 하는 동시에, 전체 네트워크의 링크 활용률을 높이기 위한 것이 필수 과제이며, 이를 로드 밸런싱이라고 한다[2]. 대표적인 기법 중 하나인 ECMP(equal cost multi-path) 는 존재하는 모든 경로에 동일한 가중치를 두고 트래픽을 분산시킨다.

데이터센터 관리자가 의도한 로드 밸런싱을 잘 구현하기 위해서는 실제 트래픽 전송 및 분배를 수행하는 스위치의 역할이 중요하다. 스위치는 관리자가 작성한 플로우 테이블 또는 라우팅 프로토콜을 기반으로 입력 포트로 받은 트래픽을 경로 별 가중치대로 여러 개의 출력 포트로 내보내는데, 이를 트래픽 분배(traffic splitting)라고 한다. 대표적인 트래픽 분배 기법에는 해싱(hashing)과 라운드 로빈(round robin)이 있다. 이를 기반으로 ECMP와 WCMP(weighted cost multi-path), Presto, CLOVE, TALON 등과 같은 다양한 로드 밸런싱 기법들이 있으며, 현재까지도 활발하게 연구 및 사용되고 있다[3][4].

본 논문은 로드 밸런싱을 위한 대표적인 트래픽 분배 기법인 해싱 기법과 라운드 로빈 기법을 비교 및 분석한다. 본 연구에서는 스위치에서 플로우 단위의 트래픽 분배를 가정하고, 데이터센터에 사용되는 가상 스위치 중 하나인 Open VSwitch(OVS)를 이용하여 두 기법의 성능을 트래픽 분배 정확도, 트래픽 분배에 걸리는 시간, 워크로드의 평균 전송 완료 시간을 측정한다.

Ⅱ. 해싱 기법과 라운드 로빈 기법

해싱 기법은 플로우들을 분류하기 위해 IP 출발 주소, 도착 주소, Port 번호, 전송 프로토콜 등의 플로우의 튜플 값들을 입력 값으로 하는 해시 함수를 사용하며, 해시 값과 경로 별 가중치를 고려하여 플로우가 출력될 경로를 선택하여 내보낸다. 해싱 방식은 복잡도가 낮은 해시 함수를 사용하여도 동일한 플로우, 즉 같은 튜플 값을 가지는 경우 같은 해시 값을 출력하기 때문에, 같은 플로우는 같은 경로로 분배하는 것을 보장한다. 그러나 경로 별 가중치를 고려하여 경로를 선택하는 알고리즘이 스위치 마다 다르며, 스위치 벤더들이 블랙박스로 공개하지 않아, 정확한 동작 방식을 알기 어려운 단점이 있다. 본 연구에서는 OVS 에 기본으로 제공하는 해싱 및 경로선택 방식을 기준으로 한다.

라운드 로빈 기법은 스위치로 들어오는 순서대로 트래픽이 경로들을 차례로 선택하여 내보낸다. 이 때, 트래픽이 들어온 순서만 고려하기 때문에, 같은 플로우 내 패킷이여도 경로가 다르게 선택될 수 있다. TCP 처럼 플로우 내 패킷들의 순서(sequence number)가 중요한 전송 프로토콜에서는 한 플로우 내 패킷들이 다른 경로로 분배

될 경우, 패킷 도착 순서와 전송되는 순서가 다른 패킷 out-of-order 문제가 발생할 수 있다. 패킷 순서가 섞이는 경우, 호스트는 패킷 재전송을 요청하고, 이로 인해 생기는 혼잡을 막기 위해 윈도우 크기를 반으로 줄여, 네트워크 처리량 손실을 발생시킨다. 이를 방지하기 위해 라운드로빈 기법은 플로우 정보와 선택된 경로를 저장하는 별도의 테이블을 관리하여, 동일 플로우에 속하는 패킷이 동일한 경로로 전송될 수 있도록 한다.

Ⅲ. 실험 및 분석

3.1. 실험 환경

토폴로지. 두 분배 기법의 성능을 측정하기 위해 다음과 같은 실험 환경을 설정한다. 두 기법을 OVS 에 구현하고, 네트워크 토폴로지 에뮬레이터인 Mininet 에 적용하여 그림 1 과 같은 토폴로지를 구성한다. 2 개의 말단(edge) 스위치에 각각 컨테이너 호스트 1 개씩 (서버와 클라이언트)을 연결한다. Edge 스위치 사이에 코어 스위치의 개수를 4 개로 구성하여, 4 개의 경로를 설치한다.

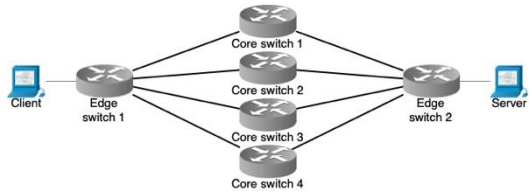


그림 1. 실험 토폴로지

측정지표. 먼저, 각 기법의 트래픽 분배 정확도 및 트래픽 분배에 걸리는 시간을 측정한다. 두 호스트 사이에 실제 데이터 센터 트래픽인 CAIDA 데이터셋을 이용해 TCP 트래픽을 생성한다. 10000 개의 연결에 대해서 실험하고, 경로 별 동일한 가중치를 설정한 경우와 경로 별 다른 가중치를 설정한 경우에서 각각 측정한다. 경로 별 가중치가 다른 경우에서는 가중치를 1 부터 10 사이의 정수 값을 무작위로 설정한다. 이 때 실제 경로 별 선택된 플로우 수를 측정하고, 분배된 비율과 가중치의 평균 오차율을 비교하여 트래픽 분배 정확도를 측정한다. 또한 각 기법이 트래픽 분배에 걸리는 시간을 측정한다.

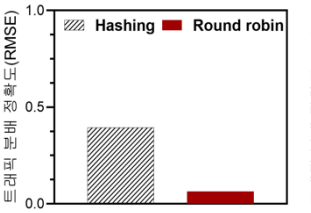
다음으로 데이터센터에서 수집된 웹 서치(검색엔진) 워크로드에 대한 전송 완료 시간을 측정한다. 워크로드를 기준으로 총 400 개의 TCP 통신을 생성한다. 토폴로지는 그림 1 처럼 구성하고, 각 경로의 가중치를 순서대로 1:2:3:4 로 설정한다. Edge 스위치에 서버와 클라이언트의 수는 각각 4 개를 설치하고, 각 클라이언트는 서버로 트래픽을 전송한다.

3.2. 실험 결과

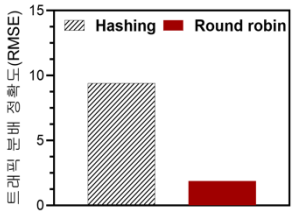
그림 2a 는 가중치가 동일한 경우 해싱과 라운드 로빈의 트래픽 분배 정확도에 대한 평균 오차율을 보여준다. 10000 개의 플로우의 트래픽 분배에 대한 평균 오차율은 해싱에서는 0.3948 이며 라운드 로빈에서는 0.0639 이다. 또한, 그림 2b 는 가중치가 다른 경우의 해싱과 라운드 로빈의 트래픽 분배 정확도에 대한 평균 오차율을 보여준다. 10000 개의 플로우의 트래픽 분배에 대한 평균 오차율은 해싱에서는 9.4201 이며 라운드 로빈에서는 1.8887 이다. 이 결과는 실제로 여러 개의 경로가 존재할 때 해싱 기법을 이용할 경우 트래픽 분배가 정확히 이루어지고 있지 않다는 것을 의미한다.

그림 3 은 해싱과 라운드 로빈의 처리 시간을 보여준다. 해싱의 처리 시간은 0.040s 이며 라운드 로빈의 처리 시간은 1.457s 이다. 이는 해시함수의 결과값인 해시값을 기반으로 빠르게 패킷을 전송하기 때문이다. 반대로, 라운드 로빈은 처리 시간이 크며 이는 트래픽 분배를 할 때 오버헤드가 굉장히 크다는 것을 의미한다.

그림 4 는 400 개의 TCP 플로우로 이루어진 웹 서치 워크로드의 평균 전송 완료 시간을 나타낸 것이다. 해싱의 평균 전송 완료 시간은 3418.906ms 이며 라운드 로빈에서는 2019.288ms 이다. 따라서, 해싱이 라운드 로빈보다 약 1.693 배 높은 평균 전송 완료 시간을 보인다.



(a) 가중치가 동일한 경우



(b) 가중치가 다른 경우

그림 2. 트래픽 분배 정확도 비교

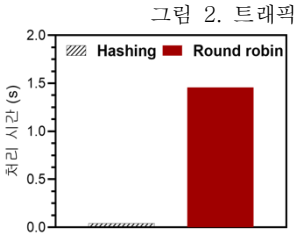


그림 3. 처리 시간 비교

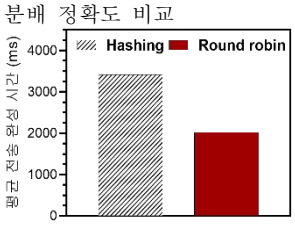


그림 4. 평균 전송 완료 시간

Ⅳ. 결론

본 논문은 로드 밸런싱을 위한 대표적인 트래픽 분배 기법인 해싱과 라운드 로빈 기법을 비교 분석한다. 이를 위해, 데이터센터에 사용되는 가상 스위치 중 하나인 Open VSwitch(OVS)에 각 기법을 구현하고, 트래픽 분배 정확도와 오버헤드, 전송 완료 시간을 비교 분석한다. 실험 결과, 해싱은 낮은 오버헤드를 갖지만 패스 분배 정확도가 낮고, 평균 전송 완료 시간이 높으며 라운드 로빈은 높은 오버헤드를 갖기는 하지만 패스 분배 정확도가 높고, 평균 전송 완료 시간이 낮은 것을 확인하였다. 이러한 측면에서, 향후에는 낮은 오버헤드를 가지면서 높은 패스 분배 정확도와 낮은 평균 전송 완료 시간을 보낼 수 있는 기법에 관한 연구가 필요하다.

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. NRF-2021R1A6A1A13044830)과 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. 2015-0-00280, (SW 스타랩) 성능 및 보안 SLA 보장이 가능한 차세대 클라우드 인프라 SW 개발)을 받아 수행된 연구임.

참 고 문 헌

[1] Pawan Kumar et al. " Issues and challenges of load balancing techniques in cloud computing: a survey", ACM Computing survey, 2019

[2] C. Hopps et al. "Analysis of an Equal-Cost Multi-Path Algorithm." RFC 2992, IETF, November 2000.

[3] Naga Katta et al. "CLOVE: How I learned to stop worrying about the core and love the edge." ACM workshop on Hot topics in networks, 2016

[4] Heesang Jin et al. "TALON: Tenant throughput allocation through traffic load-balancing in virtualized software-defined networks" IEEE International Conference on Informantion Networking, 2019