학내망에서 발생한 페이스북 SNS 트래픽 분류 및 변화량 분석

윤아현, 이수강, 김명섭
고려대학교
{dbsdkgus94, sukanglee, tmskim}@korea.ac.kr

Classification and Analysis of the Facebook SNS Traffic in Campus Network

Ah-Hyeon Yoon, Su-Kang Lee, Myung-Sup Kim
Korea Univ.

요 약

오늘날 스마트폰의 대중화와 다양한 종류의 소셜 네트워크 서비스(이하 SNS)의 등장으로 SNS의 이용자수가 증가하고 있으며 이로 인해 SNS 트래픽 또한 기하급수적으로 증가하고 있다. 이러한 이유로 SNS 트래픽을 분류하고 SNS 트래픽의 특징을 파악하는 것이 중요하다. 본 논문에서는 대표적인 SNS중 하나인 페이스북 서비스를 분류할 수 있는 헤더 시그니쳐를 정의하고, 이를 바탕으로 학내망 페이스북 트래픽의 변화량을 분석하였다. 페이스북 헤더 시그니쳐 검증 과정에서 페이스북 IP대역이 아닌 Akamai, BTN, NTT와 같은 CDN 업체의 IP대역에서도 페이스북의 페이로드 패턴이 발생하는 것을 확인하였다. 이렇게 생성한 페이스북 헤더 시그니쳐와 CDN 헤더 시그니쳐를 실제 학내망에서 발생한 트래픽에 적용하여 2010년도부터 2015년도까지의 변화량을 분석하였다. 결과적으로 시간이 지남에 따라 페이스북 트래픽이 증가하였고, CDN 트래픽 또한 페이스북 트래픽의 증가량과 관계가 있는 것을 알 수 있었다.

Ⅰ. 서 론

오늘날 스마트폰의 대중화와 다양한 유형의 소셜 네트워크 서비스의 등장으로 SNS의 이용자수가 기하급수적으로 증가하고 있다. 페이스북은 페이스북 자체 2015년 2분기 실적에 따르면 월 활동 사용자수가14억 9천만 명을 넘어섰고, 일 활동 사용자는 9억 6천을 기록한 세계 최대의 소셜 네트워크 서비스이다. 대다수 기업이나 웹 마켓 업체들은 페이스북을 온라인 마케팅에 이용하는 사례가 점점 많아 졌고 페이스북을 자사 웹사이트의 유입통로로 이용하는 경우가 많아짐에 따라 페이스북의 응용 트래픽 양이 급격히 증가하고 있는 추세이다. 이러한 추세에 맞춰 학내에서 소셜 네트워크 서비스를 이용하거나 또는 소셜 네트워크 서비스의 부정적인 영향을 확인하기 위해 실제 학내 망의 트래픽에서 SNS트래픽의 특징을 파악하는 것이 중요하다.

인터넷 트래픽에서 특정 응용 및 서비스를 탐지하는 방법들은 사용하는 시그니쳐의 종류에 따라 나누어진다. 시그니쳐 종류에 따라 페이로드[1] 기반 분석방법, 헤더 정보 기반 분석 방법[2], 통계 정보 기반 분석 방법[3]이 있다. 페이로드 기반 분석 방법은 패킷의 응용계층 페이로드 데이터로부터 고유한 스트링을 추출하고 이를 기반으로 스트링매칭을 통하여 트래픽의 응용을 결정하는 방법이다. 페이로드 기반 시그니쳐 분석은 다른 방법론에 비해 정확도와 분석률이 높다는 장점이 있지만 패킷의 페이로드 중 특정 문자열을 찾는 과정에서 오버헤드가 발생하여 분석시간이 다른 분석 방법들에 비해 속도가 느리다. 헤더기반 분석 방법은 특정 응용에서만 사용되는 헤더 정보들의 조합을 사용하여 시그니쳐를 생성하고 분류하는 방법이다. 헤더 정보란 Flow를 정의하는 기본단위인 source address/port, destination address/port, transport protocol이 있다. 이 방법은 헤더만을 비교하여 분석함으로서 페이로드 기반 분석 방법보다 속도가 빠르다는 장점이 있다.

본 논문에서는 헤더 정보 기반의 분석방법을 이용하여 페이스북 트래픽을 분류하였으며 이를 검증하기 위해 헤더 시그니쳐로 분류된 Flow에 페이로드 분석 방법을 이용하여 페이스북 서버인지 검증하였다. 본 논문의 2장에서는 페이스북 트래픽 분석을 위한 헤더 시그니쳐 생성 과정과 생성된 시그니쳐의 검증 과정을 설명하고 3장에서는 검증이 완료된 시그니쳐를 실제 학내망 트래픽에 적용한 실험 결과를 서술한다. 마지막으로 4장에서는 결론과 향후 연구를 기술한다.

Ⅱ. 연구내용

본 장에서는 페이스북 트래픽의 분류를 위한 헤더 시그니쳐 생성 과정을 설명하고, 생성된 헤더 시그니쳐의 검증을 통해 페이스북 시그니쳐로 정한 헤더 시그니쳐 리스트를 기술한다.

1. 헤더 시그니쳐 생성

전체 트래픽에서 페이스북 트래픽을 분류하기 위해 헤더 시그니쳐를 생성하는 과정은 다음과 같다. 윈도우즈 환경에서 프로세스 단위로 패킷 수집이 가능한 Network Monitor 프로그램을 사용하여 페이스북에 접속하였을 때의 해당 프로세스(Internet Explorer)의 패킷을 수집한다. 수집된 패킷파일은 해당 시간대에는 Internet Explorer를 사용하여 페이스북 서비스만 이용하였다. 수집된 패킷의 모든 IP를 국가 인터넷주소관리기관인 “whois”에 모두 검색하여 나온 IP의 범위, 소유주 명, 관리기관 명, 도메인 이름 등 을 참고하여 페이스북과 관련된 IP주소의 범위을 찾고, 이를 페이스북 헤더 시그니쳐로 정의하였다. 또한 수집된 트래픽에는 IP주소의 관리기관 명이나 소유주 명이 페이스북이 아닌 IP대역대도 존재하였고, 해당 대역대는 대부분 페이스북을 위한 CDN 서비스를 제공하는 CDN 업체의 서버대역이었다. CDN에 관련한 내용은 다음 절에서 자세히 서술한다.

이 논문은 BK21 플러스 사업(No. T1300573), 2015년도 정부(교육부)의 재원으로 한국연구재단의 기초연구사업(No.2015R1D1A3A01018057), 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터 지원(No.B0101-15-0300)을 받아 수행된 연구임.

2. 헤더 시그니쳐 검증

생성한 헤더 시그니쳐를 검증하기 위해 패킷의 페이로드 내에 페이스북 트래픽 이라고 단정 지을 수 있는 특정 패턴(“facebook”,“fbcdn”, “fbstatic”)을 발견하였고, 이것을 페이로드 시그니쳐로 정의하였다.

도메인 소유자	IP주소 범위 시작	IP주소 범위 끝
Facebook	xxx.13.64.0	xxx.13.127.255
	xxx.252.64.0	xxx.252.127.255
	xxx.171.244.0	xxx.171.255.255
	xxx.220.144.0	xxx.220.159.255
Akamai	xxx.192.0.0	xxx.192.255.255
	xxx.52.0.0	xxx.52.63.255
	xxx.16.0.0	xxx.17.255.255
	…	…
BTN	xxx.217.21.9	xxx.217.21.43
NTT	xxx.87.182.191	xxx.87.182.255
	xxx.254.0.0	xxx.254.255.255

표 1. 수집한 트래픽에서 생성한 헤더 시그니쳐

표 1은 수집한 트래픽에서 페이스북과 관련된 IP주소를 바탕으로 생성한 헤더 시그니쳐를 나타낸 표이다. “whois”검색결과 도메인 소유주나 관리 기관명이 페이스북인 것과 CDN 업체(Akamai, BTN, NTT)인 것들을 추려 작성하였다.

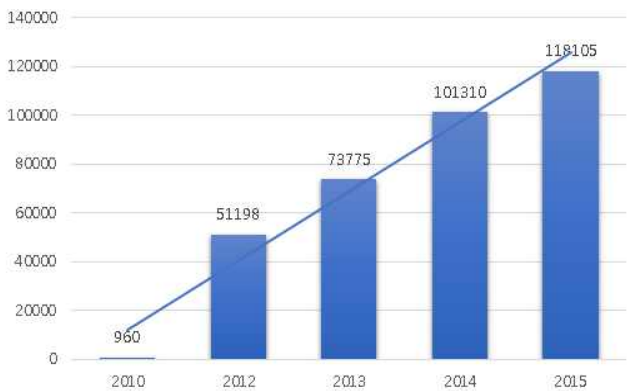
	FB	AK	BTN	NTT	Total
전체 Flow	230	562	175	233	1405
전체 Pkt(K)	156	152	13	21	360
전체 Byte(MB)	127	130	9	14	294
패턴있는 Flow	219	493	175	233	1139
패턴있는 Pkt(K)	156	145	13	21	336
패턴있는 Byte(MB)	127	125	9	14	294

표 2. 페이로드 시그니쳐가 발견된 트래픽의 정량적 수치

표 2는 생성한 헤더 시그니쳐로 분류된 Flow의 패킷들 중에서 페이로드 시그니쳐가 발견된 트래픽의 정량적 수치를 나타낸 것이다. 표2와 같이 페이스북 IP대역대가 아닌 CDN업체의 IP대역대 에서도 페이스북 페이로드 시그니쳐가 발생한 것을 알 수 있다. 따라서 페이스북이 CDN 서비스를 사용하여 서비스를 제공하는 것을 알 수 있었다. Flow기준 패턴이 있는 각각 페이스북 Flow는 전체의 95%, Akamai는 88% BTN과 NTT는 100%를 차지하는 것을 확인하였다. 결과적으로 표1에서 나타낸 헤더 시그니쳐를 페이스북 트래픽 분류를 위해 학내망에서 발생한 트래픽에 적용하였다.

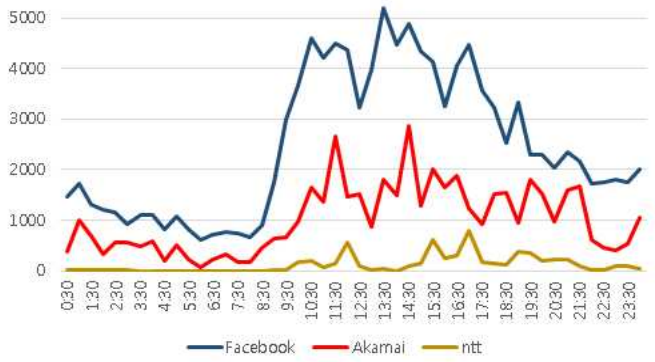
Ⅲ. 실험 결과

본 장에서는 검증이 완료된 헤더 시그니쳐를 2010년부터 2015년도 평일(5일)에 학내망에서 발생한 트래픽에 적용한 결과를 나타낸다.



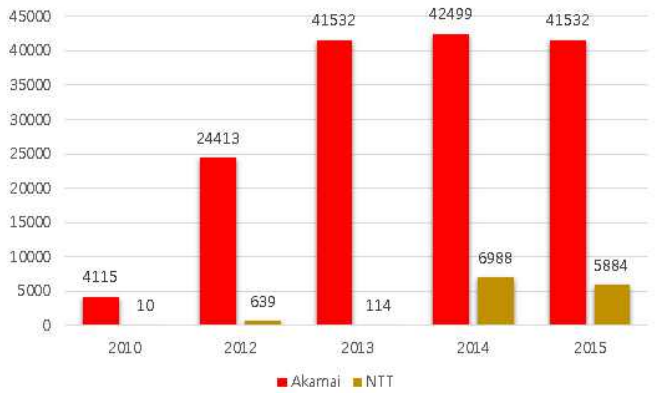
그래프 1. 연도별 페이스북 트래픽 변화량

그래프 1은 학내 망에서 2010년부터 2015년까지의 트래픽에서 페이스북으로 분류된 트래픽의 변화량을 나타낸 것이다. 그래프 1과 같이 학내망의 페이스북 트래픽은 꾸준히 증가하는 것을 알 수 있다.



그래프 2. 시간대 별 페이스북과 CDN트래픽의 Flow개수

그래프 2는 2015년 6월 17일 수요일의 트래픽을 Facebook과 CDN으로 분류된 Flow수를 30분 간격으로 나타낸 것이다. 그래프 2의 결과로 봤을 때, 10시부터 19시까지 페이스북 트래픽이 많이 발생한 것을 알 수 있으며, CDN 트래픽(Akamai, NTT)도 페이스북 트래픽 변화량과 비슷한 양상을 보였다.



그래프 3. CDN트래픽의 연도별 변화량

그래프 3은 CDN(Akamai, NTT) 트래픽의 연도별 변화량을 나타낸 것이다. 그래프 3에서와 같이 Akamai와 NTT는 시간이 변화함에 따라 그래프 1의 페이스북 변화량과 마찬가지로 트래픽량이 증가하는 것을 알 수 있다.

Ⅳ. 결론 및 향후 연구

본 논문에서는 페이스북 트래픽을 분류하기 위한 헤더 시그니쳐를 생성하고 이를 실제 학내망에서 발생한 트래픽에 적용하여 페이스북 트래픽과 CDN 트래픽의 변화량을 확인하였으며 결과적으로 페이스북 트래픽과 CDN 트래픽은 꾸준히 증가하였다. 향후 연구로는 학내에서 주로 사용하는 응용서비스와 CDN 트래픽의 상관관계를 분석할 예정이다.

참 고 문 헌

[1] 박준상, 윤성호, 안현민, 김명섭, "페이로드 시그니쳐 기반 인터넷 트래픽 분류", 2014년 통신망운용관리 학술대회 (KNOM 2014), 충남대학교, 대전, May. 15-16, 2014, pp.10-14.

[2] 윤성호, 김명섭, "헤더 기반 인터넷 응용 트래픽 분석을 위한 시그니쳐 관리 방법에 관한 연구", 한국인터넷정보학회 Vol.12 No.6, 12. 2011, pp. 19-33.

[3] 박진완, 윤성호, 박준상, 이상우, 김명섭, "통계 시그니쳐 기반의 응용 트래픽 분류", 통신학회논문지 Vol.34 No.11, , Nov. 2009, pp.1234-1244.