Deep Learning in Intrusion Detection Perspective: Overview and Further Challenges

Kwangjo Kim
School of Computing
KAIST
Daejeon, South Korea
Email: kkj@kaist.ac.kr

Muhamad Erza Aminanto
School of Computing
KAIST
Daejeon, South Korea
Email: aminanto@kaist.ac.kr

***Abstract*—Deep learning techniques are famous due to its capability to cope with large-scale data these days. They have been investigated within various of applications *e.g.*, language, graphical modeling, speech, audio, image recognition, video, natural language and signal processing areas. In addition, extensive researches applying machine-learning methods in Intrusion Detection System (IDS) have been done in both academia and industry. However, huge data and difficulties to obtain data instances are hot challenges to machine-learning-based IDS. We show some limitations of previous IDSs which uses classic machine learners and introduce feature learning including feature construction, extraction and selection to overcome the challenges. We discuss some distinguished deep learning techniques and its application for IDS purposes. Future research directions using deep learning techniques for IDS purposes are briefly summarized.**

***Keywords*—Intrusion detection system, Wi-Fi network, feature selection, artificial neural network, decision tree.**

I. Preliminaries

Cyber-attacks have becoming an immense growing rate as Internet of Things (IoT) are widely used these days [1]. More than half of Internet traffics are anticipated coming from wireless network traffics which are more prone to be exploited by adversaries [2]. IBM [3] reported an enormous account hijacked during 2016 and spam mails are four times higher than previous year. Common attacks noticed in the same report including bruteforce, malvertising, phishing, SQL injection, DDoS, malware, *etc*. Majority of malwares are accounted as a ransomware (85% of malwares existed in a year are a ransomware). These attacks might leak sensitive data or disrupt normal operations which leads to an enormous financial loss. The most prevailing companies impacted by security incidents are financial services-related companies. Followed by information and communications, manufacture, retail and healthcare [3]. This situation forces us to strengthen our security measures in our system.

Intrusion Detection System (IDS) becomes a standard security measure in computer networks. Unlike firewall, IDS usually located inside the network to monitor all internal traffics. One may consider to have both a firewall and IDS to protect the network. IDS is defined as an automation of intrusion detection process which is a process of finding events of violation of security policies or standard security practices in computer networks [4]. Besides identifying the security incidents, IDS also has other functions: documenting existing threats and deterring adversaries [4]. IDS requires particular properties which acts as a passive countermeasure, monitors whole or part of networks only and aims high attack detection rate and low false alarm rate.

We can divide IDSs based on their placement in the network and methodology used. By the positioning of the IDS module in the network, we might distinguish IDSs to 3 classes: network-based, host-based and hybrid-based IDSs. The first IDS, network-based IDS, puts the IDS module in the network which can monitor whole the network traffics. This IDS has a big picture of the network makes it has a better understanding the network in overall. On the other hand, the host-based IDS places the IDS module on each client of the network. The module can only see the ingoing or outgoing traffics of the corresponding client leads to detail monitoring of the particular client. Two types of IDSs have specific drawbacks– the network-based IDS might burden of the workload then misses some malicious activities, while the host-based IDS does not have the overview of the whole network but having less workload than the network-based IDS. Therefore, the hybrid-based IDS places IDS modules in the network as well as clients to monitor both specific clients and network overview at the same time.

In the latter case, based on the detection method, IDSs can be divided into 3 different types: misuse, anomaly, and specification-based IDSs. A misuse-based IDS, known as a signature-based IDS [5], looks for any malicious activities by matching the known signatures or patterns of attacks with the monitored traffics. This IDS suits known attack detection; however, new or unknown attacks (also called as a zero-day exploit) are difficult to be detected. An anomaly-based IDS detects an attack by profiling normal behavior and then triggers an alarm if there is any deviation from it. The strength of this IDS is its ability for unknown attack detection. However, misuse-based IDS usually achieves higher detection performance for known attacks than anomaly-based IDS. A specification-based IDS manually defines a set of rules and constraints to express the normal operations. Any deviation from the rules and constraints during execution is flagged as malicious [6]. Table I summarizes the comparison of IDS types

TABLE I
COMPARISON OF IDS TYPES BASED ON THE METHODOLOGY

	Misuse-based	Anomaly-based	Specification-based
Method	Identify known attack patterns	Identify unusual activity patterns	Identify violation of pre-defined rules
Detection Rate	High	Low	High
False Alarm Rate	Low	High	Low
Unknown Attack Detection	Incapable	Capable	Incapable
Drawback	Updating signatures is burdensome	Computing any machine learning is heavy	Relying on expert knowledge during defining rules is undesirable

TABLE II
COMPARISON BETWEEN SUPERVISED AND UNSUPERVISED LEARNING

	Supervised	Unsupervised
Definition	The dataset are labeled with pre-defined classes	The dataset are labeled **without** pre-defined classes
Method	Classification	Clustering
Example	Support Vector Machine (SVM), Decision Tree (DT)	K-means clustering, Ant Clustering Algorithm (ACA)
Known Attack Detection	High	Low
Unknown Attack Detection	Low	High

based on the methodology.

An IDS that leverages machine-learning method is an example of an anomaly-based IDS [7]. There are two types of learning namely supervised and unsupervised learning. The unsupervised learning does not require a labeled dataset for training which is crucial for huge network traffics recently, while the supervised learning requires a labeled dataset. Unsupervised learning capability is of critical significance as it allows a model to be built to detect new attacks without creating costly labels or dependent variables. Table II outlines the comparison between supervised and unsupervised learning.

II. MACHINE LEARNING-BASED IDS OVERVIEW

A combination of two typical methods are commonly used to build an IDS such as learning or training and classification as shown in Fig. 1. It is difficult and costly to obtain bulk of labeled network connection records for supervised training in the first stage. Then feature learning or clustering might become the solution in the first place. The clustering analysis has emerged as an anomaly detection recently [8]. Clustering

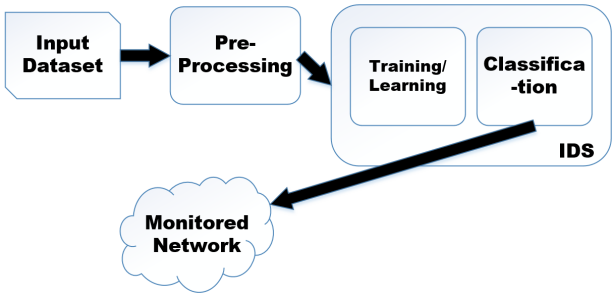

Fig. 1. IDS Typical Scheme

TABLE III
COMMON IDSS WITH A COMBINATION OF LEARNING AND CLASSIFICATION

Publication	Learning	Classification
AKKK17 [12]	ACA	FIS
HKY14 [13]	ATTA-C	ATTA-C + label
KKK15 [14]	ACA	AIS
KHKY16 [15]	ACA	DT, ANN

is an unsupervised data exploratory technique that partitions a set of unlabeled data patterns into groups or clusters such that patterns within a cluster are similar to each other but dissimilar to other clusters' pattern [8]. Meanwhile, feature learning is a tool for improving the learning process of a machine-learning algorithm. It commonly consists of feature construction, extraction, and selection. Feature construction expands the original features to enhance their expressiveness, whereas feature extraction transforms the original features into a new form and feature selection eliminates unnecessary features [9]. The classification task is a supervised method to distinguish benign and malicious traffics based on provided data which usually comes from previous step as shown in Fig. 1.

We can see in the Fig. 1 that the pre-processing step is required before entering the IDS module. The pre-processing module commonly consists of normalization and balancing steps. Data normalization is a process to output all value ranges of each attribute are equal, which is important for proper learning by any machine learning algorithm [10]. Meanwhile, the nature of real-world network is having benign traffics much larger than malicious traffics. This properties could make it difficult for the IDS module to learn the underlying patterns correctly [11]. Therefore, a balancing process which creates the dataset with equal ratio for both benign and malicious instances, is a required step for a training. However, we should use original ratio, which is unbalanced, for testing purposes to validate the IDS can be implemented in the real-world networks.

As mentioned, we explored several common IDSs with a combination of learning and classification as shown in Table III.

Ant Clustering Algorithm (ACA) is one of the most widely used clustering approaches which is originated from swarm intelligence. ACA is an unsupervised learning algorithm that is

able to find near-optimal clustering solution without predefined number of clusters needed [8]. However, ACA is rarely used in intrusion detection as the exclusive method for classification. Instead, ACA is combined with other supervised algorithms such as Self Organizing Map (SOM) and Support Vector Machine (SVM) in order to provide better classification result [16]. In AKKK17 [12], we proposed a novel hybrid IDS scheme based on ACA and Fuzzy Inference System (FIS). We applied ACA for training phase and FIS for classification phase. We chose FIS as classification phase, because fuzzy approach can reduce the false alarm with higher reliability in determining intrusion activities [17]. Meanwhile, we also examined the same ACA with different classifiers in KKK15 [14] and KHKY16 [15] by using Artificial Immune System (AIS) and Decision Tree (DT) as well as Artificial Neural Network (ANN), respectively. AIS is designed for the computational system and inspired by Human Inference System (HIS). AIS has the capability to differentiate between the "self" (cells that are owned by the system) and "non-self" (foreign entities to the system). We show that ANN can learn more complex structure of certain unknown-attacks due to a characteristic of ANN. In addition, we also investigated an improved ACA which is Adaptive Time Dependent Transporter Ants Clustering (ATTA-C) in HKY14 [13], which is one of the few algorithms that have been benchmarked on various datasets, and is now publicly available under GNU agreement [13].

In addition to above-mentioned common IDSs, we further examined other IDS models taking benefits of Hadoop framework [18] and Software Defined Networking (SDN) environment [19]. In [18], we proposed a method utilizes the advantages of Hadoop as well as behavioral flow analysis. This framework is particularly useful in the case of P2P traffic analysis due to inherent flow characteristics of this type of applications. Meanwhile, we proposed a novel IDS scheme that operates lightweight intrusion detection that keeps a detailed analysis of attacks [19]. In this scheme, a flow-based IDS detects intrusions, but with low operating cost. When an attack is detected, the IDS requests the forwarding of attack traffic to packet-based detection so the detailed results obtained by packet-based detection can be analyzed later by security experts.

III. Deep Learning Overview

Deep learning originally comes from the advancements of Neural Network (NN) algorithm. Various methods have been applied in order to overcome the limitations of one hidden layer only in NN. Those methods employ consecutive hidden layers which are hierarchically cascaded. Due to vast of methods belong to deep learning, we classify several deep learning methods based on their approach [20]. Deng [21] differentiates deep learning into three sub-groups, generative, discriminative and hybrid. The classification is based on the intention of architectures and techniques, *e.g.*, synthesis/generation or recognition/classification. The classification of the deep learning methods is shown in Fig. 2.

A. Unsupervised Learning

Unsupervised learning or so called generative architectures, uses unlabeled data. The main concept of applying generative architectures to pattern recognition is unsupervised learning or pre-training [21]. Since learning the lower levels of subsequent networks are difficult, deep generative architectures are needed. Thus, with limited training data, learning each lower layer in layer-by-layer approach without relying on all the layers above is important.

There are number of methods that classified as unsupervised learning as follows:

1) Auto Encoder (AE)-Stacked Auto Encoder (SAE): AE is an ordinary Artificial Neural Network (ANN) with the same neuron number of both input and output layers. Meanwhile, the nodes in the hidden layer are representing new feature set which is low-dimensional. This architecture leads to an ability that can reconstruct the data after complicated computations. AE aims to learn a compact set of data efficiently and can be stacked to build a deep network. Training results of each hidden layer are cascaded. This structure is called Stacked Auto-Encoder (SAE) which can provide new transformed features by different depths. In order to train more precisely, we can append an additional layer with labels once we have large amount of tagged samples [22]. In addition, a Denoising Auto Encoder (DAE) is trained to reconstruct a clear correction input from a corrupted by noise input [23]. The DAE may be also stacked in order to build deep networks as well.

2) Boltzman Machine (BM): BM is a network of binary units that symmetrically paired [24]. BM has a structure of neuron units that makes stochastic decisions about whether active or not [25]. If one BM result is cascaded into multiple BMs, called Deep BM (DBM). Meanwhile, Restricted Boltzmann machine (RBM) is a customized BM without connections among the hidden units [24]. RBM consists of visible and hidden variables such that their relations can be figured out. If multiple layers are stacked, layer-by-layer scheme, called as Deep Belief Network (DBN). DBN could be used as a feature extraction method for dimensionality reduction when unlabeled dataset and back-propagation are used (which means unsupervised training). In contrast, DBN is used for classification when appropriate labeled dataset with feature vectors are used (which means supervised training) [26].

B. Supervised Learning

Supervised learning or discriminative deep architecture is intended to distinguish some parts of data for pattern classification [21]. An example of the discriminative architecture is Convolutional Neural Network (CNN) which employs a special architecture particularly suitable for image recognition. The advantage of CNN is fast to train because of its structure. CNN can train multilayer networks with gradient descent to learn complex, high-dimensional, nonlinear mappings from large collections of data [27]. CNN uses three basic concepts: local receptive fields, shared weights, and pooling [28]. One

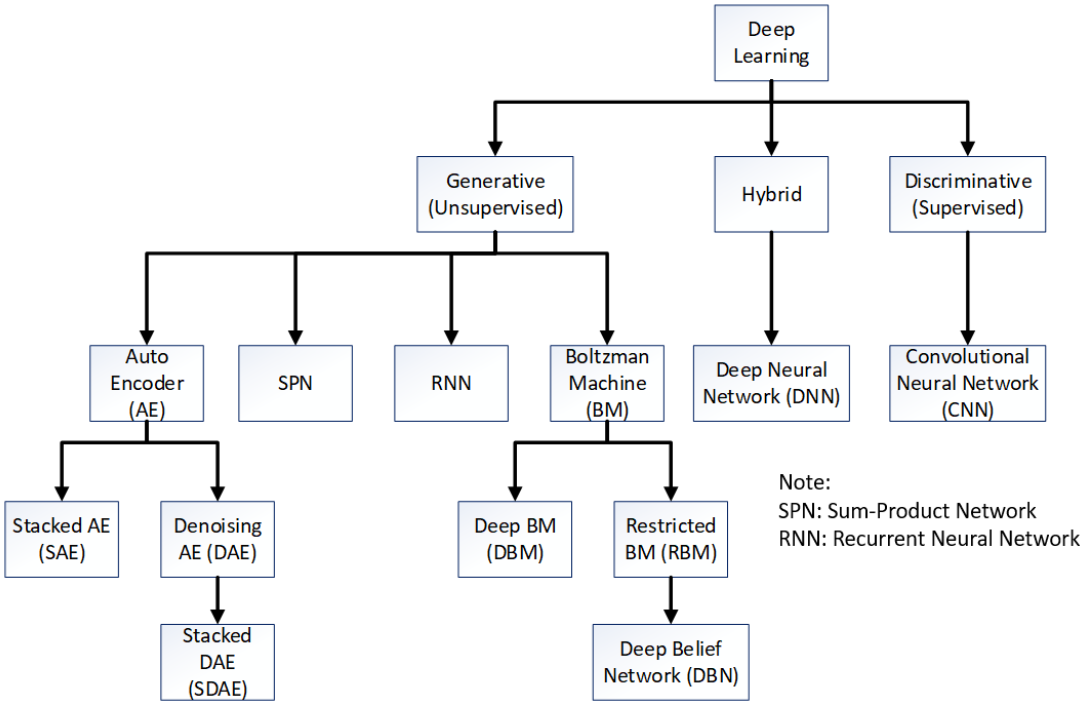

Fig. 2. Classification of Deep Learning Methods [20]

extensive research that successfully deployed using CNN is AlphaGo by Google [29].

C. Hybrid

Hybrid deep architecture combines both generative and discriminative architectures. The hybrid architecture aims to distinguish data as well as discriminative approach. However, in the early step, it has assisted in a significant way with the generative architectures results. An example of hybrid architecture is Deep Neural Network (DNN). However, some confusion terms between DNN and DBN happens. In the open literatures, DBN also uses back propagation discriminative training as a "fine-tuning." This concept of DBN is really similar to Deep Neural Network (DNN) [21]. According to Deng [25], DNN is defined as a multilayer network with cascaded fully connected hidden layers, and is often use stacked RBM as a pre-training phase.

IV. SAE Implementations in IDS

The goal of deep learning method is learning feature hierarchies from lower level to higher level features [30]. The method can learn features independently at multiple levels of abstraction, and thus discover complicated functions mapping between the input to the output directly from raw data without depending on customized features by the experts. In higher-level abstractions, humans often have no idea to see the relation and connection from raw sensory input. Therefore, the ability to learn complex features, also called as feature extraction, will become necessarily needed as the amount of data increased sharply [30]. SAE is one good instance of feature extractors.

TABLE IV
IDSs Leveraging SAE

Publication	Role of SAE	Combined with
AK16a [31]	Classifier	ANN
AK16b [32]	Feature Extractor	Softmax Regression
AK17 [33]	Clustering	K-means Clustering
ACTYK17 [34]	Feature Extractor	SVM, DT, ANN

Therefore, we discuss several previous work which implement SAE as the feature extractor and other roles as well, in the IDS module as shown in Table IV.

Feature extraction by SAE is able to reduce the complexity of original features of the dataset. However, besided as a feature extractor, we validated that SAE can also used for classifying and clustering tasks as shown in Table IV. In AK16b [32], we used semi-supervised approach for our IDS which contains feature extractor (unsupervised learning) and classifier (supervised learning). We leveraged SAE for feature extraction, and regression layer with softmax activation function for classifier. We implemented SAE as feature extractor as well in ACTYK17 [34], but we leveraged ANN, DT and SVM as a feature selection. In other words, we combines stacked feature extraction and weighted feature selections. By our experiments [34], we improved our feature learning process by combining stacked feature extraction with weighted feature selection. The feature extraction of SAE is capable of transforming the original features into a more meaningful representation by reconstructing its input and providing a way to check that the relevant information in the data has

TABLE V
COMPARISON ON IMPERSONATION DETECTION

Method	Detection Rate (%)	False Alarm Rate (%)
AK16a [31]	65.178	0.143
AK16b [32]	92.674	2.500
AK17 [33]	92.180	4.400
ACTYK17 [34]	99.918	0.012
KKSG15 [35]	22.008	0.021

been captured. SAE can be efficiently used for unsupervised learning on a complex dataset.

Unlike two previous approaches, we use SAE for other roles than a feature extractor, namely classifying and clustering methods in AK16a [31] and AK17 [33], respectively. We adopted ANN as a feature selection since the weight from trained models mimics the significance of the correspondence input [31]. By selecting the important features only, the training process becomes lighter and faster than before. In AK16a [31], we exploited SAE as a classifier, since this employs consecutive layers of processing stages in hierarchical manners for pattern classification and feature or representation learning. On the other hand, we proposed a novel fully unsupervised method [33] which can detect attacks without prior information on data label. Our method is equipped by an unsupervised SAE for extracting features and a K-means clustering algorithm for clustering task.

In order to compare those approaches, we validated those approaches using Aegean Wi-Fi Intrusion Dataset (AWID) which is a Wi-Fi network benchmark dataset built by Kolias *et al.* [35]. AWID consists of 4 classes: benign, impersonation, injection and flooding classes and provides training and test datasets. There are 1,795,575 instances in the training dataset with 1,633,190 and 162,385 benign and attack instances, respectively. While the test dataset contains 575,643 instances with 530,785 and 44,858 benign and attack instances, respectively. Kolias *et al.* [35] tested a number of existing machine learning models on the dataset in a heuristic manner. The lowest detection rate is observed particularly on impersonation attack reaching an accuracy of 22% only. Therefore, we focus to improve impersonation detection and hence compare our approaches on impersonation detection as summarized in Table V. Detection Rate (DR) refers to the number of attacks detected divided by the total number of attack instances in the test dataset while False Alarm Rate (FAR) is the number of normal instances classified as an attack divided by the total number of normal instances in the test dataset.

From Table V, we can observe that SAE is able to improve the performance of our IDS compared to KKSG15 [35]. We verified that SAE achieved high level abstraction of complex and huge Wi-Fi network data. The SAE's model free properties and learnability on complex and large scale data fit into the open nature of Wi-Fi networks. Among all IDSs, the one using SAE as a classifier achieved the lowest impersonation attack detection rate with 65.178% only. It shows that SAE is able to be a classifier but not excellent as the original role of SAE is a feature extractor. The usability of SAE as a feature extractor validated by AK16b [32] and ACTYK17 [34] which achieved highest DR. Even more, by a combination of SAE extractor and weighted selection [34], we achieved the best performance of DR and FAR among other. Besides that, we found an interesting fact that SAE can assist K-means clustering algorithm to achieve better performance with DR of 92.180% [33]. However, we need to analyze further to reduce the FAR since it achieved the highest FAR which is undesirable in IDS.

V. DISCUSSION AND FURTHER CHALLENGES

A. Discussion

We investigated various algorithms especially bio-inspired algorithms to bring significance in the field of IDS research. We believe that by adopting what nature does, we can improve current methods. We started with observing ant behavior and adopting Ant Clustering Algorithm as our clustering algorithm as shown in Table III. However, we need other methods for improving the performance of our IDSs. We believe that ACA is still limited to distinguish between benign and attack instances. Therefore, we shifted to more recent bio-inspired algorithms, deep learning, which is the advance of neural network. Incorporating deep learning methods as a real-time classifier will be a challenging task. Majority of previous work that leveraging deep learning methods in their IDS environment, they perform the feature extraction or reducing feature dimensionalities only. However, we show that deep learning methods are able to do clustering task as well.

In summary, we can conclude that SAE is very useful for following tasks:

- Feature extraction
- Clustering
- Classification

B. Further Challenges

Further challenges are left for improving IDS in the future. Based on our previous work, we recommend the followings for future directions in IDS researches.

- Deep learning have significantly improved IDSs. However, we should make it lighter to be suitable for IoT environments such as CAN (Controller Area Network) used by Unmanned Vehicle.
- Improving unsupervised approach since huge labeled data are difficult to obtained. Therefore an IDS leveraging unsupervised approach is desirable.
- Build an IDS that is able to detect zero-day attacks with high detection rate and low false alarm rate.
- A comprehensive measure not only detection but also prevention is needed in the future. Therefore, building an IDS with both detection and prevention capabilities (*e.g.* Intrusion Prevention System (IPS)) is expected.

Acknowledgements

This work was partly supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (2013-0-00396, Research on Communication Technology using Bio-Inspired Algorithm and 2017-0-00555, Towards Provable-secure Multi-party Authenticated Key Exchange Protocol based on Lattices in a Quantum World).

References

[1] C. Kolias, A. Stavrou, J. Voas, I. Bojanova, and R. Kuhn, "Learning internet-of-things security: Hands-on," *IEEE Security & Privacy*, vol. 14, no. 1, pp. 37–46, IEEE, 2016.

[2] CISCO, "Cisco Visual Networking Index: Forecast and Methodology, 2015-2020," http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-net-working-index-vni/complete-white-paper-c11-481360.html, 2016, accessed December 6, 2016.

[3] M. Alvarez, N. Bradley, P. Cobb, S. Craig, R. Iffert, L. Kessem, J. Kravitz, D. McMilen, and S. Moore, "IBM X-force threat intelligence index 2017," *IBM Corporation*, pp. 1–30, 2017.

[4] K. Scarfone and P. Mell, "Guide to intrusion detection and prevention systems (idps)," *NIST special publication*, vol. 800, no. 2007, 2007.

[5] A. H. Farooqi and F. A. Khan, "Intrusion detection systems for wireless sensor networks: A survey," *Communication and networking*, pp. 234–241, 2009.

[6] R. Mitchell and I. R. Chen, "Behavior rule specification-based intrusion detection for safety critical medical cyber physical systems," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 1, pp. 16–30, Jan 2015.

[7] I. Butun, S. D. Morgera, and R. Sankar, "A survey of intrusion detection systems in wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 266–282, 2014.

[8] C.-H. Tsang and S. Kwong, "Ant colony clustering and feature extraction for anomaly intrusion detection," *Swarm Intelligence in Data Mining*, pp. 101–123, 2006.

[9] H. Motoda and H. Liu, "Feature selection, extraction and construction," *Communication of IICM (Institute of Information and Computing Machinery, Taiwan) Vol*, vol. 5, pp. 67–72, 2002.

[10] H. Bostani and M. Sheikhan, "Modification of supervised OPF-based intrusion detection systems using unsupervised learning and social network concept," *Pattern Recognition*, vol. 62, pp. 56–72, 2017.

[11] M. Sabhnani and G. Serpen, "Application of machine learning algorithms to KDD intrusion detection dataset within misuse detection context," *Proceedings of the International Conference on Machine Learning: Models, Technologies, and Applications*, pp. 209–215, 2003.

[12] M. E. Aminanto, H. Kim, K. M. Kim, and K. Kim, "Another fuzzy anomaly detection system based on ant clustering algorithm," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 100, no. 1, pp. 176–183, 2017.

[13] K. Huseynov, K. Kim, and P. Yoo, "Semi-supervised botnet detection using ant colony clustering," *The 31th Symposium on Cryptography and Information Security (SCIS)*, 2014.

[14] K. M. Kim, H. Kim, and K. Kim, "Design of an intrusion detection system for unknown-attacks based on bio-inspired algorithms," *Computer Security Symposium (CSS)*, vol. 2015, no. 3, pp. 64–70, 2015.

[15] K. M. Kim, J. Hong, K. Kim, and P. Yoo, "Evaluation of aca-based intrusion detection systems for unknown-attacks," *The 33th Symposium on Cryptography and Information Security (SCIS)*, 2016.

[16] C. Kolias, G. Kambourakis, and M. Maragoudakis, "Swarm intelligence in intrusion detection: A survey," *Computers & Security*, vol. 30, no. 8, pp. 625–642, Elsevier, 2011.

[17] A. Karami and M. Guerrero-Zapata, "A fuzzy anomaly detection system based on hybrid pso-kmeans algorithm in content-centric networks," *Neurocomputing*, vol. 149, pp. 1253–1269, 2015.

[18] K. Huseynov, P. D. Yoo, and K. Kim, "Scalable p2p botnet detection with threshold setting in hadoop framework," *Journal of the Korea Institute of Information Security and Cryptology*, vol. 25, no. 4, pp. 807–816, 2015.

[19] D. S. Lee and K. Kim, "Improving detection capability of flow-based ids in sdn," *KAIST, Department of Computer Science, Thesis Book*, 2015.

[20] M. E. Aminanto and K. Kim, "Deep learning in intrusion detection system: An overview," *International Research Conference on Engineering and Technology 2016*, 2016.

[21] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2014.

[22] Z. Wang, "The applications of deep learning on traffic identification," *Blackhat USA 2015*, 2015.

[23] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[24] R. Salakhutdinov and G. Hinton, "Deep boltzmann machines," *Artificial Intelligence and Statistics*, pp. 448–455, 2009.

[25] L. Deng, D. Yu *et al.*, "Deep learning: methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[26] M. Salama, H. Eid, R. Ramadan, A. Darwish, and A. Hassanien, "Hybrid intelligent intrusion detection scheme," *Soft computing in industrial applications*, pp. 293–303, 2011.

[27] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[28] M. A. Nielsen, "Neural networks and deep learning," 2015.

[29] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[30] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[31] M. E. Aminanto and K. Kim, "Detecting impersonation attack in Wi-Fi networks using deep learning approach," *Information Security Applications: 17th International Workshop, WISA 2016*, 2016.

[32] M. E. Aminanto and K. Kim, "Detecting active attacks in Wi-Fi network by semi-supervised deep learning," *Conference on Information Security and Cryptography 2017 Winter*, 2016.

[33] M. E. Aminanto and K. Kim, "Improving detection of Wi-Fi impersonation by fully unsupervised deep learning," *Information Security Applications: 18th International Workshop, WISA 2017*, 2017.

[34] M. E. Aminanto, R. Y. Choi, H. C. Tanuwidjaja, P. D. Yoo, and K. Kim, "Deep abstraction and weighted feature selection for Wi-Fi impersonation detection," *In Submission to IEEE Transactions on Information Forensics & Security*, 2017.

[35] C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis, "Intrusion detection in 802.11 networks: empirical evaluation of threats and a public dataset," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 184–208, IEEE, 2015.