인터넷 서비스 식별을 위한 헤더정보 기반 자동 시그니쳐 명명 시스템

이수강, 윤성호, 심규석, 김명섭

고려대학교

{sukanglee, sungho_yoon, kusuk007, tmskim}@korea.ac.kr

Proposal of Header Signature Management System for Internet Service Identification

Lee Su-Kang, Yoon Sung-Ho, Shim Kyu-Seok, Kim Sung-Min, Kim Myung-Sup

Korea Univ.

요 약

오늘날 인터넷 장비의 발달과 새로운 응용들의 출현으로 인해 인터넷 트래픽의 양은 급격하게 증가하고 있다. 이러한 상황 속에서 안정적인 인터넷 서비스를 제공하기 위한 네트워크 관리 정책은 정확한 인터넷 응용 및 서비스 탐지를 기반으로 행해지며 그 중요성이 점차 증가하고 있다. 인터넷 응용 및 서비스 탐지를 위한 헤더 정보 기반의 탐지 방법은 다른 방법들 보다 쉽고 빠르게 특정 응용 및 서비스를 탐지할 수 있는 방법이다. 본 논문에서는 헤더 정보를 기반의 시그니쳐 자동 네이밍 시스템을 제안한다. 우리는 본 논문에서 제안한 시스템을 실제 학내망 네트워크에 적용한 결과 "nslookup" 명령 또는 "whois" IP 검색에서는 얻을 수 없는 실제 Content provider의 고유한 URI 정보를 얻을 수 있었다. 결론적으로 본 논문에서 제안한 시스템을 특정 네트워크에 적용함으로써 실제로 제공되는 서비스나 컨텐츠의 고유한 URI 정보를 이용하여 해당 네트워크에서 발생한 트래픽의 유형과 헤더 정보 쌍을 얻을 수 있었다.

I. 서 론

오늘날 인터넷 장비의 발달과 새로운 인터넷 응용 서비스의 출현으로 인해 다양한 인터넷 응용 트래픽의 양이 급격히 증가하고 있는 추세이다. 이러한 상황 속에서 안정적인 인터넷 서비스를 제공하기 위한 네트워크 관리 정책은 신속하고 정확한 인터넷 응용 및 서비스 탐지를 기반으로 행해지며 그 중요성이 점차 증가하고 있다. 인터넷 트래픽에서 특정 응용 및 서비스를 탐지하는 방법으로는 헤더 정보 기반 분석 방법, 페이로드 기반 분석방법, 통계 정보 기반 분석 방법이 있다.

인터넷 응용 트래픽 분석을 위한 분석 방법들은 사용하는 시그니쳐의 종류에 따라 나누어진다. 통계 정보 기반 분석방법은 발생한 트래픽의 통계 정보를 이용하는 방법이다. 패킷의 크기, 전송 방향, 패킷의 순서, 패킷 전송 시간 등의 통계 정보를 특정 응용을 식별할 수 있는 시그니쳐로 사용한다. 통계정보를 이용한 분석 방법은 암호화된 트래픽을 효과적으로 분석할 수 있다. 페이로드 분석 기반 분석방법은 발생한 트래픽의 패킷 페이로드 정보를 이용하는 방법이다. 패킷의 페이로드에는 특정 응용이나 서비스를 식별할 수 있는 고유한 문자열이 포함되어 있으며 이러한 문자열을 페이로드 시그니쳐로 사용하여 응용이나 서비스를 탐지한다. 페이로드 기반 분석 방법은 여러 분석 방법들 중 가장 정확한 분석 결과를 나타낸다. 하지만 패킷의 페이로드 중 특정 문자열을 찾는 과정에서 오버헤드가 발생하여 분석 시간이 다른 분석 방법들에 비해 분석 속도가 느리다. 또 다른 방법으로는 헤더 정보 기반의 분석 방법이 이다. 헤더 정보란 트래픽을 발생시킨 Host들의 출발지 IP, 출발지 Port. 도착지 IP, 도착지 Port, Protocol등이 있다. 헤더 정보 기반의 분석 방법은 이러한 헤더 정보를 이용하여 특정 응용 및 서비스를 식별하는 분석 방법이며 페이로드 기반 분석 방법보다 분석 속도는 빠르지만 분석 정확도는 떨어진다.

본 논문에서는 이러한 헤더 정보 기반의 분석방법의 한계점을 극복하기 위해 헤더 정보 기반의 자동 시그니쳐 명명 시스템을 제안한다. 본 논문은 2장에서 제안하는 시스템의 구조를 설명하고 3장에서는 본 시스템을 학내 인터넷 망에 적용한 실험 결과를 기술한다. 마지막으로 4장에서는 결론과 향후 연구를 언급한다.

II. 시스템 구조

본 장에서는 제안하는 자동 명명 시스템의 구조와 핵심 부분인 순차 패턴 알고리즘을 이용하여 헤더 시그니쳐의 이름을 추출하는 과정을 설명한다.

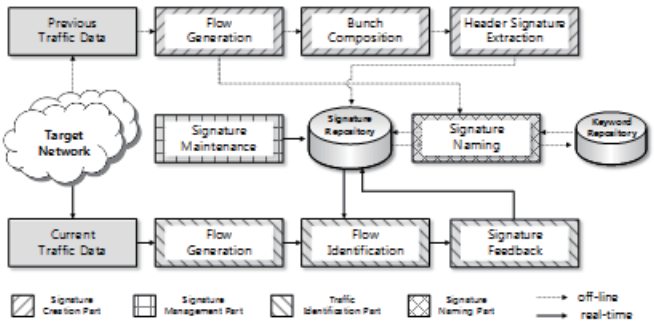

그림 1. 헤더 기반 시그니쳐 자동 명명 시스템

그림 1은 본 논문에서 제안하는 헤더 정보 기반의 자동 시그니쳐 명명 시스템의 구조이다. 본 시스템은 시그니쳐 생성부(Signature Creation Part), 시그니쳐 관리부(Signature Management Part), 트래픽 분석부(Traffic Identification Part), 시그니쳐 명명부(Signature Naming Part) 총 4개 부분으로 구성되어 있다.

시그니쳐 생성부(Signature Creation Part)에서는 해당 네트워크에서 발생하는 트래픽을 바탕으로 응용이나 서비스를 제공하는 서버의 IP, Port, Protocol을 이용하여 헤더 시그니쳐를 생성하고 저장소(Signature

Repository)에 저장한다. 이때 생성되는 헤더 시그니쳐는 이름이 명명되지 않은 상태이다. 시그니쳐 관리부(Signature Management Part)에서는 각 시그니쳐마다 갖고 있는 임계값[1]을 계산하여 시그니쳐 저장소(Signature Repository)에 계속 유지되는 헤더 시그니쳐인지 아닌지 결정하고 헤더 시그니쳐를 삭제 또는 갱신한다. 시그니쳐 명명부(Signature Naming Part)에서는 시그니쳐 관리부에 의해 유지된 헤더 시그니쳐와 실제 트래픽을 매칭한다. 이 때 매칭되는 트래픽에서 시그니쳐를 대표할 수 있는 문자열을 순차 패턴 알고리즘[2]을 사용하여 추출한다. 추출된 문자열은 해당 헤더 시그니쳐의 이름 후보 문자열이며 후보 문자열들 중에서 의미 있는 문자열을 이름으로 선택하기 위해 키워드 저장소(Keyword Repository)에 있는 키워드들과 매칭하여 해당 키워드가 포함된 후보 문자열일 경우 가중치를 부여한다. 이 가중치 값은 16진수 후보 문자열과 같은 이름으로 정해질 수 없는 후보가 이름으로 되는 것을 방지한다. 트래픽 분석부(Traffic Identification Part)에서는 기존의 페이로드 기반, 통계 기반 분석방법으로 정답지 트래픽을 만들고 해당 분석 결과와 본 논문에서 제안하는 자동 시그니쳐 명명 시스템의 결과를 서로 비교한다. 비교 과정을 통해 네이밍 시스템의 결과를 검증하고, 결과 값을 통해 해당 헤더 시그니쳐의 이름을 최종 결정하게 된다.

총 4개의 파트로 구성된 시스템을 통해 헤더 시그니쳐의 이름이 결정되고 시그니쳐 저장소에는 해당 타겟 네트워크의 시그니쳐가 유지된다. 이렇게 유지되는 헤더 시그니쳐는 서비스를 제공하는 서버 측의 헤더 정보이므로 다른 네트워크에도 동일한 헤더 시그니쳐를 적용할 수 있다.

III. 실험 결과

본 장에서는 제안하는 헤더 시그니쳐 자동 명명 시스템의 타당성을 보이기 위하여 실제 실시간으로 트래픽이 발생하는 학내망 네트워크에 적용하였다.

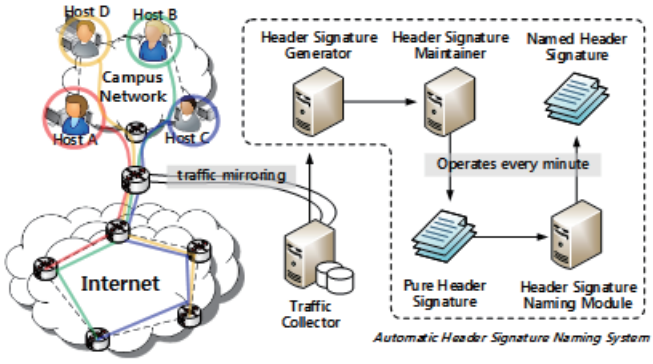

그림 2 . 실험 환경 구성

그림 2는 본 논문에서 제안하는 시스템을 실제 학내망 네트워크에 적용한 실험 환경 구성도이다. 실시간으로 발생하는 학내망의 트래픽은 Traffic Collector에서 매 분단위로 수집된다. 이렇게 수집된 트래픽은 헤더 시그니쳐 자동 명명 시스템(Automatic Header Signature Naming System)의 입력으로 사용된다. 해당 시스템 또한 매 분마다 가동되어 명명된 헤더 시그니쳐(Named Header Signature) 결과를 반환한다.

표 1은 본 논문에서 제안한 시그니쳐 자동 명명 시스템의 결과물인 명명된 헤더 시그니쳐(Named Header Signature)이다. 표에서 *표시된 항목을 살펴보면 구글의 여러 하위 서비스들이나 특정 목적에 의해 동작중인 서버의 헤더정보들을 추출할 수 있었다. 이렇게 얻어진 표 1의 IP 주소인 216.58.221.* 대역을 각각 "whois" 검색, "nslookup" 명령의 결과와 비교하였다.

표 1. 헤더 시그니쳐 자동 명명 시스템 결과

	Name	IP address	Port
*	accounts.google.com	216.58.221.141	443
		216.58.221.109	443
*	mail.google.com	216.58.221.101	443
*	clients1.google.com	216.58.221.110	443
*	www.youtube.com	216.58.221.142	80
	microsoft.com	192.229.145.200	443
	fe2.update.microsoft.com	191.232.80.62	443
	facebook.com	31.13.82.1	443

표 2, 3은 표 1에서 얻어진 명명된 헤더 시그니쳐 정보들 중에서 IP 주소(216.58.221.* 대역)의 "whois 검색", "nslookup 명령" 결과이며 이 결과는 단순히 해당 IP의 관리 기관과 대표 이름뿐 어떤 종류의 서비스나 컨텐츠를 제공하는지를 알 수 있는 정보는 포함하고 있지 않다.

표 2. 216.58.221.* 대역 Whois 검색 결과

NetRange	216.58.192.0 - 216.58.223.255
CIDR	216.58.192.0/19
NetName	GOOGLE
NetHandle	NET-216-58-192-0-1
Parent	NET216 (NET-216-0-0-0-0)
NetType	Direct Allocation
OriginAS	AS15169
Organization	Google Inc. (GOGL)
RegDate	2012-01-27
Updated	2012-01-27
Ref	http://whois.arin.net/rest/net/NET-216-58-192-0-1

표 3. 216.58.221.* 대역 nslookup 명령 결과

```
#  nslookup 216.58.221.101
Server  : 168.126.63.1
Address  : 168.126.63.1#53
Non-authoritative  answer :
101.221.58.216.in-addr.arpa  name = hkg07s01-in-f5.1e100.net.
Autoritative  answers can be found from:
221.58.216.in-addr.arpa  nameserver = ns3.google.com
221.58.216.in-addr.arpa  nameserver = ns2.google.com
221.58.216.in-addr.arpa  nameserver = ns1.google.com
221.58.216.in-addr.arpa  nameserver = ns4.google.com
ns1.google.com  internet address = 216.239.32.10
ns2.google.com  internet address = 216.239.34.10
ns3.google.com  internet address = 216.239.36.10
ns4.google.com  internet address = 216.239.38.10
```

IV. 결론 및 향후 연구

본 논문에서는 헤더 시그니쳐를 생성, 관리하고 시그니쳐의 이름을 자동으로 생성하는 시스템을 제안하였다. 제안하는 시스템의 성능을 검증하기 위해 실제 학내망 네트워크에 적용하였다. 본 시스템의 결과와 "nslookup 명령" 과 "whois 검색" 결과를 서로 비교하여 얻을 수 있는 정보의 차이를 보여줌으로써 본 시스템의 타당성을 입증하였다.

향후 연구로는 현재까지 명명된 헤더 시그니쳐를 검증하고 다른 종류의 시그니쳐와 비교하여 그 결과를 융합할 수 있는 방법을 연구할 예정이다.

ACKNOWLEDGMENT

본 논문은 BK21플러스 사업+ (No. T1300572) 및 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보.컴퓨팅기술개발사업(2010-0020728)의 지원을 받아 수행된 연구임

참 고 문 헌

[1] 윤성호, 박준상, 김명섭, "인터넷 트래픽 분석을 위한 헤더 시그니쳐 관리 방법," KNOM Review, Vol. 16, No. 1, Jul. 2013, pp. 11-23.

[2] 윤성호, 박준상, 안현민, 김명섭, "순차 패턴 알고리즘을 사용한 트래픽 행위 시그니쳐 생성 방법", 2014년도 한국통신학회 하계종합학술발표회, 라마다호텔, 제주도, Jun. 25-27, 2014, pp.996-997.