

SDN환경에서 트래픽 분류를 위한 특징 선택 기법

임환희[○], 김동현^{*}, 이병준^{*}, 김경태^{*}, 윤희용^{**}

[○]성균관대학교 전자전기컴퓨터공학과

^{**}성균관대학교 소프트웨어대학 소프트웨어학과

e-mail: {lhh423,kdh7263,byungjun}@skku.edu[○], kyungtaekim76@gmail.com^{*}, youn7147@skku.edu^{**}

Feature Selection for Traffic Classification in SDN

Hwan-Hee Lim[○], Dong-Hyun Kim^{*}, Byung-Jun Lee^{*}, Kyung-tae Kim^{*}, Hee-Yong Youn^{**}

[○]Dept. of Electrical and Computer Engineering, Sungkyunkwan University

^{**}Dept. of Software, Sungkyunkwan University

● 요약 ●

본 논문에서는 SDN환경에서 트래픽 분류를 위한 특징 선택 기법을 제안한다. 최근 들어 다양한 스마트폰 어플리케이션이나 다양한 IoT 기기들이 등장하고 있다. 다양한 IoT 기기나 어플리케이션은 엄청난 양의 트래픽을 발생시킨다. 엄청난 양의 트래픽은 전송속도를 저하시킬 뿐만 아니라, 높은 Quality of Service(QoS)를 보장하기가 힘들다. 또한 이러한 문제점들을 보완하기 위해 Software Defined Networking(SDN)이라는 기술이 빠른속도로 발전하고 있으며, 본 논문에서는 다양한 어플리케이션과 IoT 기기의 트래픽을 SDN환경에서 분류 하며, 트래픽의 분류 정확도와 더 빠른 분류를 위해 트래픽의 특징 선택 기법을 제안한다. 특징 선택을 수행한 뒤, 트래픽 분류를 진행하면 트래픽 분류 정확도를 높이고, 트래픽 분류시간은 줄어드는 효과를 보이며, 높은 QoS를 보장함으로써, 기존 네트워크 트래픽의 부하가 줄어드는 우수한 성능을 보인다.

키워드: IoT(Internet of Things), 트래픽 분류, 특징 선택, SDN(Software Defined Networking)

I. Introduction

오늘날 대한민국 인구대비 이동통신 보급률은 122%를 상회했으며, 2017년 3분기 기준으로 이동통신 데이터 트래픽은 305 Petabyte를 기록했으며, 국내 이동통신 트래픽은 10배 급증해 2023년 말 3.2 Exabyte에 이를 것으로 예상된다. 또한, 2021년까지 IoT 기기의 연결이 137억 건까지 증가될 것으로 예상된다. 이와 같이 엄청난 양의 트래픽을 관리, 제어하는 것은 매우 힘든 일이며 트래픽 관리, 제어하는 기술이 필수적이 될 것이다. 따라서 본 논문에서는 IoT 기기와 스마트폰에서 발생하는 트래픽을 분류하며, 트래픽을 특징 선택기법을 이용해 빠른 분류, 높은 QoS 보장을 목표로 한다. 2장에서는 트래픽 분류 및 특징 선택에 관한 관련연구를 서술하며, 3장에는 트래픽 특징 선택에 관한 기법을 제안하며 마지막으로 4장에서는 결론을 서술한다.

II. Preliminaries

1. Related works

1.1 Feature Selection

특징 선택 기법이란 분류 정확도를 향상시키기 위해 원본 데이터가 주어졌을 때, 가장 좋은 성능을 보여줄 수 있는 데이터의 부분 집합을 원본 데이터에서 찾아내는 방법이며, 분류 목적에 가장 밀접하게 연관되어 있는 특징만을 선택한다.

특징 선택 기법에는 대표적으로 세 가지 기법이 존재한다. 첫 번째로는 Filter Method, 두 번째로는 Wrapper Method, 세 번째로는 Embedded Method가 있다[1].

Filter Method는 각각의 특징에 통계적인 점수를 부여하여 Rank가 높은 순으로 특징을 선택하며, 계산량이 적어 빠르다는 장점이 있지만, 정확도가 떨어질 수 있는 단점이 있다. FCBF, mRMR 알고리즘 등이 있다.

Wrapper Method는 가장 이상적인 특징 조합을 찾는 것이며, 특징들의 조합을 다르게 하면서 학습을 진행하며, 완료된 모델 중 가장 점수가 높은 조합을 선택한다. 하지만 계산량이 많아 느리다는 단점이 있다. Genetic 알고리즘, SVM 등이 있다.

Embedded Method는 filter와 wrapper method이 혼합되어 있으며, 모델의 학습 및 생성과정에서 최적의 특징을 선택한다. 가장 유명한 방법으로는 특징에 penalty를 가하는 방식인 Regularization 이 있다. Lasso, Ridge Regularization 등이 있다.

III. The Proposed Scheme

본 논문에서는 Filter Method와 Embedded Method를 결합해 각 트래픽 특징들의 결과들을 중복되는 것만 뽑아 트래픽 분류를 진행한다. 아래의 그림 1은 본 논문에서 제안한 특징 선택 알고리즘이다.

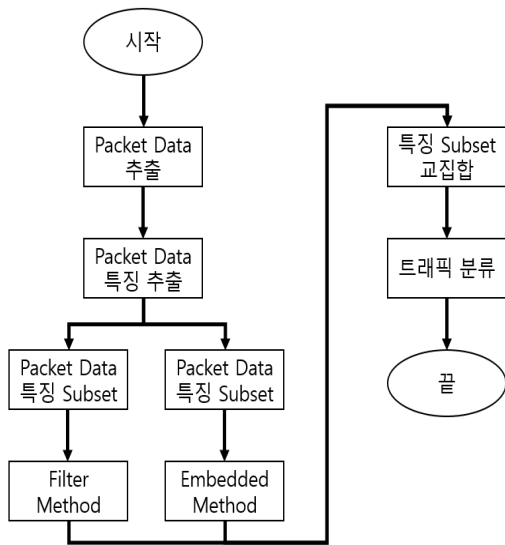


Fig. 1. 제안한 알고리즘

SDN Controller에서 Packet-in message를 이용해 패킷 데이터를 추출한다. 그다음 패킷의 특징을 추출한 뒤, Filter, Embedded Method 각각 Feature selection을 수행한 뒤에 수행되어 나온 Feature 들을 교집합을 한다. 최종적으로 교집합 된 Feature들을 가지고 트래픽 분류를 수행한다.

IV. Conclusions

본 논문에서는 SDN 환경에서 특징선택을 한 뒤 트래픽 분류를 수행하였다. 아래의 그림2는 각 알고리즘별로 분류 정확도를 측정한 결과이다. 기존의 특징선택을 하지 않고 트래픽 분류를 수행하거나, Filter 방법만 가지고 특징 선택을 한 뒤 트래픽분류를 수행한 것보다 제안한 알고리즘이 분류정확도가 가장 높은 것을 확인할 수 있다.

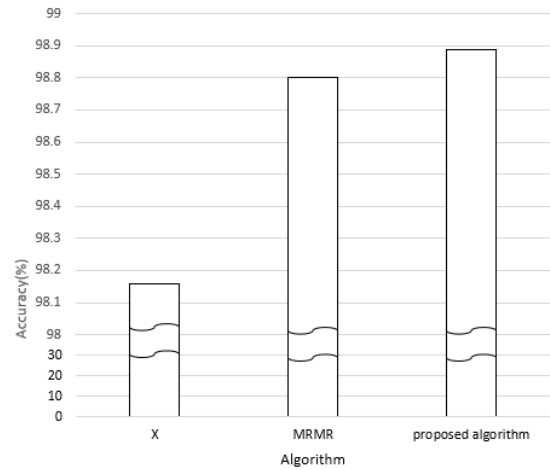


Fig. 2. 분류 정확도

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신-방송연구 개발 사업(No. 2016-0-00133, 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] Girish Chandrashekar, Ferat Sahin, "A survey on feature selection methods," Computers & Electrical Engineering, Vol. 40, No. 1, pp. 16-28, Jan 2014.