

SDN 환경에서의 TrAdaBoost 기반 Flow 규칙 구분 기법

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

[○]성균관대학교 정보통신대학 전자전기컴퓨터공학과

^{**}성균관대학교 소프트웨어대학 소프트웨어학과

e-mail: {kimmw95, lhh423, byungjun}@skku.edu[○], kyungtaekim76@gmail.com^{**}, youn7147@skku.edu^{*}

TrAdaBoost-based Flow Rule Classification Technique in SDN Environment

Min-Woo Kim[○], Hwan-Hee Lim^{*}, Byung-Jun Lee^{*}, Kyung-Tae Kim^{**}, Hee-Yong Youn^{*}

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

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

● 요약 ●

기존의 Flow 규칙 구분을 위해 연구되었던 기법들은 적응적 또는 사전 처리의 접근법이 제안되었으나 각각의 장단점을 기반으로 효율적인 접근법이 연구되어야한다. 본 연구에서는 Flow 규칙을 삽입하기 전에, 스위치의 계산 작업을 완화하기 위하여 전이 학습 기법인 TrAdaBoost를 이용함으로써 Flow 규칙들을 구분하는 접근법을 제안한다.

키워드: TrAdaBoost, SDN(software defined networking), Flow 규칙 구분(flow rule classification)

I. Introduction

SDN(Software Defined Networking)의 빠른 발전으로 Flow table에 대한 규칙을 분산하는 방법에 대한 연구가 활발히 진행되어왔다. 기존 OpenFlow 프로토콜의 기능은 저장 공간의 제약이 없고 높은 작업 처리 계산 능력을 가지고 있다는 가정을 기반으로 한다. 그렇기 때문에 실제 OpenFlow 스위치의 전력, 비용, 메모리 공간의 제약을 가진 SDN 환경에서 어떻게 Flow table에 Flow 규칙을 적용할 수 있을지에 대한 연구가 매우 중요하다[1]. 본 연구에서는 Flow 규칙을 삽입하기 전에, 스위치의 계산 작업을 완화하기 위하여 전이 학습 기법인 TrAdaBoost[2]를 이용함으로써 Flow 규칙들을 구분하는 접근법을 제안한다. 제안된 기법은 Flow 배치의 기본이 되는 단계이며 Table miss가 자주 발생하지 않으면서 높은 대역폭을 차지하는 Flow 규칙에 적용된다. Flow 규칙을 두 카테고리 나눈기 위하여 TrAdaBoost가 사용되며 구분 정확도를 통해 Flow 규칙의 배치 성능이 결정된다.

II. Preliminaries

1. Related works

1.1 Reactive Flow 규칙 배치 기법

Reactive Flow 규칙 배치는 SDN에서 기본적인 Flow 규칙 배치 기법으로 Table miss와 같은 Flow event를 기반으로 한다. Table

miss는 Flow가 Flow table에서 대응되는 Flow entry를 찾지 못할 때 발생한다. Reactive Flow 규칙 배치는 지속적으로 최근 네트워크 상태를 기반으로 컨트롤러와의 상호작용을 하며 이러한 규칙 배치는 장치의 더 많은 부하와 지연을 야기한다. 또한 컨트롤러의 계산속도 한계로 인하여 실시간으로 Flow를 관리하는 것은 합리적이지 않아 방대한 Flow의 경우에만 이러한 배치를 적용한다.

1.2 Proactive Flow 규칙 배치 기법

작은 Flow의 경우 Reactive Flow 기법의 사용이 적절하지 않다. Proactive Flow 규칙 배치는 사전에 Flow 규칙을 사용하거나 설치하는 확률을 예측하며 이를 통해 컨트롤러와 스위치의 부하를 완화하고 SDN의 지연 발생을 줄여준다.

III. The Proposed Scheme

기존의 머신 러닝은 구분자를 생성하기 위하여 학습 데이터 세트를 이용함. 이를 통해 학습된 구분자는 Flow 구분은 방대한 Flow와 소량의 Flow를 매핑하기 위하여 적용된다. 이때 학습 데이터 세트의 분포는 테스트 데이터 세트와 같으며 분포가 다른 데이터 세트는 배제된다. 기존의 머신러닝 알고리즘과는 다르게 제안된 전이 학습(Transfer Learning) 기법은 효율적인 구분자를 구현하기 위해 분포가 다른 데이터 셋을 조합한다. TrAdaBoost는 학습데이터를 이용하여

기존의 구분자와는 다르게 약한 여러 구분자를 구현한다. 이렇게 구현된 구분자들은 구분 오차에 따른 가중치를 부여받고 마지막에 강력한 구분자가 가중치를 이용하여 나머지 구분자를 결합한다. TrAdaBoost는 학습 데이터 세트와 테스트 데이터 세트의 분포가 같다는 측면에서 기존의 머신 러닝 알고리즘과 비슷하지만 전이 학습을 적용시켜 다른 분포를 가진 데이터세트까지 확장시켜 더 효율적인 성능을 얻게된다. Fig. 1은 TrAdaBoost 알고리즘의 학습 과정을 나타내고 있다.

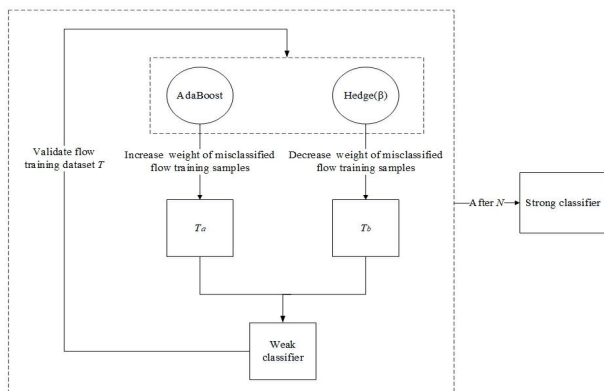


Fig. 1. TrAdaBoost 알고리즘의 학습 과정

그림에서 보이듯이 T_a 와 T_b 는 시작 단계의 입력이며 T_a 는 같은 분포를 가진 Flow 학습 데이터이고 T_b 는 다른 분포를 가진 Flow 학습 데이터를 나타낸다. 이들은 T 로 합쳐지고 이 때 각 Flow 샘플에 대한 가중치와 모델 β 의 인자가 초기화된다. 약한 구분자가 생성되고 난 후 다른 분포를 가진 Flow 학습 데이터가 잘 못 구분되면, 진행 과정의 수식에 따라 가중치가 감소하게 된다. 그러나 다른 분포를 가진 Flow 학습 샘플과는 반대로 같은 분포의 Flow 학습 샘플들의 가중치는 잘 못 구분되었을 때 증가하며 몇 번의 반복 과정 이후, 약한 구분자의 결과와 같은 결과를 같은 다른 분포의 Flow 학습 샘플들은 높은 가중치를 가지게 된다. 이러한 과정을 통하여 높은 구분 정확도를 가진 강한 구분자가 구현된다.

IV. Conclusions

향후 Flow 규칙 배치에 대한 Reactive 기법과 Proactive 기법의 장점을 효율적으로 결합시켜 적용하는 접근법이 연구할 것이며 규칙 배치 예측을 사용할 때 정확도를 향상시키기 위한 다른 모델 또는 알고리즘이 연구되어야한다.

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), SW중심

대학지원사업(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), BK21PLUS 사업의 일환으로 수행되었음.

REFERENCES

- [1] S.Zhang, F.Ivancic, C.Lumezanu, Y.Yuan, A.Gupta, S.Malik, "An Adaptable Rule Placement for Software-Defined Networks", Annual IEEE/IFIP International Conference on Dependable Systems and Networks, pp.88-99, Jun. 2014
- [2] G.Sun, L.Liang, T.Chen, F.Xiao, F.Lang, "Network traffic classification based on transfer learning", Computers & Electrical Engineering, Volume 69, pp. 920-927, Jul. 2018