

# 네트워크 트래픽 플로우의 3 가지 특징을 활용한 다차원적 트래픽 분석

이성호, 이수강, 김명섭  
고려대학교

{gaek5, sukanglee, tmskim}@korea.ac.kr

## Multi-dimensional Traffic Analysis using Three-characteristics in Traffic Flow

Sung-Ho Lee, Su-Kang Lee, Myung-Sup Kim  
Korea Univ.

### 요 약

오늘날 네트워크 기능을 사용하는 다양한 응용이 새롭게 생성되거나 기존에 있던 응용의 트래픽 패턴들이 변화하는 추세가 이어지고 있다. 따라서 네트워크 관리자 입장에서 보다 효율적이고 다차원적인 트래픽 분석 방법이 요구 된다. 그 동안 트래픽의 특징에 따른 다양한 분류 방법과 관점들이 제안되어 왔지만, 트래픽 분석 측면에서 이러한 특징들을 통합해 종합적이고 전체적인 관점에서의 새로운 분석 방법이 필요하다. 본 논문에서는 트래픽 플로우의 3 가지 특징(크기, 지속 시간, 속도)에 따른 프레임에 대해 정의하고 각 프레임을 바탕으로 한 통합적인 트래픽 분석 방법을 제안한다. 본 방법을 통해 이전의 독립적인 트래픽 분석 방법보다 시각적으로 향상된 분석력과 각 특징들 간의 상관 관계에 대해 알 수 있었다.

### I. 서 론

오늘날 네트워크 기능을 사용하는 다양한 응용들이 새롭게 생성되고 있다. 이러한 변화와 함께 네트워크 관리자 입장에서 네트워크 트래픽 모니터링 및 QoS(Quality of Service) 정책 설정에 있어 트래픽 플로우의 특징에 대한 이해는 필수적인 요소이다. 이전의 연구들에서 이미 트래픽 플로우의 특징에 따른 정의가 이루어져 왔다.[1,2,3] 플로우의 크기에 대한 관점에서 전체 플로우 중 몇몇 소수의 플로우가 실제 대부분의 데이터(Byte)를 전송한다는 결과를 얻을 수 있었다. 또 다른 관점인 플로우의 지속 시간 측면에서는 대부분의 플로우는 짧은 반면 긴 지속 시간을 갖는 몇몇 플로우가 전체 데이터의 50-60%를 전송한다는 사실을 알 수 있었다. 마지막으로 플로우의 속도 측면에서 지속 시간과의 상관 관계를 통한 분석 방법이 있었다. 그러나 이러한 연구들은 대부분 플로우의 한 가지 혹은 두 가지 측면에만 초점을 맞춘 연구가 대부분이었다. 이러한 트래픽 데이터의 특징과 특징들 간의 상관 관계에 대해 연구하고 정의하기 위해서는 이런 관점들을 통합해 다차원적인 측면에서의 분석 방법이 요구된다.

따라서 본 논문에서는 이러한 3 가지 특징을 모두 반영할 수 있는 통합적이고 다차원적인 트래픽 분석을 제안한다. 이러한 방법을 위해 각각 플로우의 크기 측면에서 Elephant-mice, 지속 시간 측면에서는 Turtle-dragonfly, 속도 측면에서 Cheetah-snail로 분류 프레임을 정의하고 각 프레임들 간의 상관 관계를 3D-plot 을 이용해 표현한다. 이를 통해 이전의 플로우의 크기, 지속시간, 속도에 따른 데이터 전송량에 대한 연구가 최근 트래픽에도 반영되는지 확인해 볼 수 있었다. 실험을 위해 2010, 2014, 2015년의 교내 트래픽 데이터를 수집했고 이를 통해 보다 자세하고 포괄적인 결과를 얻을 수 있었다.

본 논문은 1 장 서론에 이어, 2 장 본문에서 트래픽 데이터 수집 및 분석 시스템에 대해 간략히 소개하고, 플로우의 프레임 정의를 위한 기준 값 계산 방법에 대해 서술한다. 3 장 실험결과에서 수집한 트래픽에 각 프레임을 적용하고 3D-plot 을 통해 그들간의 상관 관계에 대해 알아본다. 마지막 4 장에서 결론 및 향후 연구에 대해 언급하고 본 논문을 마친다.

### II. 본론

본 장에서는 트래픽 데이터 수집 방법 및 분석 시스템에 대한 간략한 소개와 기존 연구인 트래픽 플로우의 특징에 따른 3 가지 프레임 분류 방법과 기준 값 설정 및 계산 방법에 대해 서술한다.

트래픽 데이터 수집 방법은 교내의 트래픽 데이터 수집 시스템을 사용한다. 외부 인터넷망과 연결된 라우터에서 미러링을 통해 데이터 머신으로 트래픽 정보를 가져오고 이렇게 수집된 데이터를 바탕으로 양방향 플로우 정보를 생성한다. 생성된 플로우 정보를 3D Flow Checker 를 이용해 각 플로우의 크기, 지속 시간, 속도, 프로토콜 등 기본적인 정보를 수집해 csv 파일로 분석 결과를 생성하고, csv 파일을 통계 분석 프로그램인 R 을 사용해 통합하고 플로우의 특징에 따른 프레임을 적용해 3D-plot 을 생성한다. 전체적인 분석 시스템은 그림 1 과 같다.

각 플로우에 3 가지 특징에 대한 프레임을 적용시키기 위해서는 플로우의 크기, 지속 시간, 속도에 따른 기준 값을 정의해야 한다. 기준 값을 정의하고 기준 값 이상의 값을 갖는 플로우에 대해서 앞서 정의한 Elephant-mice, Turtle-dragonfly, Cheetah-snail 에 따라 각각 Elephannt, Turtle, Cheetah 플로우라 정의한다.

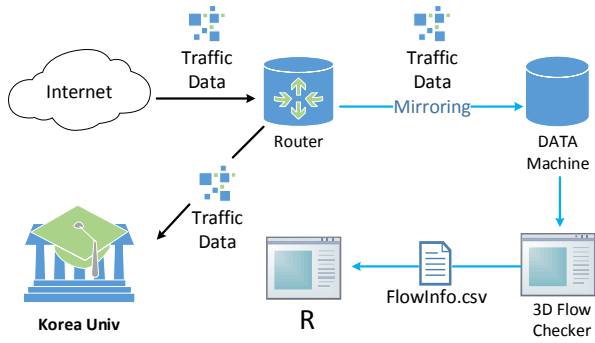


Figure 1. Traffic analysis system

플로우 데이터가 정규 분포를 따른다 가정하고, 경험적인 규칙에 따라 샘플링된 데이터의 평균에 3 표준편차를 더한 값을 기준 값으로 정의한다. 이를 통해 전체 데이터의 상위 0.3%의 플로우가 각각 Elephant, Turtle, Cheetah 가 된다.

$n$  을 전체 플로우 개수,  $x_i$  를  $i$  번째 플로우의 크기, 지속 시간, 속도라 정의하고  $m$  을 각 크기, 지속 시간, 속도의 평균이라 정의했을 때 기준 값 계산은 수식 1 과 같다.

$$Threshold_i = m + 3 \sqrt{\frac{\sum_{i=1}^n (x_i - m)^2}{n-1}} \quad (1)$$

### III. 실험 결과

본 장에서는 수집한 트래픽에 대한 자세한 정보와 프레임을 적용한 결과를 3D-plot 을 통해 나타내고 각 결과들 간의 상관 관계에 대해 알아본다.

수집한 트래픽 데이터에 대한 정보는 표 1 과 같다. 총 3 개의 트레이스로 각 트레이스는 매년 9 월 개강 후 첫째 주에 교내 망에서 3 시간동안 수집한 트래픽이다.

Trace ID	Date	Measurement Duration [hour]	Flows [million]	Packets [million]
DATA 2010_	2010 9.13	3	11.4	421.77
DATA 2014	2014 9.1	3	6.4	367.98
DATA 2015	2015 9.7	3	6.2	499.35

Table 1. Collected traffic data

표 1 의 수집한 트래픽 트레이스에 위에서 정의한 기준 값을 통해 프레임일 적용 시켰을 때 나타난 분석 결과는 그림 2 와 같다.

트래픽의 크기는 2010 년에 가장 크고 2014 년과 2015 년에서는 비슷한 크기를 나타냈다. 평균 지속 시간 측면에서는 2010 년부터 2015 년 까지 지속적으로 증가했다. 플로우의 속도는 2015 년 트레이스의 평균이 가장 빠르게 나타났다. 결과적으로 시간이 지남에 따라 트래픽의 크기는 크게 증가하지 않았지만, 지속 시간과 속도는 지속적으로 증가하는 것을 확인 할 수 있었다.

3 개의 트레이스를 모두 합친 데이터 세트에 Elephant, Turtle, Cheetah 프레임 적용했을 때 플로우의 크기는 평균 12MB 로 분포는 4KB 에서 15MB 사이에 80%이상의 플로우가 포함되었다. Elephant 프레임의 전체 크기는 수집한 데이터 전체 트레이스 크기의 약 80%

이상으로 Elephant 플로우가 대부분의 데이터를 전송하는 것을 알 수 있다.

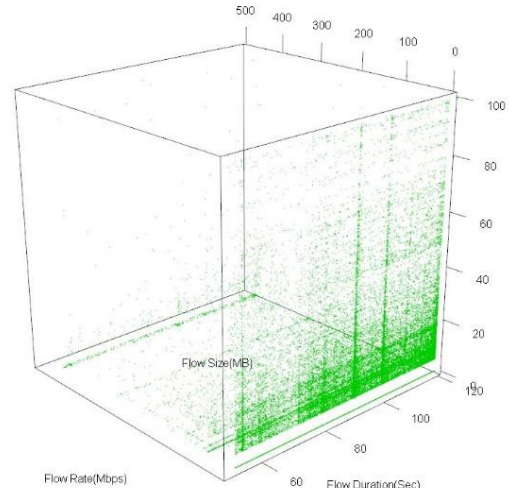


Figure 2. Applying frame result

지속 시간 측면에서는 평균 94 초로 53 초에서 120 초 사이에 전체 Turtle 프레임 플로우가 포함된다. 수집한 전체 트레이스의 지속 시간 평균이 7 초라는 것을 고려했을 때 Turtle 플로우가 많은 양의 데이터를 전송하는 것을 알 수 있다.

마지막으로 속도 측면에서는 평균 17Mbps 로 300Kbps 와 100Mbps 사이에 Cheetah 프레임 플로우의 80%가 포함된다. 전체 트레이스의 플로우 속도 평균은 108Kbps 로 Cheetah 프레임과 비교해 보았을 때, 평균 17 배 빠르고 대부분 지속시간은 길지만 플로우 크기가 상대적으로 작은 특징을 보였다.

### III. 결론

본 논문에서는 트래픽 플로우의 3 가지 특징을 정의하고 이에 따른 프레임을 적용시켜 트래픽을 분류하는 다차원적인 트래픽 분석을 제안하였고, 실험 결과가 기존의 연구 내용과도 부합하는지 살펴보았다. 또한, 3D-plot 을 이용해 플로우의 크기, 지속 시간, 속도 3 가지 측면의 상관 관계에 대해서 알 수 있었다. 본 분석 방법은 추후 트래픽 응용 분석 시스템에 적용시켜 볼 계획이다.

### 참 고 문 헌

- [1] K. Thompson, G. Miller, R. Wilder, "Wide Area Internet Traffic Patterns and Characteristics", IEEE Network Magazine, 11(6):10-23, November 1997.
- [2] D. Papagiannaki, N. Taft, S. Bhattacharyya, P. Thiran, K. Salamatian, C. Diot, "A Pragmatic Definition of Elephants in Internet Backbone Traffic", Proceedings of ACM SIGCOMM Internet Measurement Workshop 2002, pp. 175-176, Marseille, France, November, 2002
- [3] M. R. de Oliveira, A. Pacheco, C. Pascoal, R. Valadas, P. Salvador, "On the Dependencies between Internet Flow Characteristics", Lecture Notes in Computer Science Volume 5464/2009, pp 68-80, 2009