

SDN환경에서 머신러닝을 이용한 트래픽 분류방법

임환희⁰, 김동현*, 김경태*, 윤희용**

⁰성균관대학교 정보통신대학 전자전기컴퓨터공학과

**성균관대학교 소프트웨어대학 소프트웨어학과

e-mail: {lhh423,kdh7263}@skku.edu⁰, kyungtaekim76@gmail.com*, youn7147@skku.edu**

Traffic classification using machine learning in SDN

Hwan-Hee Lim⁰, Dong-Hyun Kim*, Kyoung-Tae Kim*, Hee-Yong Youn**

⁰Dept. of Electrical and Computer Engineering, Sungkyunkwan University

**Dept. of Software, Sungkyunkwan University

● 요약 ●

Software Defined Networking(SDN)은 데이터 부와 컨트롤 부를 나눠 관리하는 혁신적인 방식이다. SDN 환경에서가 아닌 기존의 IP 네트워크에서의 트래픽 분류는 많은 연구가 진행되어 왔다. 트래픽 분류 방법에는 Port 번호를 이용한 트래픽 분류 방법, Payload를 이용한 트래픽 분류 방법, Machine Learning을 이용한 트래픽 분류 방법 등이 있다. 본 논문에서는 Port 번호, Payload, Machine Learning을 이용한 트래픽 분류 방법들을 소개 및 장단점을 설명하고 SDN 환경에서 Machine Learning을 이용한 좀 더 정확한 트래픽 분류 방법을 제안한다.

키워드: Software Defined Network(SDN), 트래픽 분류, 머신러닝

I. Introduction

오늘날 인터넷은 어디서든지 연결되어 있고, 사용하며, 빠른 정보를 습득하게 해주었다. 인터넷은 전 세계 어느 곳이든 사용할 수 있지만 복잡하고 관리하기가 어렵다. 그래서 수년간 네트워크에 대한 연구가 활발히 이루어 졌다. 많은 연구가 진행되어 왔으며, 그 중 Software Defined Network(SDN)이 최근 연구주제로 떠오르고 있다[1]. SDN은 데이터 패킷을 담당하는 데이터 부와 데이터를 제어하는 제어 부로 나누고 이를 소프트웨어로 관리하는 방법이다. SDN은 OpenFlow Protocol을 사용한다. OpenFlow는 제어 부와 데이터 부 간의 통신 Protocol이다. 네트워크 관리자는 SDN을 통해 좀 더 쉽게 네트워크 트래픽을 관리한다. 그리고 오늘날 네트워크 이용 종류가 다양해지고 있으며, 각자 다른 패턴의 트래픽을 발생시킨다. 그래서 트래픽 분류는 필수적이다. 트래픽 분류를 하게 되면 최적의 라우팅 경로를 얻을 수 있으며, 전체 네트워크 성능을 최적화시킬 수 있다. 본 논문에서는 서론에 이어 2장에서는 국내외 트래픽 분류에 대한 연구 동향에 대해서 알아보며, 3장에서는 Machine Learning을 이용한 트래픽 분류 방법을 제안한다. 그리고 마지막으로 4장에서는 결론 및 기대효과에 대해 설명한다.

II. Related works

1. 국내외 동향

트래픽 분류 연구에 대한 연구는 국내외에서 활발히 진행 중이다. 그 중 Port 번호 기반, Payload 기반, 머신러닝 기반 트래픽 분류 방법을 설명한다.

1.1 Port Number 기반 트래픽 분류

Port Number 기반 트래픽 분류 방법은 과거에 가장 많이 사용된 분류 기법 중 하나이다. 기관이 할당한 고정 포트번호를 사용한다. 하지만 이 방법은 고정 포트번호를 사용하지 않는 많은 응용프로그램의 등장과 IP 계층의 암호화, 포트번호를 알지 못하는 문제점이 있다.

1.2 Payload 기반 트래픽 분류

대표적으로 Deep Packet Inspection(DPI)이 있으며, 현재 가장 많이 사용되는 기술 중 하나이다. 데이터 패킷 내부까지 분석하는 기술이며 분류 정확도가 높으나 대량의 메모리를 사용하며, 심각한 확장성 문제가 발생할 수 있다.

1.3 Machine Learning 기반 트래픽 분류[2]

머신러닝을 이용한 트래픽 분류는 Internet Protocol Flow Information eXprt 표준에 기반한 데이터에 적용하며, 별도의 모듈에서 머신러닝을 적용한 트래픽 분류를 한다. 짧은 초기 패킷을 수신한 후, 빠르게 분류를 할 수 있으며, 사용자가 정의할 수 있다.

III. The Proposed Scheme

2장에서 각각의 트래픽 분류 방법들을 설명하였다. 본 논문에서는 SDN 환경에서 트래픽 분류 방법은 제안한다. 아래의 그림 1은 본 논문에서 제안한 트래픽 분류 방법의 알고리즘이다.

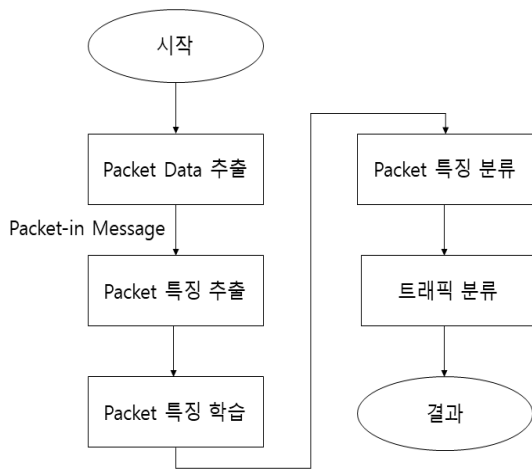


Fig. 1. 제안한 알고리즘

SDN환경에서 Node나 Host의 패킷을 SDN Controller로 Packet-in 메시지를 보낸다. 그리고 난 뒤에는 Packet의 특징을 추출해 내며 Packet의 특징을 학습한다. 학습된 데이터를 가지고 특징을 분류하고 트래픽 분류를 한다. 트래픽 분류는 가장 기본 머신러닝 분류 알고리즘인 Naive Bayes을 사용한다.

Naive Bayes 알고리즘의 식은 아래와 같다.

$$P(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)}$$

IV. Conclusions

본 논문에서는 SDN 환경에서 머신러닝을 이용한 트래픽 분류 방안을 제시하였다. 기존의 분류 방법들은 여러 가지 문제점이 있지만 머신러닝을 이용한 트래픽 분류 기법에는 학습을 해야 한다는 단점이 있지만 장점이 많아 앞으로 많은 연구가 진행될 것이다. 또한, 제안한 트래픽 분류방법은 기존의 머신러닝 기법에서 한발 더 나아가 Packet 특징 또한 머신러닝을 이용해 특징을 추출하기 때문에 좀 더 높은 정확도로 분류가 될 것으로 예상된다.

Acknowledgment

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신방송연구 개발 사업(No.B0717-17-0070, 초연결 IoT 노드의 군집 지능화를 통한 Edge Computing 핵심 기술 연구), SW중심대학지원사업(2015-0-00914), 한국연구재단 기초연구사업 (No. 2016R1A6A3A11931385, 실시간 공공안전 서비스를 위한 소프트웨어 정의 무선 센서 네트워크 핵심기술 연구, 2017R1A2B2009095, 실시간 스트림 데이터 처리 및 Multi-connectivity를 지원하는 SDN 기반 WSN 핵심 기술 연구), 삼성전자, BK21PLUS 사업의 일환으로 수행되었음.

REFERENCES

- [1] Diego Kreutz, Fernando M. V. Ramos, Paulo Esteves Verissimo, "Software-Defined Networking: A Comprehensive Survey" Proceedings of the IEEE, Vol. 103, pp. 14-76, December 2014.
- [2] Mohammad Reza Parsaei, Mohammad Javad Sobouti, Seyed Raouf khayami, Reza Javidan, "Network Traffic Classification using Machine Learning Techniques over Software Defined Networks", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 7, pp. 220-225, Jan. 2017.