네트워크 텔레메트리 기반 통합 네트워크 관리 시스템 연구

남석현°*, 임지윤*, 유재형*, 홍원기*

*포항공과대학교 컴퓨터공학과
{obiwan96, limjiyoon, jhyoo78, jwkhong}@postech.ac.kr

Network Telemetry-based Integrated Network Management System

Sukhyun Nam°*, Jiyoon Lim*, Jae-Hyoung Yoo*, James Won-Ki Hong*

*Department of Computer Science and Engineering, POSTECH

요 약

실시간으로 패킷 단위의 세부적인 네트워크 정보를 제공하는 INT (In-band Network Telemetry) 를 이용하면 우수한 성능으로 실시간 네트워크 이상 탐지와 로드 밸런싱이 가능하다. 본 논문은 INT 기반 통합 네트워크 관리 시스템을 제안하며, 제안한 시스템은 INT 를 이용하여 네트워크 상태 정보를 추출하며, 추출한 정보를 머신러닝을 이용하여 분석한다. 또한 실시간으로 네트워크 공격을 탐지하고 탐지된 공격에 대해선 완화 작용을 할 수 있으며, 네트워크 공격이 존재하지 않을 때는 네트워크 경로 별 가중치를 조정하여 네트워크 이용률을 높인다.

Ⅰ. 서 론

SDN (Software Defined Networking) 기술은 네트워크 관리에 있어 많은 것을 가능하게 한다. 네트워크 추상화, 네트워크 제어 기능의 논리적 중앙화가 SDN 의 중요한 특징이지만 특히, 최근에는 네트워크를 구성하는 스위치를 프로그래머블하게 만들어 컨트롤러에서 스위치에 대한 세부 조작이 가능하게 되었다. 프로그래머블 스위치의 언어로서 개발된 P4 [1]는 기존의 단순한 제어 평면의 트래픽 전달 동작 제어뿐만 아니라 각 스위치에서 헤더 포맷을 새로 정의하고 파싱하거나 match-action 테이블을 조절하는 등 더 복잡한 행동을 가능하도록 한다. P4 를 이용해 개발된 응용 기능 중 하나인 INT (In-band Network Telemetry)[2]는 실시간으로 네트워크의 상태 정보를 데이터 패킷에 실어 전달할 수 있어 네트워크에 대한 정보를 더 상세히 파악할 수 있도록 한다.

네트워크 관리에 있어서 가장 중요한 것은 로드 밸런싱과 네트워크 이상 탐지이다. 로드 밸런싱은 네트워크의 이용률을 높이기 위해 트래픽을 분산시켜 병목현상을 방지하는 기술이다. 네트워크 이상 탐지는 네트워크 상의 플로우에 대한 정보를 수집하여 네트워크에서 발생하는 악의적인 공격을 실시간으로 탐지하는 기술이다.

본 논문에서는 INT 를 이용한 네트워크 통합 관리 시스템을 제안한다. INT 를 활용하면 네트워크 상태에 대한 상세한 정보를 실시간으로 수집할 수 있어 로드 밸런싱과 이상 상태 탐지에 사용하기에 매우 적합하다. 따라서 INT 기반 네트워크 관리 시스템은 기존의 시스템보다 우수한 성능을 가진다.

본 논문의 구성은 다음과 같다. 2 장에서는 기존의 네트워크 이상 탐지 기술과 로드 밸런싱 기술, 그리고 INT 기술에 대해 기술한다. 3 장에서는 INT 를 활용한 네트워크 통합 관리 시스템 구조를 제안한다. 마지막으로 4 장에서는 결론을 기술한다.

Ⅱ. 관련 연구

1. 이상 상태 탐지 기술

NIDS (Network Intrusion Detection System)은 네트워크 상의 유해한 공격을 탐지하고 대응하는 시스템이다. NIDS 는 크게 두 가지 방식으로 연구되었는데, 첫째는 시그니처 기반(signature-based) 방식이고 둘째는 이상 상태 기반(anomaly-based) 방법이다. 두 방법은 현재의 네트워크 피처를 수집해서 기존에 학습된 이상 상태의 특징과 비교하느냐, 정상 상태의 특징과 비교하느냐 에 차이가 있다. 이상 상태의 특징과 비교하여 네트워크 공격을 탐지하는 시그니처 기반 방법은 학습 데이터에 없는 새로운 유형의 공격은 탐지하지 못하는 단점이 있어 최근에는 정상 상태와 비교하는 이상 상태 기반 탐지 기술이 주로 연구되고있다[3].

이상 상태 기반 탐지 기법은 주로 플로우 기반(flow-based)으로 연구가 이루어지고 있다. 현재 플로우의 특징을 수집하여 기존의 정상 상태 데이터 셋과 얼마나 다른가를 판단하기 위해 초기에는 Fuzzy logic 기반의 연구가 존재하였다[4]. 해당 연구는 DoS 공격에는 94%의 높은 F1 score 를 보였으나 그 외의 공격에는 낮은 탐지율을 보였다. 최근에는 분류(Classification) 알고리즘을 적용한 연구들이 주를 이루었다. SVM (Support Vector Machine)을 이용하여 정확도 98.29%를 기록한 연구가 존재한다[3]. 인공 신경망을 이용한 연구도 활발히 진행되고 있는데, NSL-KDD 데이터 셋의 네 가지 공격을 탐지 대상으로 심층 신경망(Deep Neural Network)을 사용하여 F1 score 75.57%의 성능을 보여준 연구가 존재한다[5].

해당 연구들은 주로 인입 포트 및 TCP 플래그 정보를 사용하였기 때문에 현재의 네트워크의 상태에 대한 자세한 정보를 사용하지 못하였다. NSL-KDD 데이터 셋을 사용한 연구의 경우에는 다양한 피처가 존재하여

비교적 높은 성능을 보이긴 하나 이는 네트워크 상에서 실시간으로 수집할 수 없는 정보이다. INT 를 이용하면 네트워크의 정보를 실시간으로 상세히 수집 가능하며, 이를 이용한 선행연구에서는 DoS 공격 탐지에 대해 92.41%의 높은 F1 score 를 보였다[6].

2. 로드 밸런싱 기술

데이터 센터에서 주로 사용하고 있는 로드 밸런싱 알고리즘으로는 ECMP (Equal-cost Multipath)[7]가 있다. ECMP 는 출발지에서 목적지까지의 최단 경로가 여러 개일 때 해당 경로들에 플로우를 균등하게 분배하는 알고리즘이다. 하지만 이는 데이터 센터에 여러 개의 대용량 트래픽이 생성될 경우 같은 경로로 할당되는 문제가 있어 네트워크의 혼잡 정도를 실시간으로 파악하고 이를 로드 밸런싱에 활용하기 위한 연구들이 지속되고 있으며, 해당 연구들은 일반적으로 성능 측정을 할 때 FCT (Flow Completion Time)을 ECMP 와 비교한다.

각 스위치에 이웃한 노드의 트래픽 혼잡 정도를 저장하여 혼잡 정도가 가장 적은 다음 경로를 지정하여 평균적으로 ECMP 에 비하여 1.2 배의 성능을 보인 연구가 있다[8]. 프로그래머블 스위치를 활용하여 주기적으로 탐지 패킷(probe packet)을 전송하여 각 스위치 별로 최적의 다음 홉 정보를 저장하여 평균적으로 ECMP 대비 1.52 배의 성능을 보인 연구도 있다[9].

이러한 연구들은 모든 스위치에 네트워크 전체의 트래픽 혼잡 정보를 저장하여 일반적인 스위치에서는 사용할 수 없는 문제나 추가 패킷을 생성하여 네트워크에 오버헤드를 일으키는 단점이 있다. In-band 네트워크 텔레메트리 기법을 활용하면 추가적인 패킷을 생성하지 않고 P4 스위치를 사용하여 네트워크 상태 정보를 파악할 수 있다[10].

3. INT (In-band Network Telemetry) [2]

INT 는 P4 를 활용한 네트워크 모니터링 프레임워크로, 패킷 헤더에 INT 헤더를 정의하여 별도의 탐지 패킷을 생성하지 않고 네트워크 정보를 수집할 수 있다. P4 프로그램을 통해 데이터 평면 상에 INT 헤더를 삽입하는 Source 스위치, INT 헤더에 네트워크 정보를 삽입하는 Transit 스위치, INT 헤더를 추출하여 전달하는 Sink 스위치를 정할 수 있다. INT 를 통해 수집 가능한 정보는 스위치 ID, 인입 및 인출 포트 관련 정보, 플로우 지연 정보, 스위치 큐 사용량 등이 있다.

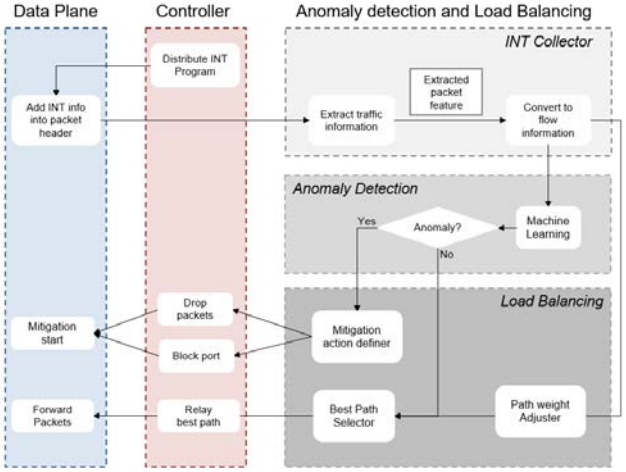

그림 1 INT 기반 네트워크 통합 관리 시스템

Ⅲ. INT 기반 통합 네트워크 관리 시스템

그림 1 은 본 연구에서 제안하는 INT 기반 네트워크 통합 관리 시스템을 나타낸다. 해당 시스템은 INT 를 기반으로 만들어져 실시간으로 네트워크에 대한 관리를 진행할 수 있다. 각 모듈에 대한 상세한 설명은 다음과 같다.

1. INT Collector

INT Collector 는 데이터 평면에서 수집된 패킷 별 INT 정보를 플로우 정보로 변환한다. INT 정보가 수집되는 알고리즘은 다음과 같다. 컨트롤러에서 수집 대상 INT 정보를 정의하여 데이터 평면에 배포한다. 데이터 평면상의 스위치들은 수집 대상 INT 정보를 패킷 헤더에 포함시켜 패킷을 전송한다. Sink 스위치에서 INT 패킷 헤더와 수집 대상 INT 정보를 추출하여 INT 정보만 INT Collector 로 전달한다. 하지만 전달된 데이터는 패킷 별 데이터이기 때문에 플로우 별 데이터로 전환해야 한다. 플로우는 같은 출발지와 목적지를 가지는 일정 시간 이내의 모든 패킷으로 정의된다. INT Collector 는 정의에 따라 패킷 데이터를 플로우 별로 모으고 각 피처를 플로우 별로 평균값을 계산하여 플로우 별 데이터로 전환한 후 Anomaly Detection 모듈과 Load Balancing 모듈로 전달한다.

2. Anomaly Detection

Anomaly Detection 모듈에서는 수집된 피처를 이용하여 해당 플로우가 이상 상태인지 아닌지를 판별한다. Anomaly Detection 모듈에는 사전에 학습된 기계 학습 모듈이 내장된다. 기계 학습 모듈은 INT Collector 에서 전달받은 플로우 별 데이터 입력 피처로 하여 이상 상태 점수를 반환하도록 학습된다. 이 때 사용하는 기계 학습 모델은 가장 많이 쓰이는 지도 학습(Supervised Learning) 모델 중 하나인 순환 신경망 (Recurrent Neural Networks)을 사용한다. 이상 상태 점수는 0 에서 1 사이의 값이다. 이상 상태일 경우 이상 상태 점수가 Load Balancing 모듈로 전달되어 이용될 수 있도록 한다.

Anomaly Detection 모듈에 대한 선행 연구에서는 네트워크 플로우 별로 INT 정보만 수집하여 순환 신경망에서 학습하였다. INT 정보와 함께 TCP 플래그와 출발 및 도착 포트 정보를 함께 사용하면 Anomaly Detection 모듈에서 네트워크 공격에 대해 F1 score 를 95% 이상의 성능을 낼 수 있을 것으로 보인다.

3. Load Balancing

Load Balancing 모듈은 이상 플로우에 대한 완화 작용 선택 모듈 (Mitigation action definer)을 포함하고 있다. 완화 작용 선택 모듈은 Anomaly Detection 모듈에서 이상 상태임을 전달받으면 해당 플로우에 대한 로드 밸런싱을 중지하고 이상 상태 점수에 따른 완화 작용을 선택할 수 있도록 한다. 완화 작용 선택에는 이상 상태 점수를 이용하는데, 이상 상태 점수가 높을 경우 해당 출발 포트를 차단하고, 낮을 경우에는 해당 플로우만 패킷 드랍(packet drop)하는 방식을 선택하여 이상 상태 정도에 맞는 완화 작용을 선택하여 컨트롤러로 전달한다. 컨트롤러는 Load Balancing 모듈에게 전달받은 완화 작용을 해당 플로우에 대해 실행한다.

Load Balancing 모듈은 이상 상태로 판별되지 않은 플로우들에 대해서 경로 가중치 계산을 한다. 가중치 계산에는 INT Collector 에서 전달받은 플로우 별 데이터를 활용한다. Load Balancing 모듈은 사전에

학습된 기계 학습 모델을 이용한다. 해당 기계 학습 모델은 강화 학습(Reinforcement Learning) 기법을 사용한다. 강화 학습 모델은 INT 정보 중 네트워크 각 링크의 처리율(throughput)을 입력 피처로 사용하여 각 출력 포트에 대한 가중치를 반환하도록 학습된다. 이 때 출력 포트에 대한 가중치들은 스위치 별로 가중치의 합이 1 이 되도록 한다. 강화 학습 모델을 학습시킬 때 FCT 를 최소화 시키도록 학습시킨다.

Load Balancing 모듈은 강화 학습의 결과를 컨트롤러에 전달한다. 컨트롤러는 출력 포트들에 대한 가중치를 스위치들에 전달하여 스위치들에서 해당 가중치에 해당하는 비율로 경로를 설정하도록 한다. 강화 학습에서 입력 피처로 네트워크 전체 정보를 사용하였기 때문에 네트워크 일부의 환경이 변하여도 전체 네트워크가 빠르게 대응할 수 있도록 한다. 또한 강화 학습을 통해 주어진 네트워크 환경에서 데이터 수집과 학습을 계속하게 되기 때문에 ECMP 대비 1.2 배 이상의 성능을 보일 수 있을 것으로 보인다.

IV. 결론

본 연구에서는 실시간으로 패킷 단위의 세부적인 네트워크 정보를 수집할 수 있는 INT 를 이용하여 네트워크 이상 상태 탐지와 로드 밸런싱을 모두 할 수 있는 INT 기반 통합 네트워크 관리 시스템을 제안하였다. 해당 시스템은 이상 상태 탐지와 로드 밸런싱에 각각 기계 학습을 적용하여 높은 성능이 기대되며, 특히 두 기능을 모두 포함하고 있기 때문에 네트워크 공격이 일어나면 로드 밸런싱에 이를 활용하여 완화작용까지 진행할 수 있는 이점을 가지고 있다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2018-0-00749, 인공지능 기반 가상 네트워크 관리기술 개발)

참 고 문 헌

[1] P. Bosshart *et al.,* "P4: Programming Protocol-Independent Packet Processors." ACM SIGCOMM Coumputer Communication Review, vol. 44, no.3, pp. 87-95, 2014.

[2] C. Kim, A. Sivaraman, N. Katta, A. Bas, A. Dixit, and L. J. Wobker, "In-band Network Telemetry via Programmable Dataplanes," ACM SOSR, 2015, pp. 2-3.

[3] P. Winter, E. Hermann, and Z. Markus, "Inductive Intrusion Detection in Flow-Based Network Data using One-Class Support Vector Machines," IFIP International Conference on New Technologies, 2011, pp. 1-5.

[4] R. Shanmugavadivu, N. Nagarajan, "Network intrusion detection system using fuzzy logic," Indian Journal of Computer Science and Engineering, 2011, pp.101- 111.

[5] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi,and M. Ghogho, "Deeplearning approach for Network Intrusion Detection in Software Defined Networking," WINCOM, 2016, pp. 258-263.

[6] 임지윤, 남석현, 유재형, 홍원기, "INT 기반 네트워크 이상 상태 탐지 기술 연구", KNOM Review, Vol. 22, No. 3, December 2019.

[7] M. Chiesa, G. Kindler, and M. Schapira, "Traffic Engineering with Equal-Cost-MultiPath: An Algorithmic Perspective," IEEE Conference on Computer Communications, 2014.

[8] M. Alizadeh et al., "CONGA: distributed congestion-aware load balancing for datacenters," Proceedings of the 2014 ACM conference on SIGCOMM,2014.

[9] N. Katta, M. Hira, C. Kim, A. Sivaraman, and J. Rexford, "HULA: Scalable Load Balancing Using Programmable Data Planes," Proceedings of the Symposium on SDN Research,2016.

[10] 임지윤, 현종환, 유재형, 홍원기, "머신러닝 기반 동적 경로 가중치 조정 로드밸런싱 알고리즘 연구," KNOM Conference 2019, Daegu, Korea, May. 30, 2019, pp. 83-86.