사이버 보안을 위한 소프트웨어 정의 네트워크에서의 트래픽 샘플링 기법

김성환, 윤승현, 하태진, 임 혁
광주과학기술원 전기전자컴퓨터공학부

{sunghwankim, seunghyunyoon, tjha, hlim}@gist.ac.kr

Traffic Sampling for Cyber Security on Software-Defined Networks

Sunghwan Kim, Seunghyun Yoon, Taejin Ha, and Hyuk Lim
Gwangju Institute of Science and Technology (GIST)

요 약

최근 모바일, IoT, 클라우드 등 정보통신환경의 변화와 함께 네트워크 트래픽 발생량이 급속히 늘어나고 있다. 이에 따라, 정보통신망의 보안에 대한 중요성도 커지고 있다. 네트워크 상의 트래픽들을 패킷 단위로 분석하여 정보통신망의 보안을 위해 사용되고 있는 대표적인 보안시스템인 침입탐지시스템(IDS, intrusion detection system)은 제한된 처리용량을 갖는다. 본 논문에서는 네트워크 IDS의 제한된 처리용량에서 플로우별 가시성 확보를 위한 소프트웨어 정의 기반 네트워크에서의 트래픽 샘플링률 결정 방법을 제안하고, 이에 대한 성능 평가를 수행하였다. 시뮬레이션 결과를 통해 제안하는 방법이 플로우별 목표 트래픽 샘플링률을 만족하는 스위치별 트래픽 샘플링률을 결정할 수 있음을 확인하였다.

Ⅰ. 서 론

오늘 날 정보통신기술의 발달로 인한 네트워크 트래픽 발생량의 증가에 따라, 네트워크를 통한 다양한 보안 위협 또한 증가하고 있다. 하루에 20,000 건 이상의 DDoS(distributed denial of service)공격이 발생하고 있으며, 최근 미국의 랜섬웨어(ransomware) 피해액은 3억 2천 5백만 달러로 추정되고 있다. 이러한 네트워크 침입을 막기 위해 대표적으로 침입탐지시스템(IDS, intrusion detection system)과 같은 트래픽 분석 장비들이 사용되고 있다. IDS는 네트워크상의 트래픽을 패킷 단위로 분석하여, 공격으로 의심되는 플로우에 대한 알람을 주는 등의 방법으로 네트워크의 안정성을 확보한다. 일반적으로 IDS와 같은 트래픽 분석 장비들은 패킷 분석을 위한 처리용량이 제한적이라는 한계를 갖는다.

네트워크 트래픽을 선택적으로 모니터링하는 트래픽 샘플링 기법을 사용하면, 트래픽 분석 장비들의 제한된 처리용량에 맞춰 네트워크 트래픽을 분석할 수 있다. 네트워크의 제어 평면과 데이터 전달 평면을 분리하여, 네트워크를 중앙의 제어기를 통해 관리할 수 있는 소프트웨어 정의 네트워킹(SDN, Software-defined Networking)기술을 사용하면, 네트워크 트래픽 모니터링을 위한 샘플링을 쉽게 구현할 수 있다 [1]. 트래픽 샘플링 기법은 모니터링해야하는 네트워크 트래픽의 양을 줄임으로써, 적은 오버헤드와 실시간 네트워크 트래픽 분석을 할 수 있는 장점이 있지만, 분석 결과의 신뢰성 측면이 부족한 단점이 있다. 즉, 버려진 트래픽에 담겨있는 잠재적으로 유용할 수 있는 패킷의 정보를 얻지 못하는 상황이 발생할 수 있다. 따라서 효과적인 네트워크 트래픽 모니터링을 위해서는 적절한 트래픽 샘플링률을 결정하는 것이 중요하다. 본 논문에서는 소프트웨어 정의 기반 네트워크에서의 IDS를 위한 트래픽 샘플링 기술에 관한 연구를 수행한다.

Ⅱ. 관련 연구

네트워크 트래픽 샘플링률 결정을 위한 소프트웨어 정의 기반 네트워크에서 트래픽 샘플링 방법으로는 플로우별 샘플링과 스위치별 샘플링이 고려될 수 있다. 플로우별 트래픽 샘플링 방법(per-flow sampling method)는 플로우마다 샘플링률을 다르게 설정할 수 있으므로, 정교한 샘플링 수행이 가능하다는 장점이 있다. 하지만, 네트워크상의 플로우들은 동적으로 변하기 때문에, 스위치들의 포워딩 테이블을 빈번하게 업데이트할 필요가 있다. SDN을 지원하는 스위치들은 제한된 메모리를 갖고 있기 때문에, 샘플링을 위해 플로우마다 포워딩 테이블을 관리하는 것은 확장성 측면에서 문제가 있다 [2].

네트워크상의 스위치들의 개수는 플로우들에 비해 동적으로 변하지 않기 때문에, 스위치별 트래픽 샘플링 방법(per-switch sampling method)은 플로우별 트래픽 샘플링 기법보다 포워딩 테이블 관리에 필요한 오버헤드가 적은 장점이 있다. 그러나, 스위치별 샘플링률을 적절하게 결정하지 못하는 경우, 몇몇 플로우들이 여러 스위치에서 지나치게 많이 샘플링되거나, 전혀 샘플링되지 않는 문제가 발생할 수 있다 [2].

샘플링을 위한 포워딩 테이블 관리 오버헤드를 줄이는 동시에 플로우별 정교한 트래픽 샘플링이 가능하도록 하는 스위치별 트래픽 샘플링 방법으로는 Rate-proportional 샘플링 방법이 있다. Rate-proportional 샘플링 방법은 플로우별 data rate의 비율에 따른 트래픽 플로우 샘플링을 수행하기 위한 스위치별 샘플링률을 결정한다. 라우팅 행렬 A와 스위치별 샘플링률 벡터 $\vec{x}$, 플로우별 목표 샘플링률 벡터 $\vec{b}$를 이용하여, 식 (1)과 같이 트래픽 플로우 샘플링을 위한 스위치별 샘플링률을 결정한다. $\vec{x}$와 $\vec{b}$의 성분은 0과 1 사이의 값을 갖는다[2].

$$\vec{x} = (A^T A)^{-1} \times A^T \times \vec{b} \quad (1)$$

Rate-proportional 샘플링 방법은 스위치별 샘플링률 벡터 $\vec{x}$를 결정하기 위해 pseudo inverse를 취하기 때문에, 작은 규모의 네트워크에서는

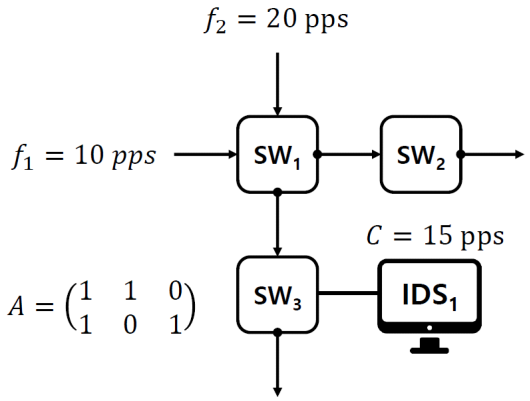

그림 1. 스위치별 샘플링률에 따른 플로우별 샘플링률 결정 예제

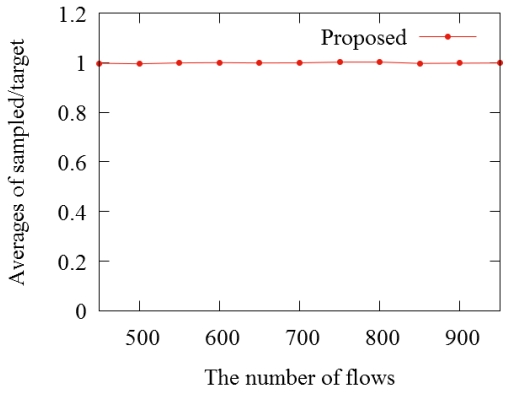

그림 2. 플로우별 목표 샘플링률 대비 실제로 샘플링된 플로우별 샘플링률의 평균

문제가 없으나 네트워크의 규모가 커질 경우 역행렬 계산의 부정확성으로 인하여 스위치별 샘플링률이 음수가 나오는 문제가 있다. 본 논문에서는 확장성 있는 네트워크 트래픽 샘플링을 위하여, 네트워크의 규모가 커지더라도 목표로 하는 플로우별 트래픽 샘플링률을 만족하는 스위치별 트래픽 샘플링률 결정 방법을 제안한다.

Ⅲ. 트래픽 플로우 샘플링을 위한 스위치별 트래픽 샘플링률 결정 방법

n개의 플로우, m개의 SDN 지원 스위치, 그리고 1개의 IDS로 이루어진 소프트웨어 정의 기반 네트워크에서 각 플로우별 data rate과 스위치별 샘플링률을 각각 $\vec{f}$, $\vec{x}$ 로 정의하고, IDS의 처리량의 한계를 C라 정의한다. $\vec{x}$의 성분은 0과 1 사이의 값을 갖는다. Data rate은 초당 패킷 개수(pps, packets per second)로 나타낸다. $n \times m$ 행렬 A는 라우팅테이블을 나타내며, i번째 플로우가 j번째 스위치를 지나가는 경우, A의 성분 $a_{i,j}$는 1이다. 플로우별 샘플링된 data rate을 $\vec{r}$로 정의하면, $\vec{x}$ 가 주어졌을 때 $\vec{r}$은 다음과 같은 식으로 구할 수 있다.

$$\vec{r} = \mathrm{diag}(\vec{f}) \times A \times \vec{x} \tag{2}$$

만약 두 개 이상의 플로우가 하나의 스위치를 동시에 지나갈 경우, 각 플로우는 플로우별 data rate의 비율에 따라 샘플링된다고 가정한다. 그림 1은 스위치별 샘플링률이 주어졌을 때 플로우별 샘플링된 data rate이 결정되는 예제이다. 네트워크 토폴로지는 각각 10 pps, 20 pps의 data rate을 갖는 2개의 플로우, 15 pps의 처리용량을 갖는 1개의 IDS, 그리고 3개의 스위치로 구성된다. $\vec{x} = [0.4\ \ 0.1\ \ 0.1]^T$인 경우, 식 (2)에 따라 각 플로우별 샘플링된 data rate을 계산하면, $\vec{r} = [5\ \ 10]^T$이다.

$\vec{\hat{r}}$을 플로우별 목표로 하는 샘플링된 data rate이라 정의하면, $\vec{\hat{r}}$과 $\vec{r}$의 차이를 최소화하는 목적을 달성하기 위한 최적화 문제를 다음과 같이 나타낼 수 있다.

$$R(\vec{x}) = \min_{\vec{x}} \sum_{i}^{n} (\vec{r} - \vec{\hat{r}})^2 \tag{3}$$

$$\text{subject to} \sum_{i}^{n} \vec{r_i} \leq C \text{ and } \vec{r_i} \leq \vec{f_i} \text{ for } \forall i$$

목적함수 $R(\vec{x})$는 $\vec{\hat{r}}$이 주어졌을 때, 식 (3)의 제약조건에 따라 총 샘플링된 패킷의 볼륨이 C를 넘지 않으며, 플로우별 샘플링된 data rate이 플로우별 data rate을 넘지 않도록 스위치별 샘플링률 $\vec{x}$를 결정한다.

Ⅳ. 시뮬레이션

제안하는 샘플링률 결정 알고리즘의 성능 평가를 위하여 스위치 400개, IDS 처리용량 1000 pps인 네트워크에서 플로우 개수를 450개에서 950개까지 증가시켜가며, 플로우별 목표 샘플링률 대비 실제로 샘플링된 플로우별 샘플링률의 평균을 측정하는 시뮬레이션을 수행하였다. 각 플로우는 평균 30 pps의 data rate을 가지며, 패킷의 사이즈는 동일하다. 네트워크 토폴로지는 networkX 라이브러리를 사용하여 생성하였다. 스위치별 샘플링률 $\vec{x}$를 결정하는 최적화 문제 식 (2)는 CVXPY 라이브러리를 사용하여 풀었다.

그림 2는 플로우 개수를 증가시켜가며, 플로우별 목표 샘플링률 대비 실제로 샘플링된 플로우별 샘플링률의 평균을 측정한 결과이다. 제안하는 방법의 평균은 그림 2와 같이 플로우 개수가 증가하여도 1에 가까운 값을 유지하였다. 분산은 0.02 이하의 값을 유지하였다. 시뮬레이션 결과, 제안하는 방법이 목표로하는 플로우별 샘플링률에 따라 실제로 플로우별 샘플링이 이루어지고 있음을 확인하였다. 또한, 네트워크가 확장되어도 정확한 플로우별 목표 샘플링률을 만족하고 있음을 알 수 있었다.

Ⅴ. 결론

본 논문에서는 소프트웨어 정의 기반 네트워크에서의 IDS를 위한 트래픽 샘플링 기술에 관한 연구를 수행하였다. 기존 소프트웨어 정의 기반 네트워크에서 트래픽 플로우 샘플링을 위한 방법인 rate-proportional 샘플링 방법은 pseudo inverse를 취하기 때문에, 네트워크 규모가 커질 경우 스위치별 샘플링률이 음수가 나오게 되는 문제가 있다. 본 논문에서는 확장성 있는 트래픽 샘플링을 위한 트래픽 샘플링률 결정 방법을 제안하였다. 시뮬레이션을 통하여 제안하는 방법이 네트워크가 확장되어도 정확한 플로우별 목표 샘플링률을 만족할 수 있음을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2017-0-00421, 새로운 보안 위협에 대처하기 위한 사이버 보안 방어 순환 기술).

참 고 문 헌

[1] T. Ha et al., "RTSS: Random traffic sampling and steering for OpenFlow-enabled networks," ACM CoNEXT Student Workshop, Irvine, California, December 12-15, 2016.

[2] S. Yoon et al., "Scalable traffic sampling using centrality measure on software-defined networks," IEEE Commun. Mag., July 2017.