사물인터넷을 위한 인공지능 기반의 침입 탐지 시스템에 관한 연구

류정현, 권병욱, 석상기, 박종혁*
서울과학기술대학교 컴퓨터공학과
e-mail:{jh.ryu, rnjsqud123, sksuk, jhpark1}@seoultech.ac.kr

A Study on Artificial Intelligence based Intrusion Detection System for Internet of Things

Jung Hyun Ryu, Byung Wook Kwon, Sang Kee Suk, Jong Hyuk Park*
Department of Computer Science and Engineering,
Seoul National University of Science and Technology (SeoulTech)

요 약

클라우드 컴퓨팅 기반 사물인터넷 환경은 급격히 증가하는 통신량, 기종 간 이질성, 지연 시간과 같은 문제점으로 인해 어려움을 겪고 있다. 이를 해결하기 위한 대표적인 방법 중 하나는 분산 모델을 통해 클라우드 컴퓨팅 환경에 집중된 네트워크 또는 컴퓨팅 파워를 분산시키는 포그 컴퓨팅 (Fog Computing) 또는 에지 컴퓨팅 (Edge Computing)을 활용하는 것이다. 그러나 이 분산형 네트워크의 단점을 보완하기 위해 사물인터넷 (IoT, Internet of Things)과 가장 가까이 존재하는 네트워크 모델로써 미스트 컴퓨팅 (Mist Computing)이 탄생하였다. 그러나 다양한 프로토콜에 의해 통신이 이루어지는 사물인터넷 환경에는 수천 가지 제로데이 공격이 존재한다. 이 공격들의 대부분은 이전에 알려진 공격의 작은 변형체이다. 이러한 공격을 효과적으로 막기 위해 사물인터넷 환경에서의 침입 탐지 시스템은 지능적이어야 한다. 따라서 본 논문에서는, 미스트 컴퓨팅 환경에서 새로운 또는 지속적으로 변화하는 사물인터넷 대상 공격을 효과적으로 방어하기 위한 인공지능 기반 침입 탐지 시스템을 제안한다.

1. 서론

사물인터넷 (IoT, Internet of Things) 환경에서 사용되는 기기의 수는 2020년 까지 약 500억 개에 달할 것으로 예상된다 [1]. 이처럼 기하급수적으로 늘어나는 모바일 기기 (모바일폰, 태블릿PC 등), 무선 센서, 액츄에이터 (Actuator) 등의 스마트 기기들을 안정적으로 관리하고 사용하기 위해서 새로운 기술이 반드시 필요하다. 대다수의 IoT장치 즉, IoT 환경에서의 네트워크 에지에 해당하는 모든 장치는 낮은 대기 시간과 방대한 양의 데이터를 분산 처리 방식으로 수행해야할 필요가 있다. 이에 따라 기존의 클라우드 컴퓨팅 (Cloud Computing) 에서 포그 컴퓨팅 (Fog Computing) 또는 에지 컴퓨팅 (Edge Computing)으로 개념이 확장되었다. 그러나 폭발적으로 증가할 스마트 기기 및 IoT 장치들을 효과적으로 관리하고 사용하기 위해 그 장치들과 가장 가까이 있어, 보다 빠른 접근 방식을 제공하는 컴퓨팅 모델로써 미스트 컴퓨팅 (Mist Computing)이 등장하였다. 이 분산형 네트워크 모델은 네트워크의 최종단에 위치하여 센서 및 액츄에이터를 포함한 스마트 기기들과 가장 가까이 위치하며 통신을 하기 때문에 통신량 분산, 쉬운 기기 간 이질성 통합, 지연 시간 최소화 등의 이점을 가지고 있어 클라우드 컴퓨팅의 문제점을 해소할 수 있다.

그러나 미스트 컴퓨팅 모델은 스마트 기기들과 가장 짧은 거리에 존재하기 때문에 보안 문제는 가장 먼저 고려되어야 할 사항이다. 사물인터넷 환경의 주요 특징 중 하나인 기종 간 이질성으로 인해 다양한 프로토콜이 존재하며, 이는 수 천 가지 제로데이 공격이 탄생하는 배경이 된다. 그러나 이런 공격의 대부분은 이전에 존재하는 공격 유형의 변형체로 알려져 있다. 공격 개체 수의 증가, 기존 공격의 변형체 즉, 알려지지 않은 사이버 공격은 대규모 사물인터넷 환경에 큰 위기를 가져올 수 있다. 사물인터넷 환경에서의 침입 탐지는 그 특수한 요구사항(낮은 대기 시간, 제한적인 자원, 확장성 및 유연성 등)으로 인해 기존의 중앙집중식 클라우드 방식과는 차이가 있다. 때문에 미스트 컴퓨팅 계층으로 확장한 침입 탐지 시스템이 연구되어야 한다.

인공지능 기반의 침입 탐지 시스템은 사물인터넷 환경에서 나타나는 다양한 공격 유형을 탐지하는데 있어 큰 가능성을 열어준다. 공격 유형의 증가 및 공격에 사용되는 컴퓨팅 파워가 증가함에 따라 기존의 기계 학습 알고리즘은 복잡한 사이버 공격을 탐지하는데 한계가 있다. 이러한 공격의 99%는 기존 공격의 변형체이며, 1%의 새로운 공격조차 이전 공격 방식의 개념에 의존한다 [2]. 사물인터넷 환경에서 인공지능 특히, 딥러닝을 이용한 공격 탐지는

작은 변화에 영향을 받지 않으면서 높은 수준의 학습 결과를 보여주어 공격을 탐지하는데 효과적이다. 사물인터넷 환경에서 딥러닝을 사용하는 것의 가장 큰 이점은 수동 피쳐 엔지니어링 (Feature Engineering)의 부재이다. 때문에 자원 제약적인 네트워크 즉, 미스트 컴퓨팅과 같은 경량 환경에서도 학습이 가능하다는 것이다. 따라서 본 논문에서는 사물인터넷 환경을 위해 미스트 컴퓨팅에서의 딥러닝을 적용한 분산 공격 탐지 시스템을 제안한다.

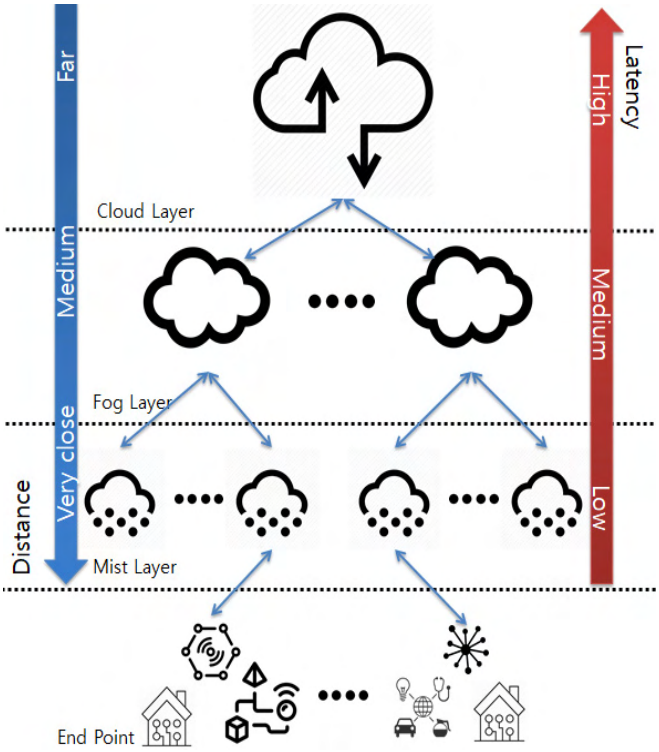
2. 미스트 컴퓨팅 (Mist Computing)

점차 규모화 되는 사물인터넷 환경으로 인해 클라우드 컴퓨팅의 한계점이 드러나게 되었다. 수 억 개의 사물인터넷 기기들이 클라우드 컴퓨팅을 통해 통신, 데이터 처리를 수행함으로써 과도한 통신량, 높은 대기 시간, 기종 간 이질성으로 인한 문제들로 인해 나타난다. 이를 해결하기 위해 클라우드 컴퓨팅과 엔드 포인트 (End Point) 즉, 사물인터넷 기기들과의 사이에서 작용하는 계층화된 모델인 포그 컴퓨팅이 등장하였다. 이 모델은 분산형으로써 클라우드에서 사물인터넷 기기로의 서비스 배포를 용이하게 하며, 포그 노드들로 구성되어 있다. 포그 컴퓨팅은 지원되는 응용프로그램과의 요청 및 응답 시간을 최소화하고 단말 기기에 대해 로컬 컴퓨팅 자원과 필요한 중앙 집중식 클라우드 서비스에 대해 네트워크 연결과 컴퓨팅 파워를 제공한다. 그러나 지속적으로 확장되는 사물인터넷 환경을 효과적으로 관리하기 위해 사물인터넷 계층 즉, 엔드 포인트에서 작동하는 모델이 등장하게 되었다. 이것이 바로 미스트 컴퓨팅 모델이다. 이 모델은 복잡한 응용프로그램을 계산하는 데 있어 효율성을 향상시키는 사물인터넷 환경에서 네트워크 에지 (Network Edge)의 가장 가까운 곳에 위치한다. 데이터를 전송 할 때 제어가 네트워크의 대기 시간을 감소시키는 엔드 노드 (End Node)로 분산되어 처리량을 증가시킨다. 이는 포그 노드와 클라우드 컴퓨팅에 정보를 공급하기 위해 마이크로 컴퓨터와 마이크로 컨트롤러를 사용하여 네트워크 에지에 사물인터넷 기기와 가까운 거리에서 존재하는 가벼운 컴퓨팅 파워를 가진다. 미스트 컴퓨팅은 네트워크 에지에 존재하는 모델로써 다음과 같은 지침 원리를 따른다 [2].

- 네트워크는 단순한 데이터가 아닌 정보를 제공해야 함
- 네트워크는 요청된 정보만을 제공해야 함
- 소비자/공급자 모델을 사용하여 작동하는 최종 장치로써 정보 필요성에 기반한 시스템을 동적 생성해야 함
- 미스트 노드는 환경을 인식해야하며 요구 정보와 네트워크 구성에 적응해야 함

클라우드와 포그 컴퓨팅 모델은 사용자의 요구와 네트워크 전역 환경에 대한 정보를 가지고 있는 반면, 미스트 컴퓨팅 모델은 물리적 환경과 지역 환경에 대한 정보만을 가지고 있으므로 주 목적은 사물인터넷 응용프로그램을 실행하는 것이다. 이를 위해서 네트워크 전역 환경에 대한 정보를 에지 즉, 미스트 노드에 전달해야 한다.

다음 (그림 1)은 미스트 컴퓨팅의 전반적인 개요를 설명한다.



(그림 1) 미스트 컴퓨팅의 개요

3. 인공지능 (Artificial Intelligence)

사물인터넷 환경에서 사이버 공격에 대한 침입 탐지를 위해 인공지능이 효과적으로 적용된다. 인공지능은 기계학습 (Machine Learning), 인공신경망 (Artificial Neural Network), 딥러닝 (Deep Learning) 등으로 분류된다. 그 중, 딥러닝은 학습 단계를 기존 인공지능 접근보다 심화한 것으로써, 작은 변화에 가장 영향을 적게 받으며 정확도가 우수하다. 때문에 본 논문에서 제안하고자 하는 침입 탐지 시스템의 공격 유형 학습을 위해 딥러닝이 가장 적합하다.

인공지능을 통한 침입 탐지 시스템에서의 공격 탐지는 시그니처 기반 (Signature based) 또는 이상동작 (Anomaly based) 기반으로 나누어진다. 시그니처 기반 탐지 방식은 들어오는 네트워크 트래픽을 데이터베이스에 존재하는, 이미 알려진 공격 유형과 비교하여 탐지한다. 반면, 이상동작 기반 탐지 방식은 정상 트래픽과의 동작 편차를 계산하여 공격을 탐지한다.

시그니처 기반 탐지 방식은 구현이 용이하고 탐지의 정확도가 높으며, 오탐지율이 낮기 때문에 침입 탐지 시스템에서 널리 사용되었지만 변형된 공격이나 새로운 공격에 대한 탐지율이 현저히 낮다. 반면, 이상동작 기반 탐지 방식은 정확도가 비교적 낮지만 변형된 공격 또는 새로운 공격을 높은 확률로 탐지할 수 있다.

인공지능 기반의 침입 탐지 시스템을 시뮬레이션 하고 성능을 평가하기 위한 수단으로써 데이터 셋을 사용한다.

네트워크에서 송수신이 이루어지는 트래픽 중 정상과 비정상을 분류해놓은, 일종의 트래픽 모음이다. 본 논문에서 침입 탐지 시스템을 위해 고려하는 데이터 셋은 KDD'99 데이터 셋과 NSL-KDD 데이터 셋이다.

KDD'99 데이터 셋은 기존 침입 탐지 시스템을 통해 수집된 데이터를 기반으로 만들어진 데이터 셋으로써 각 공격 유형을 레코드별로 분류하였다. 이는 지속기간, 프로토콜의 유형 등 다양한 속성을 포함하고 있으며 크게 네 가지 클래스로 분류한다 [3].

NSL-KDD 데이터 셋은 2009년 KDD'99의 문제점을 보완하기 위해 만들어진 것으로써 KDD'99의 확장형 데이터 셋이다. KDD'99 데이터 셋에서 중복되는 레코드를 제거하고 데이터 셋의 전체적인 크기를 줄이는 방향으로 수정하였다. 결과적으로 KDD'99 데이터 셋과 공격 유형 분포에 작은 차이를 보인다 [4].

두 데이터 셋에서의 공격 유형은 크게 Normal, DoS, Probe, R2L, U2R로 분류된다. 데이터 셋 내에 존재하는 모든 레코드는 위의 다섯 가지 유형으로 분류되며 침입 탐지 시스템이 공격 유형을 분류하는 기준이 된다.

아래의 <표 1>은 데이터 셋의 공격 유형 분류를 설명한다.

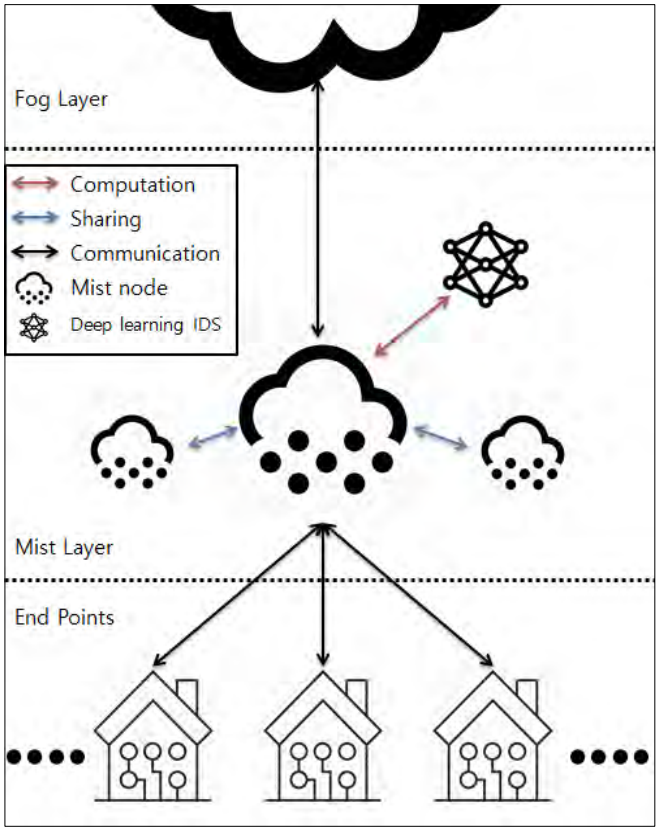
<표 1> KDD'99 및 NSL-KDD 데이터 셋의 공격 유형 분류

분류	설명
Normal	정상 트래픽
DoS	서비스 거부 공격 (Denial of Service)을 시도하는 트래픽
Probe	공격 전 정보 수집을 시도하는 트래픽
R2L	권한이 없는 상태에서 권한을 획득하려는 트래픽
U2R	일반 사용자가 루트 권한을 획득하려는 트래픽

NSL-KDD 데이터 셋은 KDD'99의 단점을 보완하기 위해 만들어진 데이터 셋으로써 전반적으로 우수한 탐지율을 보이지만 특정 공격 유형에 대한 탐지율이 KDD'99 보다 떨어진다는 연구결과가 있다 [4]. 그러므로 본 논문에서는 특정 공격 유형에 대한 탐지율 저하를 방지하기 위해 두 개의 데이터셋을 사용한다.

4. 딥러닝을 이용한 분산 침입 탐지 시스템

미스트 노드는 사물인터넷 환경의 스마트 기기들에 클라우드 및 포그 노드보다 가까이 존재하기 때문에 공격 유형을 학습하고 대응하는 데에 보다 효율적이다. 각 미스트 노드는 딥러닝 기반 침입 탐지 시스템과 통신하여 트래픽을 분석하고 분류한다. 개별 미스트 노드는 소규모의 사물인터넷 스마트 기기와 통신하기 때문에 지역적 정보만을 가지고 있지만, 인접한 미스트 노드와 학습 데이터를 공유하여 공격 유형을 분석한다. 지역적 학습 및 학습 데이터 공유를 통하여 탐지 및 대응이 보다 빠르다. 이 논문의 침입 탐지 시스템을 위한 접근 방식은 다음 (그림 2)와 같다.



(그림 2) 딥러닝 기반 분산 침입 탐지 시스템

가장 아래에 위치하는 엔드 포인트 계층은 사물인터넷 환경에서 실제 작동하는 스마트 기기의 계층이다. 가운데 미스트 계층은 엔드 포인트와 가장 인접한 거리에서 스마트 기기들과 통신하는 계층으로써 스마트 기기들로부터 들어오는 트래픽을 딥러닝 기반 침입 탐지 시스템을 통해 분석하고 판단하여 정상 혹은 비정상 트래픽으로 분류한다. 각 미스트 노드는 통신하는 스마트 기기들로부터 받은 트래픽을 인접 미스트 노드와 공유함으로써 침입 탐지 시스템을 지속적으로 업데이트한다. 가장 위의 계층은 중간 규모의 포그 계층으로, 특정 규모의 미스트 노드들과 통신하여 클라우드와의 통신을 중개한다.

5. 결론

날마다 증가하는 사물인터넷 환경의 크기는 클라우드 컴퓨팅의 도전과제이다. 과도한 통신량, 높은 대기 시간, 보안 이슈 등의 한계점은 포그, 미스트 컴퓨팅과 같은 분산형 컴퓨팅 접근으로 해결 가능하다. 그 중 미스트 컴퓨팅

은 클라우드의 한계점을 극복하기 위한 최신의 접근법으로써 각광받고 있다.

따라서 본 논문에서는 사물인터넷 환경을 위한 인공지능 중 가장 적응력 및 정확도가 높다고 평가되는 딥러닝을 기반으로 미스트 컴퓨팅 환경에서의 침입 탐지 시스템의 개요를 제안한다. 이는 분산형 침입 탐지 시스템을 적용함으로써 탐지의 정확도를 높일 수 있으며 스마트 기기들과 가장 인접한 미스트 컴퓨팅 계층에서 학습을 수행하여 보다 빠르게 학습 및 탐지가 가능하다. 가까운 미래에 본 논문에서 제시하는 침입 탐지 시스템에 적용할 학습알고리즘, 시뮬레이션 등에 대한 세부적인 연구를 진행할 것이다.

Ackwoledgement

이 논문은 2018년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임
(No. 2016R1A2B4011069).

참고문헌

[1] Michaela Iorga et al., “Fog Computing Conceptual Model”, NIST Special Publicaion 500-325, 2018

[2] Abebe Abeshu Diro, Naveen Chilamkurti, “Distributed attack detection scheme using deep learning approach for Internet of Things”, Future Generation Computer Systems, Elsevier, 82, 2018, pp. 761-768.

[3] 김경민, 김광조, “침입 탐지 시스템의 알려지지 않은 공격 탐지에 대한 최신 연구 비교”, Probe, 5, pp. 1-7.

[4] 지현정 외, “인공신경망을 통한 KDD CUP 99와 NSL-KDD 데이터 셋 비교”, 2017년 정보처리학회 춘계학술발표대회 논문집, 24(1), 2017, pp. 211-213.

[5] Manas Kumar Yogi et al., "Mist Computing: Principles, Trends and Future Direction." SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE), 4(7), 2017, pp. 19-21.

[6] Abebe Abeshu Diro, Naveen Chilamkurti, “Deep Learning: The Frontier for Distributed Attack Detection in Fog-to-Things Computing”, IEEE Communications Mag., 56(2), 2018, pp. 169-175.

[7] Nathan Keegan et al. "A survey of cloud-based network intrusion detection analysis." Human-centric Computing and Information Sciences, 6(1), 2016, 19.