SOM 알고리즘 기반 DoS 트래픽 탐지

신무곤○, 심규석, 구영훈, Huru Hasanova, 김명섭

고려대학교 컴퓨터정보학과

{tm0309 , kusuk007, gyh0808, hhuru, tmskim}@korea.ac.kr

요 약

네트워크 기반의 공격은 그 위험성과 피해의 규모가 크기 때문에 공격 초기에 빨리 탐지하는 것이 중요하다. 그러나 네트워크상의 비정상 트래픽을 탐지하는 것은 방대한 양의 데이터 전처리와 관리자의 분석이 요구된다. 또한 이러한 관리자의 분석이 정확하다는 보장이 없을 뿐 아니라 각 네트워크의 실시간 특성을 고려해야하기 때문에 탐지의 어려움이 크다. 이러한 한계를 극복하기 위해 본 논문에서는 데이터마이닝 기법 중 하나인 자기 조직화 지도(SOM)를 활용한 트래픽 분류 시스템을 제안한다. 제안된 방법론은 분류 과정에서 정확도 향상 및 실시간 탐지에 대한 기대 효과를 나타낼 수 있다.

1. 서론

네트워크 기반의 공격에 속하는 DoS(Denial of Service) 및 DDoS(Distributed Denial of Service) 공격은 타겟 시스템에 대량의 트래픽을 보냄으로써 다른 정상적인 이용자들이 이 시스템의 서비스를 제공받지 못하고, 네트워크 전체를 마비시킬 수 있는 위험한 공격이다. 이러한 공격들은 웜 바이러스 등 지능적인 공격 툴을 이용하여 빠르게 확산되고, 그 피해가 매우 커지고 있는 실정이다. 따라서 이러한 공격 트래픽이 확산되기 전에 빨리 탐지하는 것이 중요하다.

기존의 침입 탐지 시스템들은 이미 알려진 공격에 대한 시그니쳐를 수동으로 시스템에 인코딩하여 침입을 판단하였다. 그러나 수동적인 방법에 의한 시그니쳐 생성 및 업데이트는 매우 어려운 일이며 그 효율성도 떨어진다. 이러한 문제를 해결하기 위하여 인공지능, 기계학습 기법들을 침입 탐지에 적용하는 연구가 늘어나고 있다.

기존의 침입 탐지 시스템이 가지고 있는 문제를 해결하기 위한 연구가 시도되고 있으나, 아직까지 많은 연구가 지도 학습(supervised learning) 알고리즘에 근간을 두고 있다. 하지만 탐지 과정 전에 충분한 학습 과정이 이루어져야 하므로 안정적인 성능이 나오기까지 많은 비용이 든다. 그리고 학습을 위해 많은 양의 분류되어 있는 데이터를 필요로 하므로 학습 데이터의 질에 의해 탐지 성능이 크게 좌우된다. 또한 실시간 네트워크 패킷 탐지가 불가능하며 학습된 데이터 이외의 새로운 데이터에 대한 탐지가 어렵다.

이 논문은 2015 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No.2015R1D1A3A01018057)과 2017 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No. 2017-0-00513, Security Analytics 기반의 이기종 보안솔루션 위협 분석 및 대응 기술 개발)

따라서 본 논문에서는 데이터 마이닝 기법 중 하나인 Self Organizing Maps(SOM)을 소개하고 DoS 공격 패킷에 대한 SOM 알고리즘 기반 침입 탐지 메커니즘을 제안한다.

본 논문은 본 장 서론에 이어, 2 장에서 자기 조직화 지도 구조에 대해 제안하고, 마지막 3 장에서 결론 및 향후연구에 대해 언급한 후 논문을 마친다.

2. 본론

1)SOM(Self Organizing Maps)

자기 조직화 지도(Self Organizing Map) 알고리즘은 비지도 학습(unsupervised learning)을 통한 데이터 마이닝 기법으로 스스로 학습하는 인간 뇌의 학습 과정과 유사한 성격을 가진다. SOM 은 외부로부터 데이터를 받아 특정 사이즈의 노드로 구성된 맵에 데이터의 특징들을 학습한다. 여러 특징들을 가진 데이터를 반복해서 학습하면, SOM 이 갖고 있는 맵에는 비슷한 패턴을 가진 데이터를 기억할 수 있는 학습 영역이 생긴다. 충분히 많은 데이터를 학습한 맵에 새로운 데이터를 학습하게 되면, 해당 데이터가 맵의 구분된 영역 중에 어떤 부분에 속하는지에 따라 어떤 특징을 가졌는지 예측할 수 있다.

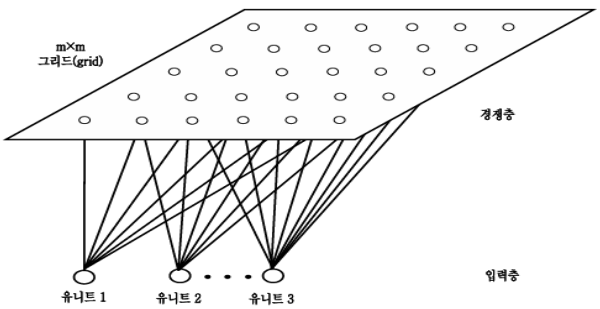

Figure 1. SOM 의 아키텍처

2)학습과정

SOM 은 입력층과 경쟁층으로 구분된 2 개의 레

이어를 기본으로 한다. 입력층에서는 입력벡터를 입력받고 경쟁층에서는 입력벡터와 경쟁층 노드와의 거리를 계산하여 경쟁에서 이긴(거리가 가까운) 데이터들이 클러스터링 된다.

입력벡터는 네트워크에서 입력받는 데이터이며, 경쟁층 노드의 가중치 값은 0 에서 1 사이의 값으로 초기화 된다. 입력벡터와 경쟁층 노드의 가중치와의 거리를 계산하여 입력벡터와 가장 유사한 가중치 값을 가진 경쟁층 노드를 선택한다. 경쟁층의 노드를 선택한 다음 해당 노드와 노드의 이웃노드들의 가중치를 수정한다. 이웃노드들의 선택은 가우시안 분포를 활용한다.

입력벡터와 노드간의 거리는 아래와 같이 정의한다.

$$D_{ij} = \sum_{i=1}^{n}(w_{ij} - x_i)^2$$

위너 노드와 그 이웃들의 가중치는 다음과 같이 수정된다.

$$w_{ij}(new) = w_{ij}(old) + \alpha(t)(x_i - w_{ij}(old))$$

우변의 두번째 항에 있는 α(t)는 학습률을 나타내며, 알고리즘이 반복될수록 값이 작아진다.

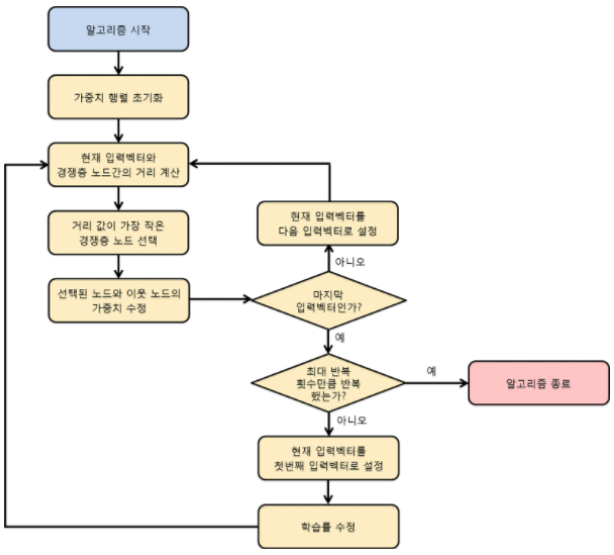

Figure 2. 전체 알고리즘 흐름도

3)SOM 구조 제안

이와 같은 SOM 알고리즘(비지도 학습 데이터 마이닝)은 입력 데이터에 대해 어떠한 정보도 주지 않으므로 그 결과에 대한 해석이 힘들다는 단점이 있다. 따라서 본 논문에서는 지도 학습(supervised learning)과 비지도 학습(unsupervised learning)을 결합한 방법론을 제안한다. 제안된 방법론은 비지도 학습을 사용하지만 패킷의 구분을 위해 일부 지도 학습의 특성을 이용하여 비지도 학습만을 사용할 때의 문제점을 일부 보완할 수 있다.

전체적인 구조는 SOM 과 동일하지만, 초기 학습을 수행 할 때에 데이터와 함께 정답지도 함께 입력함으로써 클러스터링 된 데이터가 정답을 가질 수 있게 한다. 초기 학습 과정을 거친 후에는 침입 탐지와 학습이 동시에 이루어 진다. 이와 같은 방법은 새로운 데이터가 입력되더라도 가장 유사한 클러스터에 일치시키기 때문에 유연한 탐지가 가능하다. 또한 새로운 공격 유형이 발생하더라도 탐지가 가능하다는 장점이 있다.

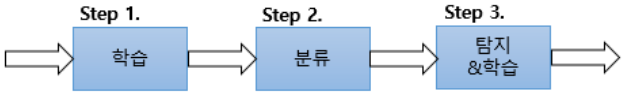

Figure 3. 탐지 단계

학습을 통하여 만들어 지는 cluster map 을 생성하는 단계에서 특정 속성이 맵 형성에 너무 많은 영향을 미치는 것을 방지하기 위해 모든 속성 값을 0-1 사이의 값을 갖도록 정규화하는 과정이 필요하다. 그리고 경쟁층 노드의 값을 초기화 시킨 후 입력벡터와 경쟁층 노드간의 거리가 가장 짧은(유사한) 노드를 선택하고 그 노드와 이웃노드의 값을 갱신한다. 입력벡터가 모두 학습될 때까지 이 과정을 반복한다. 이렇게 클러스터링 된 데이터를 가지고 공격별 규칙을 생성하여 맵상의 위치를 구분할 수 있도록 한다.

이 단계까지 마치면 공격 탐지에 필요한 맵이 형성된다. 이 맵을 바탕으로 실시간 입력 데이터가 맵에 유사한 노드에 일치하게 되고 그 부분이 공격 클러스터라면 비정상, 정상 클러스터라면 정상 트래픽으로 탐지가 가능하게 된다. 또한 실시간 탐지와 함께 학습 또한 이루어지므로 계속해서 맵이 갱신되어 실시간 네트워크 트래픽의 특성을 반영하게 된다.

3. 결론 및 향후 연구

기존의 SOM 알고리즘은 입력 데이터에 대한 어떠한 정보도 주지않기 때문에 클러스터링 된 데이터 결과에 대한 해석이 힘들다. 이러한 단점 때문에 실시간 패킷 탐지와 학습에 어려움이 있다. 하지만 본 논문에서는 이러한 단점을 극복하기 위하여 지도 학습 기법이 추가된 비지도 학습 기법을 제안하였다. 제안된 기법을 사용하면 적은 양의 학습 데이터만으로도 충분한 모듈을 구성할 수 있으며 학습에 대한 점진적 갱신이 가능하기 때문에 실시간 학습을 통한 유연한 패킷 탐지가 가능하다.

향후 연구로는 공격 패킷과 정상 패킷의 특징적인 시그니쳐 추출을 통한 데이터 셋 구성과 초기 학습 모듈 구성, 그리고 이를 바탕으로 한 실시간 공격 트래픽 탐지 시스템을 연구할 계획이다.

4. 참고 문헌

[1]황경애, “자기 조직화 지도(SOM)를 이용한 실시간 침입 탐지 메커니즘” , 2005.
[2]김민회, “분산 Self Organizing Map 기법을 이용한 효과적인 DoS 탐지 시스템”, 2015.