침입 탐지 시스템의 알려지지 않은 공격 탐지에 대한 최신 연구 비교

김경민, 김광조
카이스트 전산학부

Comparison of Recent Unknown Attack Detections in Intrusion Detection System[1)2)]

Kyung-min Kim, Kwangjo Kim
School of Computing, KAIST.

요 약

침입 탐지 시스템(IDS, Intrusion Detection System)은 통상적으로 네트워크에서 비정상적이나 악의적인 행위를 탐지하는 시스템으로, 네트워크 보안을 위해 매우 중요한 시스템이다. 대표적인 탐지 방식으로 흔적 기반 탐지(Signature-based Detection)와 비정상 행위 기반 탐지(Anomaly-based Detection)가 있다. 한편 네트워크 환경이 변화하고 인터넷이 발달하면서 새로운 형태의 공격이 지속적으로 출현하고 있다. 이에 따라 기존에 알려진 공격을 탐지하는 것 외에 알려지지 않은 공격을 탐지하는 것 또한 IDS의 중요한 필요조건이다. 하지만 기존에 알려진 공격의 패턴을 탐지하는 흔적 기반 탐지방식은 그 특성상 알려지지 않은 공격을 탐지하지 못한다. 이에 비해 사용자의 이상 행위를 탐지하는 비정상 행위 기반 탐지방식은 알려지지 않은 공격을 탐지하는 기법이라 할 수 있다. 본 논문에서는 비정상 행위 기반 탐지방식에서 알려지지 않은 공격을 탐지하는 기법을 소개하고, 이에 관한 최신 연구 결과를 비교한 후 향후 연구 방향을 제안한다.

I. 서론

네트워크 혹은 특정 시스템에서 사용자의 각종 비정상적이거나 악의적인 행동을 찾아내는 기법을 침입 탐지(Intrusion Detection)라고 한다. 침입탐지 기법은 크게 알려진 공격들의 패턴을 비교하여 탐지하는 흔적 기반 탐지방식과 정상적인 상태에서 벗어난 이상 행위를 탐지하는 비정상 행위 기반 탐지방식으로 나뉜다.

한편 다양한 네트워크의 출현과 인터넷의 발전과 함께 기존 공격 형태가 다양해지고 새로운 공격도 출현하고 있다. 하지만 흔적 기반 탐지방식은 기존 공격의 패턴을 비교하는 방식의 한계를 지니고 있어 알려지지 않은 공격을 탐지하지 못하는 단점을 지니고 있다. 이에 비해 비정상 행위 기반 탐지방식은 기존 공격 패턴이 아닌 사용자의 이상 행위를 탐지하는 방식이므로 알려지지 않은 공격을 탐지할 수 있는 가능성을 가지고 있다.

본 논문에서는 비정상 행위 기반 IDS에서 알려지지 않은 공격을 탐지하는 데에 이용되는 기법들을 소개하고, 최신 연구 결과를 살펴보고 성능 비교를 한 후 이를 바탕으로 향후 연구 방향을 제안한다.

1) 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [1391104001, 생체모방 알고리즘(Bio-inspired Algorithm)을 활용한 통신기술 연구]
2) This research was supported by the KUSTAR-KAIST Institute, under the R&D program supervised by the Korea Advanced Institute of Science and Technology (KAIST), South Korea.

II. 관련설명

1. 침입 탐지

침입 탐지는 호스트나 네트워크에서 발생할 수 있는 악의적인 공격이나 비정상 행위들을 분석하고 감시하는 기능이다[1].

IDS는 침입 탐지를 위한 기법들을 실제 시스템으로 구현한 것으로써, 네트워크 혹은 호스트를 오가는 트래픽 혹은 데이터를 측정하고 이를 이용해 침입을 탐지한다. IDS의 센서 설치 위치에 따라 네트워크 기반 IDS(NIDS), 호스트 기반 IDS(HIDS)로 나뉜다. NIDS는 센서가 네트워크 라우터 등 네트워크 전체를 감시할 수 있도록 설치된 것이며, HIDS는 센서가 호스트에만 설치된 것이다. 일반적으로 IDS는 통상 NIDS를 의미한다.

IDS가 침입을 탐지했을 경우 그 결과를 신속히 네트워크 관리자에게 보고해야한다. 이 과정에서 IDS는 각종 악의적인 공격에 대해 높은 정확도로 공격 여부를 판단해야 한다. 이러한 IDS의 성능 판단 척도로는 탐지율(Detection Rate, DR)과 거짓 양성율(False Positive Rate, FPR)이 있다. 이들은 다음과 같이 계산될 수 있다.

$$DR = \frac{\text{탐지한 공격 수}}{\text{전체 공격 수}}$$

$$FPR = \frac{\text{공격을 정상으로 판단한 수}}{\text{정상으로 판단한 수}}$$

DR이 낮을 경우 공격을 제대로 탐지하지 못한다는 뜻으로 IDS 성능의 가장 중요한 척도이다. FPR이 높을 경우 실제 공격을 정상적인 행위라고 판단하는 비율이 높다는 뜻으로 FPR 또한 성능 평가의 중요한 척도 중 하나이다.

2. KDD'99 Dataset

KDD CUP'99는 DARPA 계획의 일환으로 네트워크상에서 정상적인 연결과 비정상적인 연결을 가지고 있는 트래픽을 포함하며, IDS의 성능에 관한 객관적인 평가에 사용할 Dataset이다[2]. 이 Dataset은 1,800만개의 패킷 헤더를 가지고 있으며 크게 Probe, DoS, U2R, R2L의 4가지 공격 방식과 일반 패킷인 Normal로 나뉘어져 있다. 각 패킷의 의미는 다음과 같다.

- 일반 : 정상적인 패킷

- Probe : 실제 공격을 시도하기 전 시스템의 사전 자료(포트 등)를 수집하는 패킷

- DoS : Denial of Service. 서비스 거부 공격을 시도하는 패킷

- U2R : User to Root. 관리자(root) 권한을 얻으려 시도하는 패킷

- R2L : Remote to Local. 권한 없는 사용자가 외부에서 접근 권한을 얻으려 하는 패킷

KDD'99 Dataset에서 각 패킷의 분포는 [표 1]과 같다.

[표 1] KDD'99 Dataset 패킷 분포

패킷 유형	패킷 수	비율(%)
일반	80,767	26.0
Probe	5,356	1.7
DoS	223,488	71.9
U2R	228	0.0
R2L	1,376	0.4
합	311,029	100

3. 알려지지 않은 공격 탐지

알려진 공격의 패턴을 비교해 공격을 탐지해내는 흔적 기반 탐지방식은 패턴이 없는 알려지지 않은 공격에 대해서는 탐지할 수 없다. 이에 비해 비정상 행위 기반 탐지방식은 정상적인 범주를 벗어난 행위를 탐지하기 때문에 알려지지 않은 공격을 탐지할 수 있다.

비정상 행위 기반 탐지방식에서 알려지지 않은 공격을 탐지하기 위해서는 정상적인 행위의 범주를 설정하는 것이 중요하다. 이를 더 견고하고 정확하게 모델링하기 위해 현재 기계학습 및 데이터마이닝의 많은 기법들이 이용되고 있다.

III. 최신 연구 결과 비교

알려지지 않은 공격 탐지를 위해 현재 많은 기법들이 제안되고 있지만 그 중 기계학습과 데이터마이닝 기법들이 다양하게 이용되고 있다. 본 논문에서는 알려지지 않은 공격 탐지에 기계학습과 데이터마이닝 기법을 적용한 최신 논문 중 일부를 조사하였다.

Laskov 등[3]은 의사 결정 트리, k-근접 이웃 알고리즘, 다중 계층 퍼셉트론, k-평균 클러스터링, SVM 등 여러 가지 기계학습의 지도 학습과 비지도 학습 알고리즘을 침입 탐지에 적용하여 성능을 비교했다. 각각에 대한 성능 비교를 ROC(Receiver Operator Characteristic) 곡선으로 나타내었는데, 지도 학습 알고리즘 중에서는 최고 약 83%의 DR과 10%의 FP의 성능을 나타내었고, 비지도 학습 알고리즘은 최고 약 77%의 DR과 10%의 FPR을 나타내었다.

Bahrololum 등[4]은 인공 신경망(ANN, Artificial Neural Network)과 SOM(Self Organizing Map)을 조합해 비지도 학습을 통한 알려지지 않은 공격 탐지를 시도하였다. 이러한 기법을 통해 KDD'99 Dataset의 각 공격의 유형에 따라 최고 99.38%의 DR, 3% FPR을 나타내었고, 최저 69.56%의 DR, 28.35%의 FPR을 나타내어 공격 유형별로 성능에 많은 차이를 보였다.

Casas 등[5]은 Sub-space clustering, Evidence accumulation, Outlier detection 알고리즘을 조합한 IDS를 개발하여 알려지지 않은 공격을 탐지하였다. 이 논문에서는 전체적으로 90%의 DR과 약 4%의 FPR을 성능을 보이지만 각 공격의 유형에 따라 DR이 90%, 80%, 50%을 나타내는 등 유형별로 탐지 성능에 많은 차이가 나타났다.

Rassam 등[6]은 인공 면역 네트워크(aiNet, Artificial Immune Network)를 클러스터링에 이용하여 IDS를 구현하였는데, 81%의 DR과 19%의 FPR을 나타냈다. 그리고 Hosseinpour 등[7]은 참고문헌 [6]과 비슷한 시도로 인공 면역 시스템(AIS, Artificial Immune System)과 비지도 학습 알고리즘인 DBSCAN, k-평균 클러스터링을 각각 조합하여 성능을 비교하였다. 이 논문에서 나타난 성능으로는 DBSCAN을 이용하였을 때 DR 77.1%, FPR 0.8%를 나타내었고 k-평균 클러스터링을 이용하였을 때 DR 60.7%, FPR 15.6%를 나타내었다.

각각의 논문들에서 제안한 IDS의 알려지지 않은 공격에 탐지 성능 비교는 [표 2]와 같다. [표 2]에서 보여지는 바와 같이 기계학습 및 데이터마이닝 기법을 적용한 IDS의 대부분이 비지도학습 알고리즘을 이용하고 있다. 알려지지 않은 공격은 기계학습 알고리즘의 학습데이터 상에서 나타나지 않은 공격이기 때문에 지도학습 알고리즘을 탐지 알고리즘으로 적용하기 어렵기 때문이다.

[표 2] 제안된 IDS의 알려지지 않은 공격 탐지에 관한 성능 비교 표

Author	ML	Algorithm	일반(%)	Probe(%)	DoS(%)	U2R(%)	R2L(%)	DR(%)	FPR(%)
[3]	지도학습	SVM	-	-	-	-	-	83	10
	비지도학습	Gamma	-	-	-	-	-	77	10
[4]	비지도학습	ANN + SOM	76.21	69.56	99.38	-	94.38	84.9	26.7
[5]	비지도학습	Sub-space clustering + Evidence accumulation + Outlier detection	-	100	83	88	90	90	4
[6]	비지도학습	aiNet	-	-	-	-	-	81	19
[7]	비지도학습	AIS + k-평균 클러스터링	-	-	-	-	-	60.7	15.6
	비지도학습	AIS + DBSCAN	-	-	-	-	-	77.1	0.8

IV. 결론 및 향후 연구

본 논문에서는 비정상 행위 기반 IDS에서 알려지지 않은 공격을 탐지하는 기법을 제안하는 최신 논문들을 소개하였고, 그 기법들이 적용된 IDS의 성능을 비교하였다.

현재 IDS의 알려진 공격을 탐지하는 성능이 약 99%의 DR과 0.001%의 FPR을 나타내는 데에 비해[8] 알려지지 않은 공격을 탐지하는 성능은 [표 2]에서 볼 수 있듯이 낮은 성능을 보여주고 있다. 그리고 한 가지 알고리즘을 활용하는 것보다 여러 가지 알고리즘을 조합하여 알려지지 않은 공격을 탐지하는 기법이 더 높은 성능을 나타내는 것을 볼 수 있다. 이를 통해 현재 제안된 알고리즘들과 그 외에 ACO(Ant Colony Optimization), PSO(Particle Swarm Optimization) 등과 같은 Bio-inspired 알고리즘[9]이나 다른 여러 가지 알고리즘을 조합하여 실험해 봄으로써 알려지지 않은 공격 탐지에 더욱 높은 성능을 나타낼 수 있는 조합을 찾는 것이 중요할 것이다.

또한 본 연구는 KDD99 Dataset을 기반으로 성능 비교를 하였으나 Dataset이 1999년에 샘플링 되어서 일반 패킷의 종류와 공격 기법 등이 현재 네트워크 환경과 많이 다를 수 있다. 따라서 더욱 정확한 성능 평가를 위해 KDD'99 Dataset 이후 공개된 IDS Dataset인 UNB ISCX Dataset[10]을 이용한다면 더 유용한 결과를 얻을 수 있을 것이다.

[참고문헌]

[1] Karen Scarfone and Peter Mell, "Guide to intrusion detection and prevention systems", NIST Special Publication 800, 2007.

[2] Charles Elkan, "Results of the KDD'99 classifier learning", ACM SIGKDD Explorations Newsletter 1.2 , 2000, 63-64.

[3] Pavel Laskov, *et al.*, "Learning intrusion detection: supervised or unsupervised?", Image Analysis and Processing-ICIAP 2005, Springer Berlin Heidelberg, 2005, 50-57.

[4] M. Bahrololum, *et al.*, "Anomaly intrusion detection design using hybrid of unsupervised and supervised neural network", International Journal of Computer Networks & Communications(IJCNC) 1.2, 2009, 26-33.

[5] Pedro Casas, *et al.*, "Unsupervised network intrusion detection systems: Detecting the unknown without knowledge", Computer Communications 35.7, 2012, 772-783.

[6] Murad Abdo Rassam and Mohd Aizaini Maarof, "Artificial immune network clustering approach for anomlay intrusion detection", Journal of Advances in Information Technology 3.3, 2012, 147-154.

[7] Farhoud Hosseinpour, *et al.*, "Artificial Immune System Based Intrusion Detection: Innate Immunity using an Unsupervised Learning Approach", International Journal of Digital Content Technology & its Applications 8.5, 2014.

[8] O. Y. Al-Jarrah, P. D. Yoo, and K. Kim, "Large-Scale Network Intrusion Detection", UAE Forum on Information and Communication Technology Research (ICTRF), May 11-13, 2014, Abu Dhabi, UAE.

[9] Dario Floreano and Cluadio Mattiussi, "Bio-inspired artificial intelligence: theories, methods, and technologies", The MIT Press, 2008.

[10] Ali Shiravi, *et al.*, "Toward developing a systematic approach to generate benchmark datasets for intrusion detection", Computers & Security, Vol.31.3, 2012, 357-374.