Signature Hierarchy 를 통한 트래픽 분류 시스템의 처리 속도 향상을 위한 연구

최지혁O, 김명섭

고려대학교 컴퓨터정보학과

{jihyeok_choi, tmskim}@korea.ac.kr

Study on Improvement of Processing Speed of Traffic Classification System based on Signature Hierarchy

Ji-Hyeok Choi, Myung-Sup Kim

Korea Univ.

요 약

최근 급격한 인터넷의 발전으로 효율적인 네트워크관리를 위해 응용 트래픽 분석의 중요성이 강조되고 있다. 응용 트래픽을 분석하는 방법 중에 가장 널리 사용되고 있는 방법은 시그니쳐 기반 분석 방법이다. 점점 증가하는 응용에 비례하여 시그니쳐의 수 역시 증가 하고 있지만, 트래픽 분류 시스템의 처리 속도는 향상 되고 있지 않다. 이러한 문제점을 해결하기 위해 본 논문에서는 시그니쳐 계층화를 통해 트래픽 분류 시스템의 처리 속도를 향상 시키는 방법을 제안한다. 기존의 1 차원적인 시그니쳐들을 대표 할 수 있는 대표 시그니쳐라는 개념을 제안하고, 대표 시그니쳐를 통해 시그니쳐들을 계층화 한다. 트래픽 분류 시스템은 계층화된 시그니쳐를 토대로 트래픽 분류를 할 수 있으며, 새로운 매칭 알고리즘을 통해 이전의 방법보다 더 빠르게 트래픽을 분류 할 수 있게 된다.

1. 서론

초고속 인터넷의 보급과 인터넷 기반의 서비스가 다양화됨에 따라 네트워크 관리의 중요성이 강조되고 있다. 네트워크를 효율적으로 관리하기 위해서는 트래픽이 어떠한 응용에서 발생 되었는지를 파악하는 것이 중요하다. 발생된 트래픽이 어떤 응용인지 파악하는 방법 중에 가장 널리 사용되고 있는 방법은 시그니쳐 기반 분석 방법이다. 이러한 시그니쳐 기반 분석 방법을 통해 현재의 트래픽 분류 시스템은 해당 트래픽이 어떠한 응용인지를 분류하고 있다. 하지만 현재의 트래픽 분류 시스템은 점점 증가하는 시그니쳐 수에 비해 처리 속도에는 큰 변화가 없는 상태다[1, 2, 3].

본 논문에서는 갈수록 많아지는 시그니쳐를 대표 시그니쳐라는 개념을 통해 효율적으로 관리 할 수 있고, 트래픽 분류 시스템에 적용하였을 때 처리 속도 또한 빠르게 만드는 방법을 제안한다. 기존에 트래픽 분류 시스템은 시그니쳐들을 하나의 계층으로 보고 1 차원적인 매칭을 하였다[3,4]. 하지만 이러한 방법은 시그니쳐 개수가 많아지면 많아질수록 매칭하는 시간이 계속 증가하게 된다.[5] 본 논문에서는 이러한 단점을 극복하기 위해 시그니쳐들을 계층화 시킨 후, 상위 계층부터 매칭을 하는 방법을 통해서 트래픽 분류 시스템의 처리속도를 향상 시키는 방법을 제안한다.

본 논문은 다음과 같은 순서로 기술한다. 2 장에서는 기존의 트래픽 분류시스템의 매칭 구조와 현재의 매칭 구조를 비교하고 3 장에서는 대표 시그니쳐의 개념에 대해 서술한다. 4 장에서는 매칭 알고리즘에 대해 살펴보고, 마지막으로 5 장에서는 결론 및 향후 연구를 언급한다.

2. 트래픽 분류 시스템의 매칭 구조

* 이 논문은 정부(교육과학기술부)의 재원으로 2010 년도 한국연구재단-차세대정보컴퓨팅기술개발사업(20100020728) 및 2012 년도 한국연구재단(2012R1A1A2007483)의 지원을 받아 수행된 연구임

현재 트래픽 분류 시스템의 매칭 구조는 그림 1 과 같다. 먼저 트래픽을 수집하는 프로그램을 통해 트래픽을 수집하고 분석 단위인 플로우 형태로 변환하는 작업을 수행한다. 변환된 플로우는 트래픽 분류 시스템이 보유하고 있는 시그니쳐들과 순차적으로 매칭을 시작한다. 매칭이 될 경우에는 해당 플로우가 어떤 시그니쳐에 매칭이 됐는지 로그를 남기고 다음 플로우를 매칭하는 구조이다. 하지만 이러한 매칭 구조는 시그니쳐의 수가 증가하면 증가할수록 매칭 횟수 또한 증가하게 된다.[6]

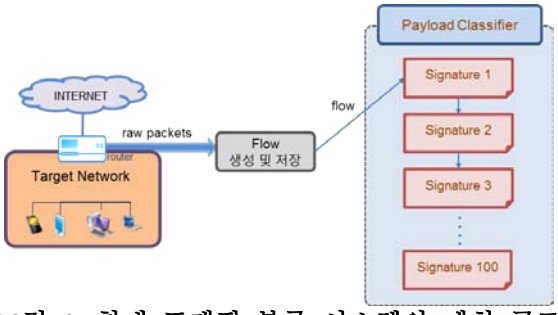

그림 1. 현재 트래픽 분류 시스템의 매칭 구조

위와 같은 문제점을 해결하기 위해 본 논문에서는 현재 보유하고 있는 시그니쳐를 그림 2 와 같이 계층화 시키는 방법을 제안한다.

본 논문에서는 현재 보유하고 시그니쳐들을 가지고 시그니쳐들끼리 서로의 공통점을 찾아서 대표 시그니쳐라는 개념으로 시그니쳐를 새로 만들어 기존의 평면 구조의 시그니쳐가 아닌 계층 구조의 시그니쳐를 만드는 방법을 제안한다. 대표 시그니쳐의 개념과 설명은 3 장에서 자세히 기술한다.

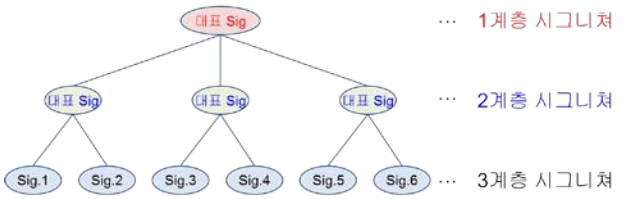

그림 2. 시그니쳐의 계층화

3. 대표 시그니쳐

대표 시그니쳐에 대해 설명하기 전에 본 논문에서 정의하는 시그니쳐에 대해 먼저 기술한다. 여기서 말하는 시그니쳐는 특정 응용을 분류하기 위한 시그니쳐일 수도 있고, 특정 응용의 하나의 기능만을 분류하기 위한 시그니쳐일 수도 있다.

그림 3. 시그니쳐 보유 현황

예를 들어 현재 트래픽 분류 시스템이 가지고 있는 시그니쳐가 그림 3 과 같이 총 8 개라고 가정하자. 8 개의 시그니쳐중에 sig.1 부터 sig.6 까지 6 개의 시그니쳐는 네이버에 관련된 시그니쳐이고 sig.7 과 sig8.은 곰티비에 관련된 시그니쳐라고 하자. 이중에서 sig.1 과 sig.2 는 네이버 맵의 기능들을 탐지하기 위한 시그니쳐이고, sig.3 과 sig.4 는 네이버 메일에 관련된 시그니쳐, 그리고 마지막으로 sig.5 와 sig.6 은 네이버 뮤직에 관련된 시그니쳐이다. 여기서 sig.1 과 sig.2 는 두 개의 다른 시그니쳐로 만들어져 있지만 네이버 맵을 사용 할 때 나온다는 공통점이 존재한다. Sig.3 과 sig.4, 그리고 sig.5 와 sig.6 도 마찬가지로 다른 시그니쳐지만 sig.1 과 sig.2 처럼 서로 공통점을 가지고 있다. 이러한 공통점을 토대로 sig.1, sig2 처럼 서로 다른 시그니쳐를 포함할 수 있는 공통 시그니쳐를 만들 수 있는데 이것을 대표 시그니쳐라고 정의한다.

위와 같은 방법으로 대표 시그니쳐를 만든 후에 다시 대표 시그니쳐끼리 공통점이 있는지 확인한다. sig.1 부터 sig.6 까지에서는 총 세 개의 대표 시그니쳐를 만들 수 있는데, 만들어진 세 개의 대표 시그니쳐는 하나의 공통점이 있다. 바로 Naver 에서 제공하는 서비스들이라는 것이다. 그것은 결국 Naver 에서 제공하는 네이버 맵, 네이버 메일, 네이버 뮤직을 대표하는 세 개의 대표 시그니쳐가 그림 4 처럼 Naver 라는 하나의 대표 시그니쳐로 표현 될 수 있다.

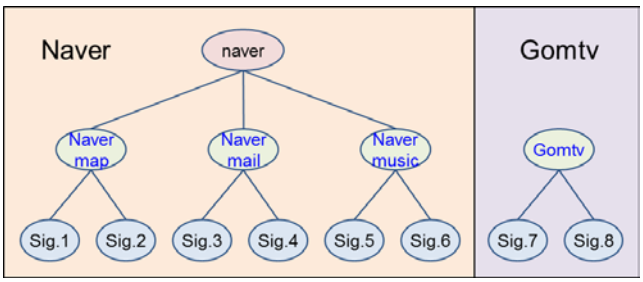

그림 4. Naver 와 Gomtv 의 계층 구조

Naver 와 같은 경우에는 3 계층으로 표현이 가능하지만 곰티비는 2 계층으로 밖에 표현 할 수가 없다. 그 이유는 곰티비에서 제공하는 서비스가 하나 밖에 없기 때문에다. 곰티비가 제공하는 기능들은 방송 다시 보기나 생중계 등 여러 가지가 있을 수 있지만 이것들은 모두 3 계층에 해당되는 시그니쳐들이다. 3 계층을 대표하는 대표 시그니쳐는 만들 수 있지만, 2 계층을 묶을 수 있는 대표 시그니쳐가 존재 하지 않기 때문에 2 계층으로 밖에 표현하지 못한다. 일반적으로 3 계층으로 표현 가능한 것들은 대부분이 대형 포탈 사이트들이다. 대형 포탈 사이트들은 각각 다양한 서비스들을 제공하고 있기 때문에 3 계층으로 표현이 가능하다. 하지만 일반적으로 사용되는 응용 프로그램들은 2 계층까지 밖에 표현하지 못한다.

본 장에서 설명하는 대표 시그니쳐를 추출하고 계층화 하는 작업을 수행하기 위해서는 몇 가지 단계가 있다. 먼저, 대표 시그니쳐를 응용 별로 추출

해야 하기 때문에 평면 구조로 되어있는 시그니쳐들을 응용 별로 분리하는 작업을 해야 한다. 그 다음에는 응용 별로 대표 시그니쳐를 추출하는 작업을 수행한다. 대표 시그니쳐를 추출 하는 방법은 응용 별로 모인 시그니쳐들의 공통된 스트링을 찾는 것이다. 공통된 스트링이 결국 2 계층 대표 시그니쳐가 되고, 2 계층 대표 시그니쳐들끼리 다시 3 계층에서 했던 방법처럼 대표 시그니쳐들끼리 공통된 스트링을 찾아서 1 계층 대표 시그니쳐를 추출할 수 있다.

4. 매칭 알고리즘

본 장에서는 계층화된 시그니쳐를 기반으로 트래픽 분류 시스템에 적용 하였을 때 트래픽 분류 시스템의 매칭 알고리즘을 기술 한다. 그림 5 는 시그니쳐 계층구조 기반 트래픽 분류 시스템의 매칭 알고리즘이다.

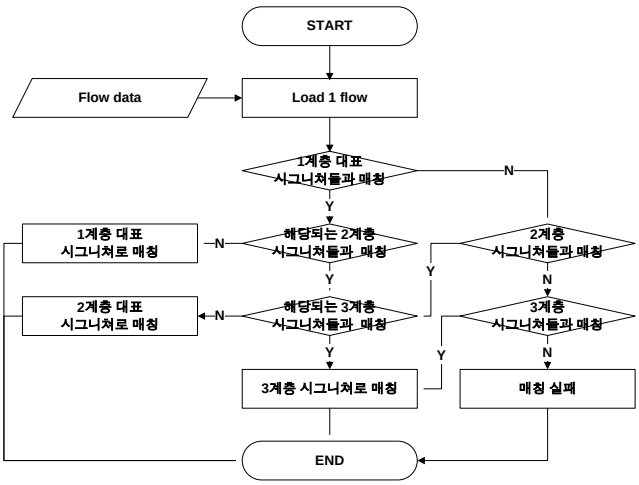

그림 5. 매칭 알고리즘

먼저, 입력으로 플로우 하나가 들어오게 되고 해당 플로우는 1 계층의 대표 시그니쳐들과 우선적으로 매칭 검사를 시작한다. 1 계층 대표 시그니쳐와 매칭이 될 경우 1 계층의 하위에 있는 2 계층 대표 시그니쳐들과 매칭 검사를 하게 되고, 2 계층의 대표 시그니쳐와도 매칭이 되면 2 계층의 하위에 있는 3 계층 시그니쳐들과 마지막으로 매칭 검사를 한다. 만약에 1 계층 대표 시그니쳐와 매칭되지 않을 경우에는 2 계층에 존재하는 모든 대표 시그니쳐와 매칭 검사를 하게 된다. 그래도 매칭이 되지 않을 경우에는 3 계층에 존재하는 모든 시그니쳐와 매칭 검사를 하게 된다. 3 계층 시그니쳐와도 매칭이 안될 경우에는 해당 플로우는 결국 매칭에 실패하게 된다. 그리고 만약 1 계층 대표 시그니쳐와 매칭이 되고 해당되는 2 계층 대표 시그니쳐까지도 매칭이 되지만 해당되는 3 계층 시그니쳐와 매칭이 되지 않을 경우 해당 플로우는 2 계층의 대표 시그니쳐와 매칭이 된 것으로 판단하고 매칭 검사를 종료된다. 위와 같은 알고리즘을 트래픽 분류 시스템에 적용하면 매칭 속도 측면에서 이전의 분류 시스템보다 더 빨라질 수 있다.

예를 들어 트래픽 분류 시스템이 가지고 있는 시그니쳐가 1000 개라고 가정하자. 그림 1 과 같이 매칭을 한다면 worst cast 로 총 1000 번이 걸릴 것이다. 하지만 그림 6 과 같이 본 논문에서 제안하는 계층화된 방법을 사용하면 매칭 검사 횟수를 훨씬 줄일 수 있다.

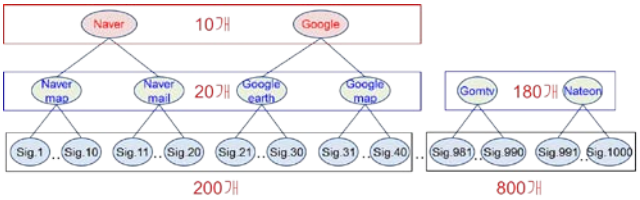

그림 6. 시그니쳐 1000 개의 계층화

그림 6 은 현재 보유하고 있는 시그니쳐로 구축한 시그니쳐 계층 구조이다. 먼저 3 계층 시그니쳐는 총 1000 개로 구성되어 있고, 이것을 바탕으로 대표 시그니쳐를 추출 한 결과 2 계층 시그니쳐는 200 개가 된다. 그리고 마지막으로 2 계층 대표 시그니쳐들을 바탕으로 추출한 1 계층 시그니쳐는 총 10 개이다. Naver 맵을 사용하여 발생 된 플로우가 입력으로 들어왔다고 가정하였을 때, 트래픽 분류 시스템은 맨 처음 1 계층 대표 시그니쳐들과 매칭을 한다. 1 계층 대표 시그니쳐가 총 10 개이기 때문에 10 번의 매칭을 하게 된다. Naver 라는 1 계층 대표 시그니쳐가 존재 하기 때문에 Naver 에 해당 되는 2 계층 대표 시그니쳐들과 바로 매칭을 시작하게 된다. Naver 에 해당되는 대표 시그니쳐는 2 개 밖에 없기 때문에 두 번만 매칭을 한 후, 3 계층 시그니쳐로 내려가서 마지막으로 매칭을 하게 된다. 3 계층 Naver map 시그니쳐는 10 개가 존재 하기 때문에 10 번의 매칭을 더 진행하게 된다. 위와 같은 매칭 과정을 다 합쳐도 표 1 과 같이 22 번밖에 되지 않지 않기 때문에 기존의 분류 시스템에서의 매칭 횟수인 1000 번과 약 50 배 가량의 매칭 속도 차이가 날 수 있다.

표 1. 평면구조와 계층구조 성능 비교

	평면구조	계층구조
시그니쳐 수	1000 개	1000 개
매칭 횟수	1000 번	22 번

6. 결론 및 향후 과제

점점 늘어나는 응용의 수에 비례하여 시그니쳐의 수 또한 증가하고 있다. 하지만 현재의 트래픽 분류 시스템은 점점 증가하는 시그니쳐 수에 비하여 처리 속도 측면에서는 발전이 없는 상태이다.

본 논문에서는 시그니쳐 계층화를 통해 시그니쳐

를 효율적으로 관리하고 트래픽 분류 시스템의 처리속도 또한 빠르게 만드는 방법을 제안하였다. 제안한 방법을 통하여 기존의 트래픽 분류 시스템의 매칭 구조 보다 더욱 빠르게 트래픽 분류가 가능하게 될 것이다.

향후 연구로써는 본 논문에서 제안한 방법들이 타당하다는 것을 실험을 통해 증명할 것이고, 구체적인 수치를 통해 기존의 방법보다 발전된 방법이라는 것을 증명할 예정이다.

7. 참고 문헌

[1] Myung-Sup Kim, Young J. Won, and James Won-Ki Hong, "Application-Level Traffic Monitoring and an Analysis on IP Networks," ETRI Journal, Vol.27, No.1, Feb. 2005, pp.22-42."

[2] Jun-Sang Park, Jin-Wan Park, Sung-Ho Yoon, Young-Seok Oh, Myung-Sup Kim, "Development of signature Generation system and Verification Network for Application Level Traffic classification", Conference of Korea Information Communication Society, Apr. 23-24, 2009, pp. 1288-1291.

[3] Liu, Hui Feng, Wenfeng Huang, Yongfeng Li, Xing " Accurate Traffic Classification", Networking, Architecture, and Storage, 2007. International Conference.

[4] Risso, F. Baldi, M. Morandi, O. Baldini, A.Monclus, P. "Lightweight, Payload-Based Traffic Classification:An Experimental Evaluation," Proc. of the Communications, 2008. ICC '08. IEEE International Conference, 2008.

[5] 윤성호, 김명섭, "인터넷 트래픽 분류 시스템의 처리속도 향상에 관한 연구", 2012 년도 한국통신학회 동계종합학술발표회, 용평, 강원도, Feb. 8-10, 2012.

[6] 박준상, 박진완, 윤성호, 김명섭, "페이로드 시그니처 기반 응용 트래픽 분류 알고리즘의 성능 향상", 2010 년도 한국통신학회 하계종합학술발표회, 제주, 라마다프라자 제주호텔, June. 23-25, 2010, pp.482.