# Swarm Intelligence를 이용한 침입 탐지 시스템의 방식 비교

이동수, 김광조

KAIST 전산학과

# Comparison of Swarm Intelligence-based Intrusion Detection Systems[1)2)]

Dongsoo Lee, Kwangjo Kim

Computer Science, KAIST.

**요 약**

침입탐지 시스템은 통상 네트워크를 통하여 비정상적이나 악의적인 행동을 찾아내는 시스템으로, 네트워크 보안을 위하여는 매우 중요한 시스템이다. 이 논문에서는 기존의 침입탐지 시스템의 탐지율을 높이기 위하여 새로운 기법중 하나인 각종 Swarm Intelligence 방식을 소개하고, 이 방식을 이용한 침입 탐지 시스템을 기술하고 각각 장단점을 비교 분석하고 향후 연구 방향을 제안한다.

키워드 : SI, IDS, ACO, PSO, ACC

## I. 소개

침입탐지(Intrusion Detection)은 통상 네트워크에서 각종 비정상적이나 악의적인 행동을 찾아내는 기법이며, 인터넷의 급격한 발전과 함께 공격 형태도 다양하게 발전하여 이에 대응하는 효과적인 탐지 기술은 대단히 중요하다.

침입탐지 시스템 (IDS : Intrusion Detection System)은 크게 Snort와 같이 패턴을 지정하여 찾는 방법과 기계 학습이나 인공지능 기법을 이용하여 Dataset에서 공격 여부를 스스로 학습하는 시스템 등 이 있다.

이 논문에서는 생체 모방 (Bio-inspired) 기술 중 하나인 Swarm Intelligence(SI)에 대한 소개와 SI를 IDS에 적용하는 방법, 그리고 SI를 적용한 다양한 IDS의 성능 비교 [KGM11]를 분석한 다음, SI 기반 IDS에 대한 추후 연구 방향을 제안한다.

## II. 관련 설명

### II.1 침입탐지

침입 탐지는 호스트나 네트워크에 발생할 수 있는 악의적인 공격이나 비정상 행위들을 분석하고 감시하는 기능이다[SM07].

IDS는 침입 탐지 기법들을 시스템화한 것으로써, 기존의 공격들의 특징에 대해 정리한 Knowledge base가 필요하며, 하나 또는 그 이상의 센서를 통해 센서를 지나는 네

1) 본 연구는 미래부가 지원한 2013년 정보통신·방송(ICT) 연구개발사업의 연구결과로 수행되었음

2) This research was supported by the KUSTAR-KAIST Institute, Korea, under the R&D program supervised by the KAIST.

트워크 패킷을 측정하고, 이를 이용해 네트워크 내의 연결을 분석하고 침입을 탐지할 수 있다. IDS는 센서의 설치 위치에 따라 네트워크 기반 IDS(N-IDS), 호스트 기반 IDS(H-IDS)로 나뉘며, 일반적으로 IDS는 통상 N-IDS를 의미한다. N-IDS는 센서가 네트워크 라우터 등  네트워크 전체를 감시할 수 있도록 설치된 것이며, H-IDS는 센서가 호스트(서버) 하나에만 설치된 것이다.

침입이 발견되었을 경우 IDS는 침입을 차단하거나 네트워크 관리자에게 신속히 보고를 하여야 하며, IDS는 각종 공격에 대하여 높은 정확도로 네트워크 내에 침입이 일어나고 있는지 아닌지를 판별할 수 있어야 한다.

이러한 IDS의 성능 판단 척도로 탐지율(Detection Rate, DR)과 오탐율(False Alarm Rate, FAR)이 있다. DR은 (탐지한 공격 건수/전체공격 건수)로 나타내며 DR이 낮을 경우 공격이 시도되고 있음에도 인식하지 못한다는 뜻으로 IDS에서 가장 중요한 척도이다. FAR은 ((공격을 정상으로 판단한 건수+ 정상을 공격으로 판단한 건수) / 공격으로 판단한 전체  건수))로 나타내며, FAR이 높을 경우 비정상적인 연결이 접속될 확률이 높아져 네트워크가 위협에 빠질 수가 된다.

II.2 KDD'99 Dataset

KDD CUP'99는 Darpa 계획의 일환으로 네트워크 상에서 정상적인 연결과 비정상적인 연결을 가지고 있는  트래픽을 가지고 있었으며, 각종 IDS에 관한 객관적인 평가에 사용할 Dataset으로 제공된 것이 KDD99 Dataset이다[El00]. 이 Dataset은 1800만개의 패킷 헤더를 가지고 있으며 크게 나누어 4가지 공격 방식인 Probe, DoS, U2R, R2L과 일반 패킷인 Normal 중에 하나로 되어 있어 분류된 패킷의 의미는 다음과 같다.

- 일반 : 공격이 아닌 정상적인 패킷

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

- DoS : Denial of Service. 서비스를 이용할 수 없도록 과도한 요청을 하는 패킷

= U2R : User to Root. 권한 없는 사용자가 관리자(root) 권한을 얻으려 시도하는 패킷

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

각 패킷에 대한 분포는 [표 1]과 같다.

[표 1] KDD'99 Dataset의 분포

| 패킷 | 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% |

II.3 Swarm Intelligence

Swarm Intelligence(군집 지능)은 생체모방 인공 지능의 일종으로 기계학습기술이나 로봇 시스템을 개발하는데 쓰인다. SI는 복잡한 문제를 해결하기 위하여 개별적으로는 지능이 낮은 에이전트가 다수가 모여 효과적인 해결 방법을 자연 현상에서 찾아 활용하는 것이다[FM08]. SI를 사용할 경우 자연 현상을 이해하고 그에 따른 모델을 수립하는 것에 어려움이 있으나, 적절히 사용했을 경우 유사 해를 매우 빠르게 구할 수 있는 장점이 존재한다.

## III. 각종 SI 기법 소개

III.1 Ant Colony Optimization

Ant Colony Optimization(ACO)은 개미가 길을 찾을 때 페로몬을 이용하는 것에 착안해 만들어진 최적화 이론이다[4]. 시작점과 도착점 사이에 여러 경로가 있을 때, 처음엔 무작위로 움직이면서 페로몬을 뿌리면 가까운 경로가 좀 더 많은 페로몬이 남으므로 적

은 페로몬이 남은 경로를 제거하면 최적 경로를 찾을 수 있다.

III.2 ACO 기반 IDS

ACO는 공격자의 라우터 경로에 따라 페로몬을 증가시키는 방법으로 공격자의 위치 추적(Traceback)하는 기술을 사용할 수 있어, 공격 탐지와 더불어 공격 차단까지 가능하다. 그러나 ACO는 침입 탐지를 하는 데 단독으로 사용되기는 어려우나 기계 학습 중 결정 트리(Decision Tree)와 그의 응용 과정을 사용하는데 필요한 분류 규칙(Classification Rule)을 결정하는데 도움이 될 수 있다. ACO는 매우 많은 분류 규칙 후보 중에서 혼선을 줄 수 있는 후보를 없애고, 중요한 후보만을 남기는 것으로 탐지율을 높일 수 있다.

III.3 Particle Swarm Optimization

Particle Swarm Optimization(PSO)은 새나 벌과 같이 떼를 지어 이동하는 것을 착안해 만들어진 최적화 이론이다[FM08]. 무리 중 한 개체를 Particle로 표현하며, 각 Particle은 잠재해가 된다. 각 Particle은 위치와 속도를 가지고 각 변수를 차원으로 표현한 다차원 공간에서 움직이며 더 적절한 위치로 이동한다. 이때 각 입자가 발견한 최적의 위치(pbest)와 모든 입자들이 발견한 최적의 위치(gbest)에 가까워지면 속도가 변하는 운동을 통해서 유사해를 찾아나갈 수 있다.

III.4 PSO 기반 IDS

PSO는 Attribute의 수가 많을 때 최적해 또는 유사해를 구하는 문제에 적용 할 수 있다. 기계 학습의 경우 Attribute의 수가 많으며 학습에 필요한 Data의 수가 많은데, PSO를 접목하면 더욱 높은 성능을 갖게 된다. PSO는 인공 신경망 네트워크에서 뉴런들의 거리와 구성을 상태를 결정하는데 적용하여 이를 이용한 IDS를 만들 수 있으며, SVM(Support Vector Machine)에서도 각종 변수를 PSO를 통하여 결정하는 형식으로도 IDS를 구성할 수 있다. 이외에도 분류 규칙 결정 등에도 PSO를 적용할 수 있다.

III.5 Ant Colony Clustering

Ant Colony Clustering(ACC)은 개미굴에서 개미가 외부의 명령 없이도 식량, 알 등을 자동으로 분류하는 것에 착안하여 만들어진 이론이다. 개미는 물건이 적게 놓여있으며 비교적 다양한 종류가 놓여있는 곳에서 물건을 집어서 움직이다가 물건이 많이 놓여 있으면서 개미가 가지고 있는 물건과 비슷한 곳일 경우 놓는 작업을 반복한다. 이 작업을 통해 서로 흩어져 있다 할지라도, 점점 같은 물건이 밀집하는 결과가 나타난다.

III.6 ACC 기반 IDS

ACC는 클러스터링 이나 분류하는데 최적화 되어있다. 자체 구성 지도(Self Organizing Map)를 만드는데 필요한 분류 작업에서 ACC를 사용하여 IDS를 구성할 수 있으며, SVM을 이용한 IDS를 만들 때 ACC를 이용한 클러스터링 기법을 조합하여 정확도를 높이는 방법도 사용할 수 있다.

## IV. 방식별 비교 분석

각종 SI 기법을 적용한 IDS 기법들의 성능을 비교하기 위해 객관성 있고 널리 사용되는 KDD99 Dataset의 10%가량을 추출하여, [표 2]과 같이 비교하였다. 각 항목별 실제 공격 횟수를 확인하기 위해 분류된 [표 1]을 같은 패킷을 이용하였다.

PSO는 신경망과 SVM 등 다양한 방법으로 혼용되었으며, 적용된 방식에 따라서도 DR과 FAR이 매우 다양하게 나오고 있어, 어떤 기법과 조합하느냐 뿐만 아니라 조합 방법에 따라 IDS의 성능에 큰 영향을 끼치는 것으로 보인다.

ACO를 적용한 경우 적용 가능한 기법이 한정되어 있고, 그 경우도 다른 기법에 비해

[표 2] SI 기반 IDS의 성능 비교 표

| Author | ML | SI | 일반 | Probe | DoS | U2R | R2L | DR | FAR |
|---|---|---|---|---|---|---|---|---|---|
| KDD99 Winner[El00] | | | 99.5 | 83.3 | 97.1 | 13.2 | 8.4 | 90.9 | - |
| [CQ09] | 신경망 기반 | PSO | - | 88.86 | 92.57 | 91.14 | 94.29 | - | - |
| [Re87] | | | 96.88 | 92.20 | 97.74 | 52.86 | 8.30 | - | 0.61 |
| [MLLW08] | | | - | - | - | - | - | 96.77 | 8.01 |
| [MLL07] | | | - | - | - | - | - | 97.3 | 4.89 |
| [LJL10] | SVM 기반 | PSO | - | 86.48 | 88.48 | 85.52 | 84.53 | - | - |
| [ZLL09] | | | - | - | - | - | - | 97.26 | - |
| [WHRL09 | | | - | - | - | - | - | 99.84 | - |
| [CW09] | 분류 규칙 | PSO | - | - | - | - | - | 92.2 | 3.97 |
| [AKEN08] | | ACO | 98.5 | 82.5 | 98.5 | 76.3 | 89 | 95.5 | 0.0018 |
| [AH10] | | | 96 | 86.25 | 98.83 | 72.8 | 33.45 | 94.33 | - |
| [RA05] | 해당 없음 | ACC | 99.64 | 98.29 | 99.98 | 64 | 99.47 | - | - |
| [TK05a] | | | 98.5 | 86.9 | 97.5 | 27.2 | 11.0 | 92.25 | 1.5 |
| [TK05b] | | | 98.8 | 87.5 | 97.3 | 30.7 | 12.6 | - | - |
| [FZYW06] | | | 99.1 | 97.18 | 99.35 | 63 | 97.79 | - | - |

크게 뛰어난 결과를 보여주고 있지는 못하다.

ACC를 이용한 경우에는 U2R과 R2L의 탐지율이 낮게 나오고 있으나, 이는 클러스터링을 이용한 방법의 특성상 항목의 수가 적을 경우 제대로 구분하지 못하기 때문이며, 가장 많은 공격 방식인 DoS의 경우 ACC를 이용한 IDS 대부분이 매우 뛰어난 DR을 보이고 있다.

## V. 결론 및 향후 연구

본 논문에서는 SI 기법인 ACO, PSO, ACC에 대해서 소개하고 IDS에 적용하는 방법에 대해서 설명하였다. 또한 실제 구현된 IDS의 성능을 비교하였다.

ACO를 이용한 기법은 이론적인 한계 값이 존재하는 문제가 있다. PSO는 다양한 기계 학습 기법들과 조합되어 사용될 수 있으나 DR을 올리려면 매우 많은 기법과 조합되어야만 한다. ACC는 많은 공격이 올수록 더욱 탐지율이 상승하는 장점이 있고, 클러스터링을 이용한 기법의 특징으로 사전 정보가 주어지지 않아도 구분이 가능한 점이 있기 때문에 미지의 공격을 대비하는 데에도 적절한 기법이라 볼 수 있다.

본 연구는 KDD99 Dataset을 기반으로 각종 성능 비교를 하였으나 패켓 샘플링 한 시점이 1999년이므로 일반 패킷의 종류와 공격 기법 등에서 현 시점의 네트워크 환경과 많이 다르기 때문에, 각종 최근의 침입 기법을 탐지에 효과적인 지는 아직 미지수이다. 따라서 KDD99 Dataset 이후 공개된 IDS Dataset인 UNB ISCX Dataset [SSTG11]를 이용하여 IDS의 성능을 비교하면 더 유용한 결과를 얻을 수 있을 것이다. 또한, 위의 세 방법 외에 인공 면역 시스템(Artificial Immune System)과 같은 다른 SI 기법을 도입한 IDS에 대해서도 조사할 필요성이 있다[FM08].

## [참고문헌]


[AH10] Mohammad Saniee Abadeh, Jafar Habibi, and Emad Soroush, "Induction of Fuzzy Classification systems via evolutionary ACO-based algorithms", Computer, Vol. 35, 2008.

[AKEN08] H. Alipour, et al., "ACO-FCR: Applying ACO-based algorithms to induct FCR", Proceedings of the World Congress on Engineering. Vol. 1. 2008.

[CQ09] Zhifeng Chen and Peide Qian, "Application of PSO-RBF neural network in network intrusion detection" Intelligent Information Technology Application, 2009. IITA 2009. Third International Symposium on. Vol. 1. IEEE, 2009.

[CW09] Zhao Chang, and Wang Wei-ping, "An improved PSO-based rule extraction algorithm for intrusion detection", Computational Intelligence and Natural Computing, 2009. CINC'09. International Conference on. Vol. 2. IEEE, 2009.

[El00] Charles Elkan, "Results of the KDD'99 classifier learning". ACM SIGKDD Explorations Newsletter 1.2 , 2000, 63-64.

[FM08] Dario Floreano and Claudio Mattiussi, "Bio-inspired artificial intelligence: theories, methods, and technologies", The MIT Press, 2008.

[KGM11] Constantinos Kolias, Kambourakis Georgios and M. Maragoudakis, "Swarm intelligence in intrusion detection: a survey." Computers & Security, Vol.30.8, 2011, 625-642.

[LJL10] Huaping Liu, Yin Jian, and Sijia Liu, "A new intelligent intrusion detection method based on attribute reduction and parameters optimization of SVM", Education Technology and Computer Science (ETCS), 2010 Second International Workshop, Vol. 1. IEEE.

[MLL07] Ruhui Ma, Yuan Liu, and Xing Lin, "Hybrid QPSO based wavelet neural networks for network anomaly detection", Digital Media and its Application in Museum & Heritages, Second Workshop, IEEE, 2007.

[MLLW08] Ruihui Ma, et al., "Network anomaly detection using RBF neural network with hybrid QPSO", Networking, Sensing and Control, 2008. ICNSC 2008. IEEE International Conference, IEEE, 2008.

[RA05] Vitorino Ramos and Ajith Abraham, "ANTIDS: Self Orga nized Ant-Based Clustering Model for Intrusion Detection System", Soft Computing as Transdisciplinary Science and Technology,Springer, 2005, 977-986.

[Re87] Craig W. Reynolds, "Flocks, herds and schools: A distributed behavioral model", ACM SIGGRAPH Computer Graphics. Vol. 21, No. 4, ACM, 1987.

[SM07] Karen Scarfone and Peter Mell, "Guide to intrusion detection and prevention systems", NIST Special Publication 800, 2007.

[SSTG11] Ali Shiravi, et al., "Toward developing a systematic approach to generate benchmark datasets for intrusion detection", Computers & Security, Vol.31.3, 2012, 357-374.

[TK05a] Wilson Tsang and Sam Kwong, "Unsupervised anomaly intrusion detection using ant colony clustering model", Soft Computing as Transdisciplinary Science and Technology, Springer, 2005. 223-232.

[TK05b] Chi-Ho Tsang and Sam Kwong, "Multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised feature extraction", Industrial Technology, 2005. ICIT 2005. IEEE International Conference, 2005.

[WHRL09] Jun Wang, et al., "A real-time intrusion detection system based on PSO-SVM", International Workshop on Information Security and Application, 2009.

[ZLL09] Tie-Jun Zhou, Yang Li, and Jia Li, "Research on intrusion detection of SVM based on PSO", Machine Learning and Cybernetics, 2009 International Conference on. Vol. 2. IEEE.