Data Fusion-Based Network Anomaly Detection towards Evidence Theory

Cong Thanh Bui
Software R&D Department
Hi-tech Telecommunication Center
Communications Command
congthanhttmt@gmail.com

Van Loi Cao
Faculty of Information Technology
Le Quy Don Technical University
loi.cao@lqdtu.edu.vn

Minh Hoang
Head Office
Institute of Science Technology and Innovation
hoangminh@most.gov.vn

Quang Uy Nguyen
Faculty of Information Technology
Le Quy Don Technical University
quanguyhn@lqdtu.edu.vn

***Abstract*—We propose a fusion model of deep neural networks and traditional algorithms for anomaly detection. The proposed model inherits the advantages of both these methods to create a robust anomaly detection algorithm. We employ the Dempster-Shafer theory (D-S) of Evidence, a very reliable and flexible data fusion technique, to form a fusion-based network anomaly detection (FuseNAD) by applying a basic probability assignment (BPA) function and modifying the D-S theory's rule. FuseNAD fuses four anomaly detection methods consisting of a deep learning technique, namely Shrink Auto-Encoder, and three traditional ones such as One-class Support Vector Machine (OCSVM), Kernel Density Estimation (KDE) and Local Outlier Factor (LOF). The experimental results show increases in detection rate and overall accuracy in comparison to the individuals on several public network anomaly detection datasets.**

***Index Terms*—Deep learning; Auto-Encoder; Anomaly Detection; D-S Theory; Data Fusion.**

I. Introduction

The rapid development of the computer network in all aspects of its infrastructure has put it under the pressure of the modern cyber attacks. Seeking solutions to identify and prevent cyberattacks is a crucial task in network security, which has attracted the research community for the last few decades. Recently, machine learning has been seen as the primary approach for anomaly detection in network security [1], [2]. Amongst machine learning approaches, semi-supervised learning-based anomaly detection methods are suitable for detecting network anomalies [3], [4]. The reason is that these methods can construct anomaly detection models from only one class of data, typical normal class, and any querying data points that do not fit the models are indicated as anomalies [5]. There are many well-known semi-supervised algorithms for anomaly detection including One-class Support Vector Machine (OCSVM) [6], Local Outlier Factor (LOF) [7] and Kernel Density Estimation (KDE). These algorithms have shown the excellent results on many anomaly detection domains. However, they suffer from the challenge in dealing with high-dimensional network data [3]. Fortunately, deep learning-based anomaly detection methods have been demonstrated as powerful algorithms for complex anomaly detection problems, and can complement to the weaknesses of the density/distance-based techniques. Auto-Encoders (AEs), deep neural networks that learn to reconstruct the input data at the output layer, are known as the state-of-the-art anomaly detection method [8], [9]. Several variances of AEs, such as Denoising AEs and Shrink AE, make deep learning-based approaches very popular in anomaly detection domain.

However, these single methods might not be fully adapted to complicated and changing network systems [10]: network intrusions can be launched by "real person", and may be more complex than other destructive actions [11]. Thus, all the single anomaly detection methods tend to be only useful for some intrusions, but might not be ideal for others. This may result in the inefficient performance of network anomaly detection algorithms, such as low accuracy, higher false alarm rate (FAR), and lower detection rate (DR). To mitigate this problem, we propose a model by fusing the strength from different methods.

When constructing a fusion model, three important issues should be considered [12]. The first issue is the level of fusion model, i.e. how to choose the base detectors, and the fusion mechanism [11], [13]. The second one relates to the choice of the level fusion and the working field. The last issue is the fusion algorithms. The studies [10], [13]–[16] showed that the D-S Theory is a potential method for establishing a fusion-based model for network anomaly detection as it owns a flexible feature. However, in the D-S theory, the definition of Frame of discernment (FoD) and setting up the Basic probability assignment (BPA) function are too complicated. Another limitation is that it is difficult to solve the problem of unequal performance amongst individual models [10], [16], [17]. In this paper, we attempt to overcome the above issues by adjusting the D-S theory for network anomaly detection. In other words, we introduce a new D-S theory's rule, called DSR_AD, and employ FoD and the BPA function to efficiently fuse individual anomaly detection methods together. We then use the mechanism to combine four individual classifiers, SAE [3], OCSVM [18], KDE [19], and LOF [20] to form a fusion-based network anomaly detection method, called FuseNAD. The main contributions of the paper are:

- We develop a new D-S theory' rules, called DSR_AD, for building a robust model from single anomaly detectors. This can overcome the limitations of the classical D-S

978-1-7281-5163-2/19/$31.00 ©2019 IEEE

theory. We also adjust other elements of the D-S theory when applying the rule for anomaly detection problem.
- We propose new fusion anomaly detection model (FuseNAD) by using DSR_AD to combine a deep learning-based algorithm and three classical algorithms.
- We conduct extensive experiments on the well-known publicly datasets to evaluate and testify FuseNAD.

The remain of the paper proceeds as follows. In Section II, we briefly review some relevant work. In Section III, we present a complementary background. We detail our proposed model, FuseNAD, in Section IV and discuss our experimental results in Section V. Finally, the conclusion is presented in Section VI, which also draws some future directions.

II. RELATED WORK

This part briefly presents recent machine learning approaches for network anomaly detection and some fusion-based models in this domain. These research can be divided into two main categories: classical methods and deep learning techniques.

In the first group, some well-know anomaly detection methods (ADs), such as OCSVM, KDE and LOF have been demonstrated as powerful methods. Support vector machines have been succeeded for anomaly detection since the 1990s. OCSVM is sometimes known as a novelty detection technique. Based on the structure of the positive training data, OCCSVM creates a boundary in a feature space to best present the region containing the positive class. The margin is concreted by a set of data points in the feature space, called support vectors. OCSVM was proposed by Schölkopf et al. [21]. The aim is to find a hyper-plane that maximizes the margin between the region containing normal data and the origin in the feature space. The remaining space, including the origin, is indicated as the anomalous region. The trade-off between maximizing the margin and minimizing the number of target vectors dropping into the anomalous region is controlled by an outlier fraction $\nu \in (0, 1)$.

MP Wand et al. [19] introduce Kernel density estimation (KDE) that is a probability density-based approach. KDE fits a large number of kernels over the data, one per data point. These kernels share a single co-variance called bandwidth. As applying KDE for anomaly detection domain, only normal data is used to build a probability density function. Every output is classified as an anomaly if its density produced from the function is lower than a predestined threshold. A Lazarevic et al. [20] proposed Local Outlier Factor (LOF) as a novelty detection method. The algorithm works based on the amount of density of each data point w.r.t its neighbors. A local outlier factor score indicates the anomaly degree of each data point. The larger score a data point has, the higher the probability the data point is classified as an anomaly.

Recently, Cao et al. [3] proposed Shrink AutoEncoder (SAE) for anomaly detection. Their model forces the hidden representation of normal data into a very compact area centered at the origin by attaching a new regularizer to the formal loss function of AE. The norm of hidden vectors at the bottleneck layer of these trained SAE can be considered as anomaly score [3]. For the fusion-based NAD approach, Thomas and Balakrishnan [22] improve the performance of IDS using the D-S theory to fuse multiple IDS together. Zhao et al. [11] use the D-S theory to combine multiple anomaly detection methods to create a more efficient fusion-based model. The fusion task works at the information layer of the 3-level structure, including the basic detection layer, information layer, and knowledge layer. However, these fusion models [11], [22] are evaluated on the out-date datasets [23]. Liu et al. [10] proposed a new way of optimizing the D-S theory to fuse 6 sensors. They introduced weights to control the balance between each AD algorithms. They conduct the BPA function based on assuming that the distance among the normal records is less than it's in the abnormal data. Recently, Mattar et al. [17] introduce a network anomaly detection model using the D-S theory. They provide some ways to determine a BPA function to fuse four different methods. In the recent survey, Li et al. [13] present some results in using data fusion techniques for network intrusion detection. They suggest we need to have more investigation on applying the D-S theory for anomaly detection.

III. BACKGROUND

This section presents fundamental backgrounds for understanding our proposed model. This consists of some anomaly detection algorithms and the Dempster-Shafer theory.

A. Dempster-Shafer Theory of Evidence

The Dempster-Shafer Theory of Evidence (the D-S Theory), also called the mathematical evidence of theory, is proposed by Arthur Dempster and improved by Glenn Shafer [24]. The theory is used to calculated the probability of an event by combining the evidence from multi-sources information. The proposition can be a subset of the given set of finite hypotheses, named FoD and flag by Θ, which is the set of all hypotheses that might happen in the entire system. Let $E_1, E_2, ..., E_n$ $(n > 2)$ be a number of evidence sources, so each E_i output in the set of k states can be written as follows:

$$\Theta = \{H_1, H_1, ...H_k\} \tag{1}$$

The relative concept to the D-S theory describes as following: The FoD for the fusion problem under the consideration having n exclusive and exhaustive hypotheses. The sets of all subsets of Θ is called as power set of Θ and are denoted by 2^Θ. Basic probability assignment (BPA) over a Θ is a function $m : 2^\Theta \rightarrow [0, 1]$ such that

$$\sum \{m(H))|H \subseteq \Theta\} = 1, m(\theta)) = 0 \tag{2}$$

where m(H) is represents the belief exactly committed to hypothesis H. In the concept of the D-S theory's rule, how to reach a conclusive decision based on each anomaly detectors in the fusion model is difficult. This problem can be solved by applying the Dempster's rule, a tool to combine mass assignments from multiple sources of information. When two mass assignments are combined, they might produce a null

set, or they might have an intersection point. In the first case, the mass assignment is considered to have a zero value and the mass assignment of the non-empty set is boosted by the factor K, commonly known as the conflict factor.

The D-S theory's rule in case of combine from many sources of information $E_1, E_2, ..., E_m$ can be presented as:

$$(m_1 \oplus ...m_n)(H) = \frac{\sum_{\cap_i E_i = H} m_1(E_1)\, m_2(E_2) ...m_n(E_n)}{\sum_{\cap_i E_i \neq H} m_1(E_1)\, m_2(E_2) ...m_n(E_n)} \quad (3)$$

B. Support Vector Machine - SVM

Typically of SVM for AD is OCSVM [18], the algorithm is aimed to find a decision function that returns a positive value in the specific region containing most of normal training data (called the normal region), and a negative value in the region encompassing the origin in a feature space. The idea behind this is to allocate the region containing the origin for anomalies to appear.

C. Kernel Density Estimation

KDE is used for estimating the probability density function of a sample in data [19]. KDE can be used for forming an AD model, as presented in [25], [3]. The main drawback of the model is to work with large datasets.

D. Local Outlier Factor

LOF [8] views the datapoints that have a considerably lower local density than their neighbors as anomalies. It estimates a density deviation score, called LOF. The larger the LOF score a given data point has, the higher the probability the data point is anomalous. The algorithm has shown its power on NAD domain [20]. In practice, however, it has some limitations when dealing with high-dimensional data. [3].

E. Shrink Autoencoder

Shrink Autoencoder is an extension of AutoEncoder (AE), which is a type of artificial neural network used to learn a representation (encoding) for a set of data, usually for dimensional reduction [3], [25]. Shrink AutoEncoders(SAE) introduced newly by Cao et al. [3]. By adding regularizer to the loss function of an AE to encourages the AE to form a reproduction of normal data in the latent space, this considers as an anomaly score.

IV. PROPOSED MODEL

This section presents our proposed model called fusion network anomaly detection (FuseNAD) based on the D-S theory. As mentioned in section I, single anomaly detectors (ADs) are assumed to perform efficiently on some certain attacks, but may yield poor results on others. In order to complement this drawback, the D-S theory will be use to fuse traditional anomaly detection methods and a deep learning one to form a robust anomaly detection model.

Fig. 3 briefly describes our model. It consists of four components: (1) Network data records; (2) Pre-processing module; (3) Four single ADs: SAE [3], OCSVM [18], KDE [19], and LOF [20]; and (4) Fusion module using the D-S theory (D-S Unit). In the training phase, the four anomaly detection methods were trained on only normal independently to obtain the four corresponding models. For each model, a classification threshold is also estimated so as to correctly classify a certain number percent of normal training data (say 90%). The threshold will split the output scores of these models into three parts as described in Fig. 2: Normal area (N), Anomaly area (A) and uncertain area (N/A). In the testing phase: each query data point will be fed into each of these single anomaly detection models, and their anomaly score outputs will be use as the inputs for the main part of the fusion model, D-S Unit. The D-S Unit component will do two main task in order to yield the final decision of our model. These task will be carried out through the BPA_AD function showed in Algorithm 1, and the DRC_AD rule described in Algorithm 2 that is the modification of Dempster's rule showed in Eq. 3. Based on the output of the DRC_AD function, the final decision will be made. The following paragraph will explain how the D-S Unit component work.

In order to apply the D-S theory for anomaly detection domain, we need to define system state, hypothesis and sources of evident information. In this domain, we have two system state: N representing normal status; and A representing anomaly status. The FoD function in Eq. 1 can be defined as, $\Theta = \{A, N\}$, so that a set of hypotheses $P(\Theta) = (A, N, NA, \emptyset)$, where $N \cap A = \emptyset$. Now we define a BPA function for the anomaly detection domain, BPA_AD, based on Eq. 2 as presented in Algorithm 1. We use terms s_i^{SAE}, s_i^{SVM}, s_i^{KDE}, s_i^{LOF} to refer to the output anomaly score of the data point x_i produced by SAE, OCSVM, KDE and LOF respectively. We normalize the output anomaly score produced by the four classifiers in the same format. This means that the larger score a model assigns for a data point, the higher probability the point is classified as anomaly. The anomaly score range produced by these ADs is split into three parts A, NA and N by using a threshold t and sigma parameter σ as showed in Fig. 2. The uncertain area, NA, is determined within the range $[t - \sigma, t + \sigma]$. Threshold t is determined so that ADs can classify correctly 90% of the normal training data. The sigma σ is set equal to $d * bw$, where d is the distance from the smallest anomaly score to threshold t, and bw is a scale parameter estimated based on each dataset. In the BPA_AD function, we introduce mass functions, namely cal_a, cal_b, with the input initial mass $x_0 \approx y_0 \approx 0.5$ and a very small value $z_0 = 10^{-5}$. This makes sure that the confident values $m(A), m(NA), m(N)$ will satisfy the criteria $m(A) + m(NA) + m(N) = 1$ and $m(\varnothing) = 0$ in Eq. 2.

When applying conventional D-S theory's rule in Eq. 3, all ADs should have the same role in despite of their different performance [10], [26]. In order to solve this problem, we modified the D-S theory's rule by adding weights to the confident values of hypotheses on each anomaly detectors to create a new D-S theory's rule. This is the main contribution of our paper. Specifically, we assign weights w_N and w_A for $m(N)$ and $m(A)$ respectively. This weights not only amongst

Algorithm 1 BPA_AD function

Input: anomaly score s_i, threshold t, sigma σ và Θ
Output: $m(A), m(NA), m(N)$

$a = cal_a(s_i, \sigma, x_0, z_0)$
$b = cal_b(a, x_0, y_0, z_0)$
if $s_i \in$ the area of s(1) **then**
  $m(N) = x_0 * a; m(A) = z_0; m(NA) = y_0 * b$
else
  $m(A) = x_0 * a; m(N) = z_0; m(NA) = y_0 * b$
end if
return $m(A), m(N), m(NA)$

single anomaly detectors, but also between the confidence values of the hypotheses. This is very different from a previous work [10] that only adds weights to support anomaly detectors involving in a fusion model. The Dempster's rule of combination function that combines the four anomaly detectors (DRC_AD) can be described in Algorithm 2.

Algorithm 2 DRC_AD function

Input: Θ, $m(A), m(N), m(NA), w_A, w_N$ of each AD
Output: $m(A), m(N), m(NA)$ of the Fusion model

Recalculate the mass based on their weight of each AD
$m(N) = m(N) * w_N$
$m(A) = m(A) * w_A$
$m(NA) = 1 - m(N) * w_A - m(A) * w_N$
Calculate $m(N), m(A), m(NA)$ of FuseNAD by Eq. 3
return $m(A), m(N), m(NA)$ of FuseNAD

In the Algorithm 2, w_A and w_N are used to weight the confident value of the state A and N respectively. These values are calculated as in [10]. Finally, the decision of the entire system can be made by using the rule: if the mass value of the state N, $m(N)$, is higher than that of the state A, $m(A)$, the state of the system will be is Normal, otherwise it is Anomaly.

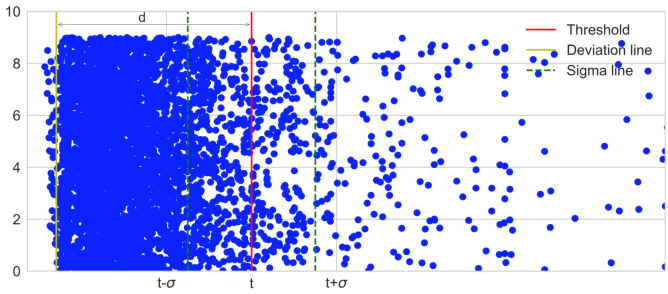

Fig. 1. Simulation of choosing threshold (t) and deviation (σ) from the score of training data (using SAE).

V. Experimental Settings

A. Data sets

The experiments are conducted on two publicly datasets, namely NSL-KDD [23] and UNSW-NB15 [27]. The normal network traffic presented in the UNSW-NB15 and NSL-KDD datasets is considered as normal data, while all network attacks are treated as anomalies.

NSL-KDD is created from the KDD Cup'99 dataset by removing redundant information and making it more difficult for classifying than KDD Cup'99. The dataset is split into two parts: the training set (67343 examples) and the testing set (9711 normal examples and 12833 anomalies). Each record consists of 41 features and a label (normal or a specific attack type). UNSW-NB15 is released recently, which attempts to solve some drawbacks in the previous public datasets. UNSW-NB15 consists of a training set (56000 records) and a testing set (37000 normal instances and 45332 anomalies). Each connections is constructed from 47 features, and can be assigned a normal label or one of the nine attack types. These datasets have some categorical features that are processed by using one-hot-encoding resulting in higher dimension versions, 122 for NSL-KDD and 196 for UNSW-NB15). During our experiments, both the label of the training set in NSL-KDD and UNSW-NB15 are removed, testing dataset including 3000 records of Normal and 3000 records of Abnormal was randomly selected from the original testing data.

B. Parameter Settings

The hyper-parameters of classical ADs and SAE are configured followed the previous work [3]. KDE and OCSVM use the same Gaussian kernel function and the scaling parameter γ (other form of kernel bandwidth) is set by a default value, $\gamma = \frac{1}{nf}$, where nf is the number of input features. The trade off parameter ν is set 0.5. The number of nearest neighbors in LOF, k, is chosen as 10% of the training size.

The network architecture and its hyper-parameters of SAE are set as in [3]. The weights of SAE are initialized by following [28] and the activation function is $tanh$. We train the model over 1000 epochs by an adaptive SGD algorithm (ADADELTA). For the D-S Unit, we choose $bw \in \{0.01, 0.1, 0.5\}$ and report the best results.

VI. Results and Discussion

We conduct two sets of experiments. The first is to compare the performance of each single ADs against an anomaly detection model based on the classical D-S theory's rule, called FuseNAD_C. The second one is the primary experiment, which compare the performance of the proposed model, FuseNAD, with single anomaly detectors. The performance of these models are measured by common parameters, accuracy (ACC), detection rate (DR), and false alarm rate (FAR).

It can be seen from Tables I and II that the fusion model based on the classical D-S theorys's rule (FuseNAD_C) yields quite good performance. Its performance is often better than those of single ADs. In terms of accuracy, however, it produces lower values than those of SAE and KDE on NSL-KDD, and also SAE on UNSW-NB15. These show that single ADs tend to have different performance on the same problem. When comparing our proposed model, FuseNAD, with FuseNAD_C and the four single ADs, we can see better performance (FAR, DR and ACC) on the two datasets. This indicates that the

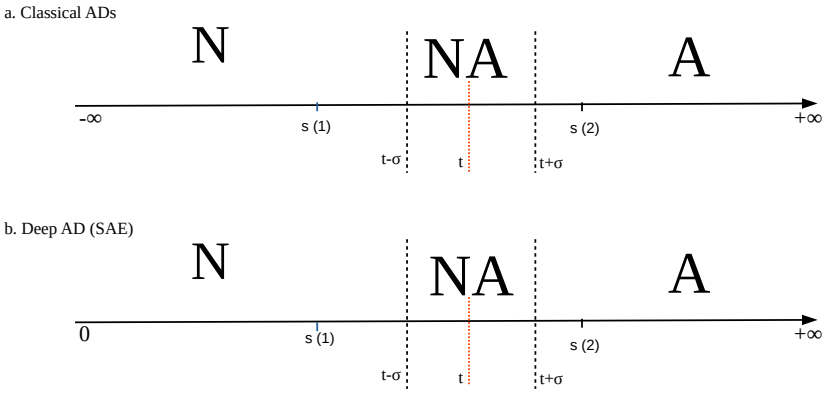

Fig. 2. The output anomaly score of anomaly detectors: a threshold split the score range into two areas s(1) and s(2).

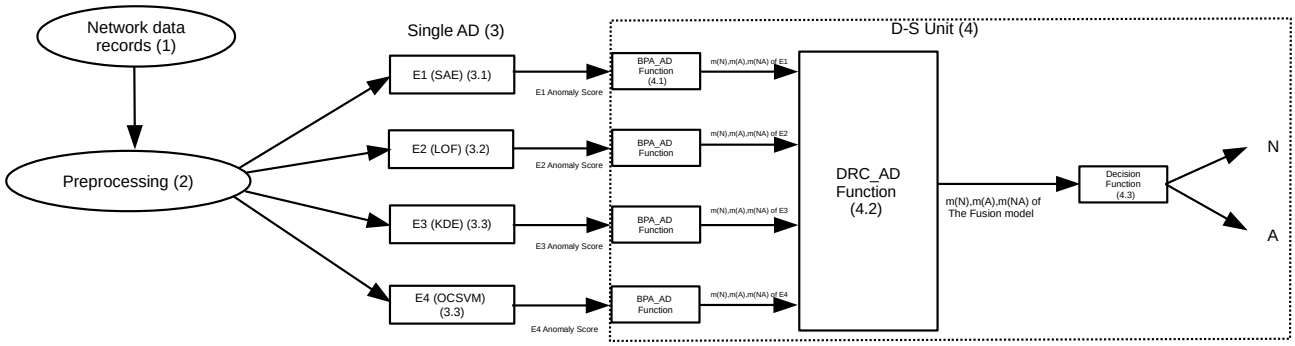

Fig. 3. Our proposal fusion network anomaly detection model (FuseNAD).

new D-S theory's rule can combine these single ADs more efficiently than the classical D-S theory's rules: taking the advantages from both of these anomaly detector.

TABLE I
COMPARISON OF INDIVIDUAL ADS AND FUSENAD ON NSL-KDD

Tested AM	NSL-KDD Dataset						
	TP	**FP**	**FN**	**TN**	**FAR**	**DR**	**ACC**
SAE	2773	227	399	2601	0.133	0.924	0.895
LOF	1145	1855	1181	1819	0.394	0.382	0.813
KDE	2923	77	554	2446	0.185	0.974	0.894
OCSVM	2158	842	215	2785	0.072	0.719	0.823
FuseNAD_C	2751	249	412	2588	0.917	0.137	0.889
FuseNAD	2786	214	383	2617	**0.128**	**0.929**	**0.901**

VII. CONCLUSIONS AND FUTURE WORK

This paper proposed a novel fusion-based network anomaly detection, called FuseNAD. The main purpose of FuseNAD is to take the advantages from both single anomaly detectors to increase classification accuracy and detection rate, and also reduce false alarm rate. We employ the Dempster-Shafer theory (D-S) of Evidence to construct a fusion-based network anomaly detection. In other words, we introduce a new D-S theory's rule that determines how to combine the confidence

TABLE II
COMPARISON OF INDIVIDUAL ADS AND FUSENAD ON USNW-NB15

Tested AD	UNSW Dataset						
	TP	**FP**	**FN**	**TN**	**FAR**	**DR**	**ACC**
SAE	2316	684	388	2612	0.129	0.772	0.821
LOF	1373	1627	1402	1598	0.467	0.458	0.495
KDE	2398	602	541	2459	0.18	0.799	0.809
OCSVM	1112	1888	244	2756	0.081	0.371	0.644
FuseNAD_C	2145	855	237	2763	0.079	0.715	0.818
FuseNAD	2265	735	170	2830	**0.057**	**0.755**	**0.849**

values of hypotheses produced by single ADs to make a final decision. We also adjust FoD, and a basic probability assignment (BPA) function to implement the rule. FuseNAD is constructed from four anomaly detection methods: a deep learning technique and three traditional ones.

We conducted the experiments on two publicly datasets to compare our proposed model with a fusion-based model using the classical D-S theory's rule, and four well-known anomaly detection methods. The results illustrates that our model outperforms both the four single ADs and the fusion-based model using the classical D-S theory's rules. This suggests that the new D-S theory's rule is more suitable for constructing a fusion-based anomaly detection model from single anomaly detection methods.

In the future, we plan to carry out an extensive experiment to investigate the performance of FuseNAD on a wide range of anomaly detection problems. We also develop new D-S theory's rules for the case the system state is larger than two (normal and anomaly), such as normal and a number of specific attack types.

ACKNOWLEDGEMENT

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.05-2019.05.

REFERENCES

[1] Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019.

[2] Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60:19–31, 2016.

[3] Van Loi Cao, Miguel Nicolau, and James McDermott. Learning neural representations for network anomaly detection. *IEEE transactions on cybernetics*, 49(8):3074–3087, 2019.

[4] Sarah M Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning. *Pattern Recognition*, 58:121–134, 2016.

[5] Mary M Moya, Mark W Koch, and Larry D Hostetler. One-class classifier networks for target recognition applications. *NASA STI/Recon Technical Report N*, 93, 1993.

[6] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

[7] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *ACM sigmod record*, volume 29, pages 93–104. ACM, 2000.

[8] Nathalie Japkowicz, Catherine Myers, Mark Gluck, et al. A novelty detection approach to classification. In *IJCAI*, volume 1, pages 518–523, 1995.

[9] Mayu Sakurada and Takehisa Yairi. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, page 4. ACM, 2014.

[10] Yuan Liu, Xiaofeng Wang, and Kaiyu Liu. Network anomaly detection system with optimized ds evidence theory. *The Scientific World Journal*, 2014, 2014.

[11] Xiaofeng Zhao, Hua Jiang, and LiYan Jiao. A data-fusion-based method for intrusion detection system in networks. *International Journal of Information Engineering and Electronic Business*, 1(1):32, 2009.

[12] Tim Bass. Intrusion detection systems and multisensor data fusion: Creating cyberspace situational awareness. *Communications of the ACM*, 43(4):99–105, 2000.

[13] Guoquan Li, Zheng Yan, Yulong Fu, and Hanlu Chen. Data fusion for network intrusion detection: A review. *Security and Communication Networks*, 2018:1–16, 05 2018.

[14] Junfeng Tian, Weidong Zhao, and Ruizhong Du. Ds evidence theory and its data fusion application in intrusion detection. In *International Conference on Computational and Information Science*, pages 244–251. Springer, 2005.

[15] K Saleem Malik Raja and K Jeya Kumar. Diversified intrusion detection using various detection methodologies with sensor fusion. In *2014 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)*, pages 442–448. IEEE, 2014.

[16] Vrushank Shah, Akshai K Aggarwal, and Nirbhay Chaubey. Performance improvement of intrusion detection with fusion of multiple sensors. *Complex & Intelligent Systems*, 3(1):33–39, 2017.

[17] Ahmed Mattar and Marek Z Reformat. Detecting anomalous network traffic using evidence theory. In *Advances in Fuzzy Logic and Technology 2017*, pages 493–504. Springer, 2017.

[18] Wenli Shang, Peng Zeng, Ming Wan, Lin Li, and Panfeng An. Intrusion detection algorithm based on ocsvm in industrial control system. *Security and Communication Networks*, 9(10):1040–1049, 2016.

[19] Matt P Wand and M Chris Jones. *Kernel smoothing*. Chapman and Hall/CRC, 1994.

[20] Aleksandar Lazarevic, Levent Ertoz, Vipin Kumar, Aysel Ozgur, and Jaideep Srivastava. A comparative study of anomaly detection schemes in network intrusion detection. In *Proceedings of the 2003 SIAM International Conference on Data Mining*, pages 25–36. SIAM, 2003.

[21] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

[22] Ciza Thomas and N Balakrishnan. Improvement in intrusion detection with advances in sensor fusion. *IEEE Transactions on Information Forensics and Security*, 4(3):542–551, 2009.

[23] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A Ghorbani. A detailed analysis of the kdd cup 99 data set. In *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, pages 1–6. IEEE, 2009.

[24] Glenn Shafer. *A mathematical theory of evidence*, volume 42. Princeton university press, 1976.

[25] Van Loi Cao, Miguel Nicolau, and James McDermott. A hybrid autoencoder and density estimation model for anomaly detection. In *International Conference on Parallel Problem Solving from Nature*, pages 717–726. Springer, 2016.

[26] Chunlin Lu, Yue Li, Mingjie Ma, and Na Li. A hybrid nids model using artificial neural network and ds evidence. *International Journal of Digital Crime and Forensics (IJDCF)*, 8(1):37–50, 2016.

[27] Nour Moustafa and Jill Slay. Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). In *2015 military communications and information systems conference (MilCIS)*, pages 1–6. IEEE, 2015.

[28] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.