

플로우의 특성을 이용한 다차원적 응용 트래픽 분석

구영훈, 이성호, 심규석, 정우석, 김성민, 김명섭

고려대학교

{gyh0808, gaek5, kusuk007, hary5832, gogumiking, tmskim}@korea.ac.kr

Multi-dimensional Application Traffic Analysis using Flow Characteristics

Young-Hoon Goo, Sung-Ho Lee, Kyu-Seok Shim, Woo-Suk Jung, Sung-Min Kim,

Myung-Sup Kim

Korea Univ.

요약

날이 갈수록 복잡, 다양해지며 고속화되고 있는 오늘날의 인터넷 환경 하에 효율적인 네트워크 설계 및 구축을 위해서는 구성하고자 하는 네트워크 트래픽의 분석이 선행적으로 이루어져야 한다. 또한, 구축한 네트워크의 올바른 정책 수립과 유지 보수를 위해서는 보다 직관적이며 다차원적인 응용별 트래픽의 특징 분석이 요구된다. 본 논문에서는 플로우의 특성을 이용한 다차원적 응용별 트래픽 분석을 제안한다. 제안하는 방법론을 학내망의 대표적인 5가지 응용에 적용하여 트래픽의 특성을 분석하였다. 본 방법을 통해 응용별의 트래픽 특성을 이전의 독립적인 방법보다 더 직관적으로 파악할 수 있었으며 응용별로 다차원적으로 명확한 차이가 있음을 확인하였다.

I. 서론

오늘날의 네트워크는 전송 속도 증가와 그에 따른 다양한 응용의 증가로 인하여 네트워크 트래픽이 매우 복잡하고 다양해지고 있다. 이러한 환경 하에 효율적인 네트워크 설계와 구축을 위해서는 구성하고자 하는 네트워크의 분석이 선행적으로 이루어져야 한다. 또한, 구축한 네트워크의 트래픽 제어와 유지 보수를 위해서는 지속적인 응용별 트래픽의 특성을 파악하고 올바른 정책을 수립하여야 한다.

이전의 많은 연구[1,2,3]에서는 플로우의 특성을 정의하고 이를 통해 트래픽의 대표 형태를 분석하는 연구가 이루어져왔다. 대표적으로 사용되는 플로우의 특성들로 플로우의 크기(size), 발생 시간(duration), 발생 빈도(rate), 급작스러운 발생빈도(burstiness)가 있으며 이러한 특성들을 동물의 특성에 비유하여 트래픽 대표 형태를 나타낼 수 있다. 발생크기로는 Elephant와 Mice, 발생 시간으로는 Tortoise와 Dragonfly, 발생 빈도로는 Cheetah와 Snail, 마지막으로 급작스런 발생 빈도로는 Porcupine과 Stringray로 비유할 수 있다. 이러한 비유적 특징은 트래픽의 발생 현상을 나타내는데 용이하게 사용될 수 있다.

그러나, 날이 갈수록 복잡, 다양해지는 오늘날의 트래픽 환경에서 응용별 특성을 효율적으로 파악하기 위해서는 다양한 플로우의 특성을 다차원적으로 분석하여 이를 직관적으로 확인할 수 있어야 한다. 이에 본 논문에서는 플로우의 특성을 이용한 다차원적 트래픽 분석을 제안한다. 제안하는 방법론에 대표적으로 상기 기술한 플로우의 발생 크기, 발생 시간, 발생 빈도 3가지 특성을 적용하고 이를 통해 학내망에서 수집한 대표적 5가지 응용에 대해 분석한다. 이를 통해 각 응용의 특성을 시각적으로 쉽게 파악하고 각 응용을 손쉽게 분류할 수 있음을 확인하였다.

본 논문은 본 장의 서론에 이어 2장 본문에서 제안하는 방법론에 대해

설명하고, 3장에서는 본 방법론을 적용한 실험 및 결과를 기술한다. 마지막으로 4장에서 결론 및 향후 연구에 대해 기술한다.

II. 본론

본 장에서는 플로우의 특성을 이용한 다차원적 응용 트래픽 분석 방법론에 대하여 기술한다. 본 방법론은 분석하고자 하는 트래픽을 플로우 단위별 공간상의 좌표로 표시하고 이를 통해 각 응용별로 여러 플로우의 특성을 손쉽게 파악하는 방법이다. 좌표를 표현하는 순서쌍은 플로우의 특성들이 된다. 본 방법론은 특정 네트워크의 트래픽을 직관적으로 분석함이 가능하게 하여 네트워크 설계 및 구축 시 비용, 자원 활용, 용량 계획 등을 수립하는 데 활용할 수 있으며, 응용별 분석을 통해 응용별 트래픽 제어 등의 정책 수립을 용이하게 한다.

본 논문에서는 제안하는 방법론을 위한 플로우의 특성으로 플로우의 발생 크기, 발생 시간, 발생 빈도의 3가지 특성을 이용한다.

트래픽은 플로우의 바이트 양을 기준으로 구분할 수 있는데, 크기가 큰 트래픽은 Elephant, 작은 트래픽은 Mice라 표현한다. 수식 1은 Elephant를 구분해내는 수식으로 n 개의 플로우에 대해서 b_j 는 j 번째 플로우의 바이트 양, \bar{b} 는 평균 바이트 양을 나타낸다. 다음 수식과 일치 하지 않으면 Mice로 구분이 되며 즉, 크기가 작은 트래픽이라 말할 수 있다.

$$b_j > \bar{b} + 3 \sqrt{\frac{\sum_{i=1}^n (b_i - \bar{b})^2}{n-1}}$$

수식 1. Elephant 조건 검사식

또한 트래픽은 플로우의 유지 시간을 기준으로 구분할 수 있다. 유지 시간이 긴 트래픽은 Tortoise, 짧은 트래픽은 Dragonfly라 표현하며, 수식 2는 Tortoise를 구분해내는 수식으로 n 개의 플로우에 대해서 d_j 는 j 번째 플로우의 유지 시간, \bar{d} 는 평균 플로우 유지 시간을 나타낸다. 수식의 조건에 맞지 않는 플로우는 Dragonfly로 구분한다.

$$d_j > \bar{d} + 3 \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}}$$

수식 2. Tortoise 조건 검사식

마지막으로 트래픽은 속도에 따라 구분할 수 있다. 속도가 빠른 트래픽은 Cheetah, 상대적으로 느린 트래픽은 Snail이라 표현하며, 수식 3은 Cheetah를 구분해내는 수식으로 n 개의 플로우에 대해서 r_j 는 j 번째 플로우의 속도, \bar{r} 는 평균 속도를 나타낸다. 수식의 조건에 맞지 않는 플로우는 달팽이로 구분한다.

$$r_j > \bar{r} + 3 \sqrt{\frac{\sum_{i=1}^n (r_i - \bar{r})^2}{n-1}}$$

수식 3. Cheetah 조건 검사식

응용 트래픽은 종류에 따라 다른 트래픽의 특성을 가지게 되며 위의 세 가지 기준에 따라 나눌 수 있다. 각각의 트래픽은 고유한 특성을 갖고 있으며 이를 이용하여 응용 및 서비스별 분류가 가능하다.

III. 실험 결과

본 논문에서는 5가지 응용을 선정하고 제안하는 방법론을 적용하여 분석하였다. 5가지 응용은 Torrent, Dropbox, Daumpot, KakaoTalk, Facebook으로 각각 파일 전송, 파일 공유, 비디오 스트리밍, 인스턴트 메신저, SNS 기능의 대표적 응용들이다. 학내망의 4개의 호스트에서 각각 응용에 대한 3분치의 정답지 트래픽 트레이스를 수집하여 이를 혼합한 트래픽을 1 Set의 응용 트래픽으로 선정하였다. 이를 5가지 응용에 대하여 각 5 Set의 트래픽을 수집하였다. 표 1은 수집한 트래픽의 정량적 정보다.

Set	1	2	3	4	5
Torrent					
Flow	582	924	630	732	579
Packet	237205	173018	204483	182642	230648
MByte	222.6	161.9	225.6	163.6	208.4
Dropbox					
Flow	39	39	45	44	39
Packet	190243	224567	196262	320014	288179
MByte	162.2	193.2	214.5	271.3	244.1
Daumpot					
Flow	1082	785	1099	778	1021
Packet	393015	439786	494106	477361	493130
MByte	373.0	433.0	500.6	472.6	485.9
Kakaotalk					
Flow	38	36	20	18	26
Packet	2119	3335	1449	1940	1324
MByte	1.0	1.8	0.4	1.0	0.5
Facebook					
Flow	136	188	132	158	127
Packet	72026	57508	83672	61714	50786
MByte	68.4	55.6	77.1	60.8	47.1

표 1. 실험에 사용된 트래픽 정량적 정보

표 1의 수집한 각 Set들의 트레이스들을 플로우 단위로 플로우의 발생 크기, 발생 시간, 발생 빈도를 분석하여 Elephant와 Tortoise, Cheetah에 해당하는 플로우의 비율을 Matlab을 통해 3차원 좌표 공간에 표현한 그림은 그림 1과 같다.

각 응용별로 플로우의 특성을 다차원적으로 분석한 결과 Torrent의 경우

Elephant 비율의 평균, Tortoise 비율의 평균, Cheetah 비율의 평균 순으로 2.81%, 10.03%, 1.74%, Dropbox의 경우 59.52%, 20.24%, 17.86%, Daumpot의 경우 2.68%, 3.59%, 2.17%, Kakaotalk의 경우 0.85%, 2.56%, 1.71%, Facebook의 경우 7.17%, 34.35%, 0.43%의 결과를 보였으며 각 응용에 해당하는 트레이스별 분산도 크지 않음을 확인하였다.

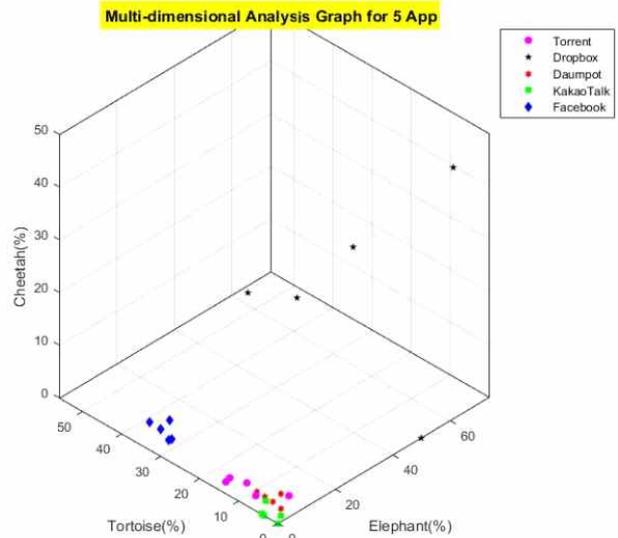


그림 1. 플로우의 특성을 이용한 3차원 그래프 분석

또한, 각 좌표별로 Dropbox를 제외하고 각 응용에 해당하는 트레이스가 3차원 좌표공간상에서 군집을 형성하는 결과를 보였다. Dropbox의 경우는 각 트레이스가 군집을 형성하지 않았지만 모두 Elephant의 비율이 매우 높게 나오는 특성을 가지고 다른 응용과는 명확히 분류됨을 알 수 있었다. 이를 통해 본 분석 방법이 트래픽 분류에 용이하게 사용될 수 있음을 증명하였으며, 본 논문에서는 Elephant, Tortoise, Cheetah의 비율 분석만 기술하였지만, Elephant, Tortoise, Cheetah에 해당하는 발생 크기, 발생 시간, 발생 빈도의 평균 등의 분석을 통해 각 응용 트래픽의 특성 또는 해당 네트워크의 특성을 다차원적으로 파악할 수 있다. 또한, 갑작스러운 발생 빈도, 플로우를 구성하는 패킷간 시간 등의 다른 플로우의 특성을 활용한 분석도 가능하다.

IV. 결론

본 논문에서는 여러 플로우의 특성을 종합적으로 고려하여 직관적으로 트래픽을 분석할 수 있는 방법으로 플로우의 특성을 이용한 다차원적 응용 트래픽 분석을 제안하였다. 본 방법론을 활용하여 학내망에서 수집한 5가지 응용 트래픽의 특성을 다차원적으로 분석하였으며, 분석 결과 중 Elephant, Tortoise, Cheetah의 비율 그래프 분석을 통해 본 방법론이 응용 분류에도 용이하게 사용될 수 있음을 증명하였다. 향후 연구로는 이를 학내망 전체 트래픽에 적용하여 다양한 분석을 할 계획이다.

참고 문헌

- [1] Sandor Molnar, Zoltan Moczar, "Three-dimensional Characterization of Internet Flows", IEEE ICC (2011)
- [2] Peter Megyesi, Sandor Molnar, "Analysis of Elephant Users in Broadband Network Traffic", EUNICE (2013)
- [3] 이성호, 이수강, 김명섭. "네트워크 트래픽 플로우의 3가지 특징을 활용한 다차원적 트래픽 분석", 2015년도 한국통신학회 추계종합학술발표회 (2015)