Payload Signature Hierarchy 를 통한 트래픽 분류 시스템의 처리 속도 향상

최지혁, 김명섭
고려대학교

{jihyeok_choi, tmskim}@korea.ac.kr

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

Ji-Hyeok Choi, Myung-Sup Kim
Korea Univ.

요 약

최근 급격한 인터넷의 발전으로 인해 효율적인 네트워크 관리가 필요하다. 그 중에서 네트워크 트래픽의 응용을 분석하는 트래픽 분석은 네트워크 관리를 위해 꼭 필요한 기술이다. 트래픽을 분석하는 방법 중에 가장 널리 사용되고 있는 방법은 시그니쳐 기반 분석 방법이다. 하지만 시그니쳐 기반 분석방법은 시그니쳐의 수가 증가하면 할수록 처리 속도가 점점 느려지는 단점이 있다. 본 논문에서는 시그니쳐의 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 분류하는 방법을 제안한다.

Ⅰ. 서 론

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. 시그니쳐 기반 분석방법은 응용마다 가지고 있는 고유의 특징을 기반으로 응용을 판별하는 방법이다. 하지만 이러한 시그니쳐 기반 분석은 응용의 수가 늘어날 수록 시그니쳐의 수 또한 증가하게 되어 트래픽 분석 시스템에 적용 하였을 때, 분석 속도가 상당히 오래 걸린다는 단점이 있다[1].

본 논문에서는 갈수록 많아지는 시그니쳐를 대표 시그니쳐라는 개념을 통해 효율적으로 관리 할 수 있게 하고, 또한 트래픽 분류 시스템에 적용하였을 때 처리 속도 역시 빠르게 만드는 방법을 제안한다. 기존의 트래픽 분류 시스템을 살펴 보면 시그니쳐들을 하나의 계층으로 보고 1 차원적인 매칭을 하였다[2]. 하지만 이러한 방법은 시그니쳐 개수가 많아지면 많아질수록 매칭하는 시간이 계속 증가하는 단점이 있다. 하지만 본 논문에서는 이러한 단점을 극복하기 위해 평면적인 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.

Ⅱ. 대표 시그니쳐와 시그니쳐 계층화 방법

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

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

그림 1. 평면적인 시그니쳐 구조

그림 1 은 트래픽 분류 시스템이 현재 가지고 있는 총 1000 개의 시그니쳐 평면 구조를 나타낸다. 만약에 사용자가 네이버 메일의 다운로드 기능을 사용 한다고 가정하였을 때, 트래픽 분류 시스템은 네이버 메일 다운로드에 관련된 시그니쳐를 찾기 위해 총 1000 번의 매칭을 시도 한다. 시그니쳐의 수가 많아 지면 많아 질수록 매칭하는 빈도수가 증가하기 때문에 이것은 결국 트래픽 분류 시스템의 속도을 저하시키는 요인이 된다.

이러한 문제점을 해결하기 위해서 본 논문에서는 대표 시그니쳐를 통한 시그니쳐 계층화 방법을 제안한다. 대표 시그니쳐를 추출하기 위해서는 먼저 1000 개의 시그니쳐들을 응용 별로 분리하는 작업을 해야 한다. 그 다음에 응용 별로 대표 시그니쳐를 추출한다. 대표 시그니쳐를 추출 하는 방법은 먼저 응용 별로 정답지 트래픽을 모은 후에, 각각의 플로우에서 공통적으로 나타나는 스트링을 찾는 것이다. 이렇게 찾아진 스트링은 해당 응용을 대표할 수 있는 대표 시그니쳐라고 부른다. 또한 대표 시그니쳐들끼리 다시 대표 시그니쳐를 뽑는 경우도 있는데, 이는 네이버나 구글 처럼 대형 포탈 사이트에서 제공하는 응용이 많을 경우 각각의 응용들의 대표 시그니쳐를 뽑은 후에 대표 시그니쳐들끼리 비교를 하여 공통적으로 나타나는 스트링을 대표 시그니쳐로 만들게 된다. 그림 2 를 보면 sig.1 부터 sig.4 까지는 하나의 응용에서 나타나는 시그니쳐 이고, sig.1 과 sig.2 의 정답지 트래픽에서 공통적으로 나온 스트링이 sig.1 과 sig.2 의 대표 시그니쳐가 된다. 같은 방법으로 sig.3 과 sig.4 의 대표 시그니쳐도 만든 후, 만들어진 두 개의 대표 시그니쳐를 비교하여 같은 스트링이 있을 경우 다시 대표 시그니쳐를 만들 수 있다. 계층적인 구조의 장점은 같은 응용의 시그니쳐들끼리 모여 있기 때문에 관리가 용이하고, 매칭 속도 또한 빨라질 수 있다. 기존에는 1000 개의 시그니쳐가 있을 경우 1000 번의 매칭 횟수를 갖지만, 계층 구조의 경우 최 상위 계층부터 매칭을 하는 구조를 갖고 있기 때문에 훨씬 적은 매칭 횟수를 보이게 된다. 자세한 매칭 알고리즘은 3 장에 기술한다.

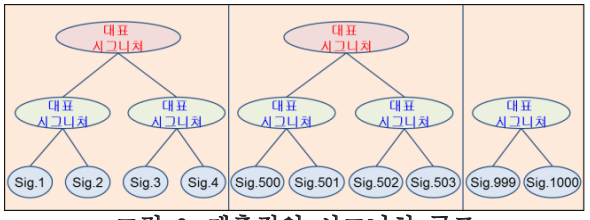

그림 2. 계층적인 시그니쳐 구조

Ⅲ. 매칭 알고리즘

본 장에서는 계층화된 시그니쳐를 기반으로 트래픽 분류 시스템에 적용 하였을 때 트래픽 분류 시스템의 매칭 알고리즘을 기술 한다. 그림 3 은 시그니쳐 계층구조 기반 트래픽 분류 시스템의 매칭 알고리즘이다.

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

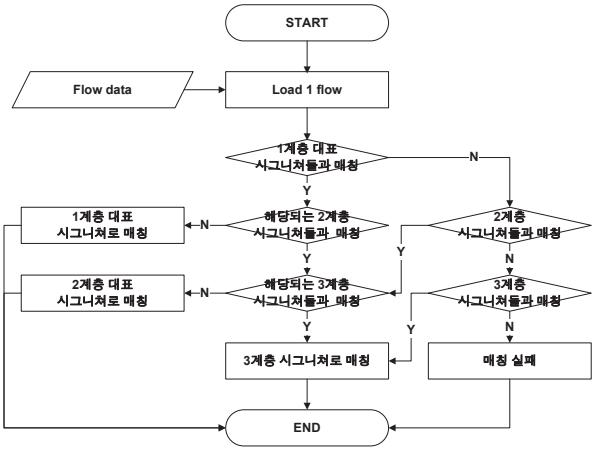

그림 3. 매칭 알고리즘

Ⅳ. 결론 및 향후 연구

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 대표 시그니쳐라는 개념을 통해 시그니쳐들을 계층화하고 이를 이용하여 시그니쳐들을 효율적으로 관리할 뿐만 아니라 트래픽 분류 시스템의 처리속도 또한 빠르게 만드는 방법을 제안하였다. 제안한 방법을 통하여 기존의 트래픽 분류 시스템의 매칭 구조보다 더욱 빠르게 트래픽 분류가 가능하다.

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

참 고 문 헌

[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] 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.