A Data-Driven Approach for Network Intrusion Detection and Monitoring Based on Kernel Null Space

Thu Huong Truong[1(✉)], Phuong Bac Ta[1], Quoc Thong Nguyen[2], Huu Du Nguyen[3], and Kim Phuc Tran[4]

[1] School of Electronics and Telecommunications, Hanoi University of Science and Technology, Hanoi, Vietnam
huong.truongthu@hust.edu.vn
[2] Division of Artificial Intelligence, Dong A University, Da Nang, Vietnam
[3] Faculty of Information Technology, Vietnam National University of Agriculture, Hanoi, Vietnam
[4] GEMTEX Laboratory, Ecole Nationale Sup des Arts et Industries Textiles, BP 30329 59056, Roubaix Cedex 1, France

Abstract. In this study, we propose a new approach to determine intrusions of network in real-time based on statistical process control technique and kernel null space method. The training samples in a class are mapped to a single point using the Kernel Null Foley-Sammon Transform. The Novelty Score are computed from testing samples in order to determine the threshold for the real-time detection of anomaly. The efficiency of the proposed method is illustrated over the KDD99 data set. The experimental results show that our new method outperforms the OCSVM and the original Kernel Null Space method by 1.53% and 3.86% respectively in terms of accuracy.

Keywords: Network security · Kernel Quantile Estimator · One-class classification · Kernel Null Space · Support vector machine

1 Introduction

Nowadays, every computer system has the security policies but these policies have not been strong enough to prevent or detect all new types of attacks. Therefore, building one monitoring system is essential to alarm novelties early. Detecting incoming intrusion early helps systems reduce the damage and protect the crucial information. Intrusion detection system (IDS) is the key to resolve these problems and attract a lot of researchers to work on the issue [3]. IDS has been used in a great number of applications such as network intrusion, fraud detection and security systems.

Currently, there are two families of mechanisms in IDS: signature-based IDS and anomaly-based IDS. In this paper, we focus on developing an anomaly-based

T. Q. Duong et al. (Eds.): INISCOM 2019, LNICST 293, pp. 130–140, 2019.
https://doi.org/10.1007/978-3-030-30149-1_11

IDS solution, in which the designed IDS system is trained based on knowledge of normal traffic only; the system does not need to be trained with attack data traces in advance to know if incoming traffic is anomaly or normal. This characteristic is good for the attack detection aspect because attack manners may vary over time; so, we may be in the situation that the system was not trained with an attack pattern before. In case of never-seen-before attacks, an IDS system based on training of attack and normal data traces may not be effective any more.

Among the anomaly-based IDS solution family, Novelty Detection is a research direction attracting a great number of researchers. A model is built from normal data to detect unknown abnormality by novelty detection algorithms such as OCSVM [6,11] and Kernel Null Space [1,2,4]. There is also an approach in intrusion detection using Statistical Process Control [7].

Our proposed solution aims at improving the performance of the Kernel Null Space method [2] in terms of accuracy. To be more specific, we propose using a Control-Chart based method called Kernel Quantile Estimator to determine the detection threshold dynamically driven by each specific training data set instead of using a fixed threshold as described in the existing Kernel Null Space solutions [1,2,4]. The Control Chart Based on a Kernel Estimator of the Quantile Function was also developed in [5]. In addition, we also optimize the kernel parameter of the kernel function to improve the performance of novelty detection.

The rest of the paper is organized as follows: Sect. 2 elaborates the related work. Our proposed Kernel Null Space solution for Novelty Detection is provided in Sect. 3, followed by the performance evaluation in Sect. 4. Finally, conclusion is given in Sect. 5.

2 Related Work

Generally, the novelty detection issues can be divided into two types based on the number of known classes during the training phase: one-class and multi-classes. Since our work focuses on one-class classification, we will review the state of the art for the family of one-class novelty detection. To the best of our knowledge, Kernel Null Space has the highest performance in novelty detection and there are only three studies dealing with one-class classification in novelty detection using this method [1,2,4]. In [2], the authors proposed Kernel Null Space for novelty detection but they made the experiment with a fixed threshold and a fixed kernel parameter of the kernel function. In paper [1], the authors also improved the performance of the original method. However, they only concentrated on decreasing the timing operating of the algorithm, the accuracy remains unchanged. Following this trend, paper [4] improved the solution proposed in [2] by decreasing the complexity of the kernel null space method without taking the accuracy into account.

From another approach, the OCSVM method, which detects novelty by finding the boundary of training data with maximum margin, is often used to solve the one-class novelty detection problem, for example, in [11,12]. The OCSVM

method has received more extensive attention since it can easily handle nonlinear data with kernel trick and also achieve a high level of detection accuracy [11].

As mentioned in Sect. 1, the solution we will explain throughout the paper is to improve the accuracy of the Kernel Null Space method [2] in the favor of anomaly detection. We propose a solution combining Kernel null space and Control chart to automatically define an efficient detection threshold stemming from each training data trace.

Simultaneously, we also use the optimizing parameter method proposed in [11] to increase the accuracy for the algorithm. Our proposed solution is proved to outperform the Kernel Null Space methods in [1,2,4] and OCSVM in [11,12] in terms of Accuracy.

3 Intrusion Detection Scheme Using the Enhanced Kernel Null Space Method

For an intrusion detection system to work accurately, we propose a so-called an enhanced Kernel Null Space method to improve the accuracy of detecting novelty samples. The scheme is elaborated as follows:

- Pre-process and normalize the attributes of the data set.
- Design an enhanced Kernel Null Space method to analyze data inputs.

In this method, the threshold is computed by Kernel Quantile Estimator [9] for a given probability q. Figure 1 shows the process of the detection scheme, in which Internet raw data coming to the detection system will be pre-processed, and analyzed to test if it is a novelty (i.e. anomalies).

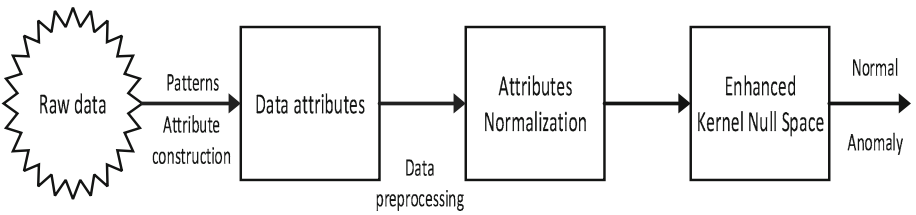

Fig. 1. Intrusion detection process

3.1 Pre-processing and Normalizing Data Attributes

In order to do the comparison with different intrusion detection methods, in the experiment, we use the NSL-KDD data set [10] which is commonly used. Each sample in this NSL-KDD corresponds to a real connection in the simulated military network, containing 41 attributes with Normal and Attack-type labels. In the data set, there are 39 types of attacks divided in 4 groups:

– DoS - Denial of services, e.g. syn flood.
– R2L: Unauthorized access from a remote machine, e.g. guessing password.
– Probing: surveillance and other probing, e.g. port scanning.
– U2R: unauthorized access to local super user (root) privileges, e.g. buffer overflow.

To make the data set simpler, reducing the redundancy without losing the information, we pre-process the data set as follows:

- Conversion from the Symbolic type to the Numeric type: there are 3 attributes in the Symbolic manner such as: Protocol, Service, Flag which are needed to be converted to the Numeric type to be compatible with the inputs of the algorithm. The symbolic values are labeled as in Table 1.

Table 1. Symbolic-typed attributes

Attribute	Symbolic	Corresponding numeric value
Protocol_type	UDP	1
	TCP	2
	ICMP	3
Flag	OTH	1
	REJ	2
	RTSO	3
	RTSOSO	4
	RSTR	5
	S0	6
	S1	7
	S2	8
	S3	9
	SF	10
	SH	11
Service	65 values	From 1 to 65

- Normalization: Normalization of data in the NSL-KDD data set is necessary since there are many big values in comparison with much smaller values in the set. We apply the Min-max normalization method to turn all values to the range [0, 1] as follows:

$$\hat{v}_i = \frac{v_i - min(v_i)}{max(v_i) - min(v_i)}, \text{ for } i = 1, 2, \ldots, 41 \quad (1)$$

where:

v_i: value of one attribute before normalization.
$\hat{v}_i$: value of one attribute after normalization.
$i = 1, ..., 41$: 41 attributes.

3.2 Enhanced Kernel Null Space

Before describing the enhanced Kernel Null Space, we briefly re-call the One-Class Classification using Kernel Null Space proposed in [2]. Let us consider a dataset of N training samples $\{x_1, x_2, \ldots, x_N\}$, with each $x_i \in R^D$, and D is the number of observed features. In the one-class setting, all the training samples belong to a single target class. The input features $X = [x_1, x_2, \ldots, x_N]$ are separated from the origin in the high-dimensional kernel feature space similar to one-class SVM [8]. As described in [2], a single null projection direction is computed to map all samples on a single target value s. A test sample x^* is projected on the null projection direction to obtain the value s^*. Figure 2 illustrates the one-class approach with kernel null space. The novelty score of x^* is the distance between s and s^*:

$$NoveltyScore(x^*) = \mid s - s^* \mid . \tag{2}$$

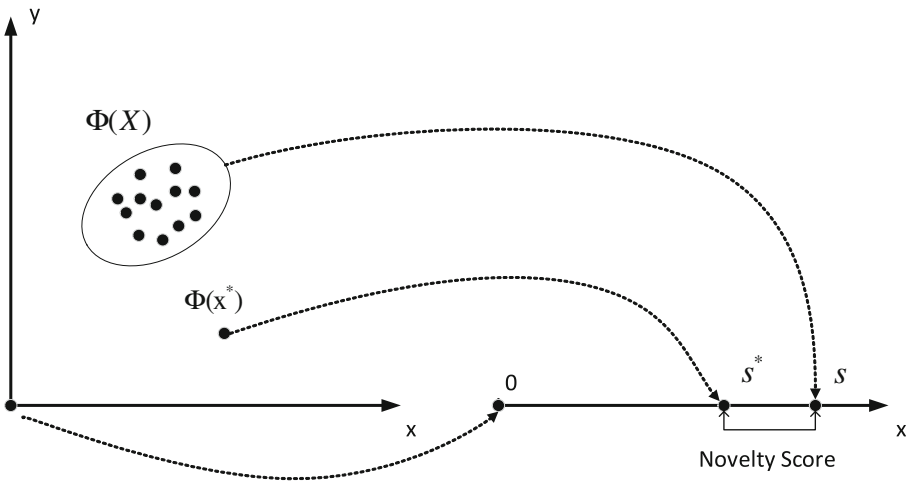

Fig. 2. The samples are separated from the origin in the kernel feature space with a mapping Φ, then mapped on a point s, and the novelty score of a testing sample x^* is the distance of its projection s^* to s.

A large novelty score indicates more likely novelty. In [2] and [1], a hard decision threshold $\theta_{threshold}$ is used to determine whether the test sample x^* belongs to the target class or not. Determining the threshold plays a very important role to the performance of the novelty detection process. To the best of our knowledge, this threshold has been selected heuristically up till now. Therefore, in this study, we propose an intrusion detection scheme based on an enhanced version of this Kernel Null Space method.

The procedure of the enhanced Kernel Null Space method is illustrated in Fig. 3 with two phases: the training phase and the detection phase.

In the training phase: training data samples $\{x_1, x_2, \ldots, x_N\}$, which have been already pre-processed, will be mapped on a point s in the Null Space F. The intrusion detection system uses another data set called the validation set that comprises other normal data samples $\{y_1, y_2 \ldots, y_M\}$. Each sample y_i of the validation set is mapped on a point $\hat{s}_i$ in the feature null space, for which $NoveltyScore(y_i)$ is calculated. After mapping all samples of the validation set and calculating Novelty scores for all of them, a set $\{NoveltyScore(y_i)\}$ is formed. Based on this set of novelty scores, we use the Kernel Quantile Estimator to derive the threshold $\theta_{threshold}$, which will be described in Sect. 3.2.

During the detection phase in real time, when a test data sample x^* comes, the system maps it on a point s^* and then calculate its $NoveltyScore(x^*)$. Then by comparing the $NoveltyScore(x^*)$ with $\theta_{threshold}$ found in the training phase, x^* can be classified as Normal or Anomaly.

In the following subsections, we will elaborate how we achieve an optimal kernel parameter on the given training data set and how to calculate threshold $\theta_{threshold}$ by Kernel Quantile Estimator.

Determination of Kernel and Kernel Parameter. In this paper, we select the Gaussian kernel (or Radial Basic Function (RBF)) for Kernel Null Space which is commonly used.

$$k(x, y) = \exp\left(\frac{-\| x - y \|^2}{2\sigma^2}\right) \tag{3}$$

where: σ stands for the kernel parameter in [0, 1].

Using the method proposed in [11], the optimal sigma σ^* is estimated from the data set $\{x_1, x_2, \ldots, x_N\}$. The optimal σ^* is the one that maximizes the objective function $J(\sigma)$

$$J(\sigma) = \frac{2}{N}\sum_{i=1}^{n} \exp\left(-\frac{Near(x_i)}{2\sigma^2}\right) - \frac{2}{N}\sum_{i=1}^{n} \exp\left(-\frac{Far(x_i)}{2\sigma^2}\right) \tag{4}$$

Denote the nearest and farthest neighbors distances as:

$$Near(x_i) = \min_{j \neq i} \| x_i - x_j \|^2$$

$$Far(x_i) = \max_{i} \| x_i - x_j \|^2$$

Threshold Calculation Based on Kernel Quantile Estimator. As mentioned, the threshold for the Novelty Score is the crucial key for the accuracy in anomaly detection. A common method to choose a good threshold that we have observed up till now is checking various discrete threshold values in the increasing order until the test system outputs highest accuracy. But when we have to cope with continuous values, that heuristic check-up hardly finds a good threshold we can not check all continuous values.

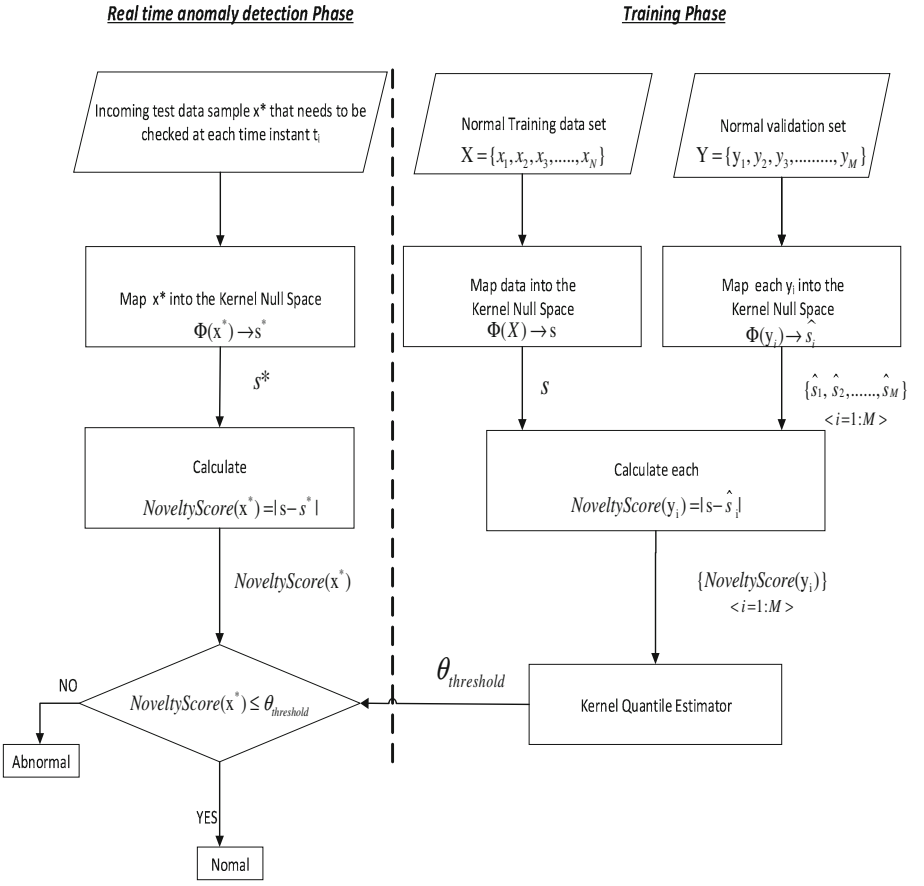

Fig. 3. Detection procedure using Kernel Null Space and Kernel Quantile Estimator

The set of the novelty scores is denoted by $\{NS_1, NS_2, \ldots, NS_M\}$ and investigated for the probability density distribution. As observed in Fig. 4, the Novelty Score values $\{NS_1, NS_2, \ldots, NS_M\}$ can not be approximated by a normal distribution, i.e. the underlying distribution of the sample is unknown. In this case, nonparametric methods could be used to explore this unknown underlying.
In this paper, we use the **Kernel Quantile Estimator** [9] to estimate $\theta_{threshold}$ over the set of Novelty Score values.

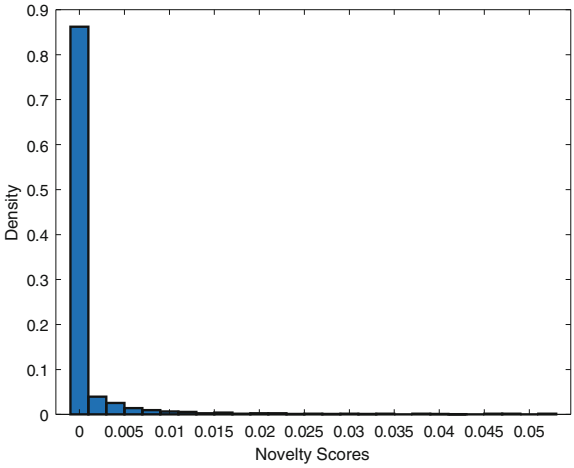

Fig. 4. Probability density distribution of novelty scores

Let $NS_{(1)} \leq NS_{(2)} \leq \ldots \leq NS_{(M)}$ denote the corresponding order statistics of the novelty scores. Suppose that $K(.)$ is a density function symmetric about Zero and that $h \to 0$ as $n \to \infty$, the Kernel Quantile Estimator can be calculated as follows [9]:

$$KQ_p = \sum_{i=1}^{N} \left[\int_{\frac{i-1}{n}}^{\frac{i}{n}} K_h(t-p)dt \right] NS_{(i)} \tag{5}$$

where $h > 0$ is the bandwidth. The bandwidth h controls the smoothness of the estimator for a given sample of size n. $K_h(.) = \frac{1}{h}K(\frac{.}{h})$. And p is the proportion of the quantile.

Here we use the standard Gaussian kernel for the resulting estimate KQ_p which is a smooth unimodal,

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}) \tag{6}$$

The selection of h is important in kernel density estimation: a large h will lead to an over-smoothed density estimate, while a small h will produce a ragged density with many spikes at the observations. As described in [9], the bandwidth computed as

$$h_{opt} = \left(\frac{pq}{n+1} \right)^{\frac{1}{2}} \tag{7}$$

where: $q = 1 - p$

For a lot of continuous distributions used in statistics, specific quantiles such as the $p = 0.95$, 0.975, and 0.99 quantiles are tabulated. Therefore, in our experiment, we have investigated 3 cases of q: 0.05, 0.025 and 0.01 respectively. These 3 q values corresponds to 3 threshold value $KQ(p = 1 - q)$ (i.e. $\theta_{threshold}$).

4 Performance Evaluation

4.1 Data Description

In this experiment, we use the NSL-KDD data set [10] to test the detection accuracy of the proposed solution. The training data set contains 13449 normal samples After training the system, the system performance is checked by using 6000 normal and abnormal samples. To test performance, we use all 41 attributes/parameters of the data set.

4.2 Performance Analysis

There are some important performance metrics in the novelty (anomaly) detection domain that have been widely used to analyze the performance of a certain detection method:

- Accuracy = $\frac{TP+TN}{TP+FP+TN+FN}$
- ReCall - True Positive Rate or Sensitivity = $\frac{TP}{TP+FN}$
- FPR - False Positive Rate: $FPR = \frac{FP}{FP+TN}$

Where TP (True Positive) is the number of anomalies correctly diagnosed as anomalies; TN (True negative) is the number of normal events correctly diagnosed as normal; FP (False Positive) is the number of normal events incorrectly diagnosed as anomalies; and FN (False Negative) is the number of anomalies incorrectly diagnosed as normal events.

In our test, we compare the performance of the enhanced Kernel Null Space with the original Kernel Null Space in which the threshold is heuristically selected and fixed at 0.05 [2] and with the One Class Support Vector Machine method (OCSVM) [11].

As mentioned in Sect. 3.2, we have tested with 3 different q values: 0.01, 0.025 and 0.05 and found out that $q = 0.025$ brings best performance. The results are shown in Table 2:

As can be seen in Table 2, with the normalized and pre-processed 41-attribute data set $\{X_1, X_2, \ldots, X_N\}$, the optimal kernel parameter estimated is $\sigma^* = 0.5957$. Subsequently, from the given data set of Novelty scores $\{NS_1, NS_2, \ldots, NS_M\}$, supposed that $q = 0.025$, the threshold is $\theta_{threshold} = 0.0233$.

Table 2. Performance comparison

$\sigma = 0.5957$	Kernel Null Space			OCSVM	Origin Kernel Null Space with fixed threshold = 0.05
	q = 0.05 $\theta_{threshold}$ = 0.0097	q = 0.025 $\theta_{threshold}$ = 0.0233	q = 0.01 $\theta_{threshold}$ = 0.0514		
Accuracy	0.9548	0.9598	0.92	0.9445	0.9212
FPR	0.0443	0.018	0.006	0.0433	0.006
Recall	0.954	0.9377	0.846	0.9323	0.8483

The obtained results show that; the enhanced Kernel Null Space slightly outperforms the OCSVM and the original Kernel Null Space methods in both terms of Accuracy and False Positive rate while a bit inferior to the Original Kernel Null Space method in terms of Recall.

5 Conclusion and Future Work

In this research, we have proposed and elaborated an Intrusion Detection System using the so-called enhanced Kernel Null Space method with data-driven threshold retrieval. The proposed solution with data-driven findings such as $q = 0.025$ and $\sigma = 0.5957$ is proved to outperform the current OCSVM and Original Kernel Null Space methods in terms of Detection Accuracy and False Positive Rate.

In the future, we would like to address the intrusion detection and the monitoring problem using deep learning, targeting on time series data with uncertainties. We also focus on the detection ability of our proposed approach for large stream data.

References

1. Bodesheim, P., Freytag, A., Rodner, E., Denzler, J.: Local novelty detection in multi-class recognition problems. In: 2015 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 813–820. IEEE (2015)
2. Bodesheim, P., Freytag, A., Rodner, E., Kemmler, M., Denzler, J.: Kernel null space methods for novelty detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3374–3381 (2013)
3. Borkar, A., Donode, A., Kumari, A.: A survey on intrusion detection system (IDS) and internal intrusion detection and protection system (IIDPS). In: International Conference on Inventive Computing and Informatics (ICICI), pp. 949–953. IEEE (2017)
4. Liu, J., Lian, Z., Wang, Y., Xiao, J.: Incremental kernel null space discriminant analysis for novelty detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 792–800 (2017)
5. Mercado, G.R., Conerly, M.D., Perry, M.B.: Phase i control chart based on a kernel estimator of the quantile function. Qual. Reliab. Eng. Int. **27**(8), 1131–1144 (2011)
6. Nguyen, Q.T., Tran, K.P., Castagliola, P., Truong, T.H., Nguyen, M.K., Lardjane, S.: Nested one-class support vector machines for network intrusion detection. In: 2018 IEEE Seventh International Conference on Communications and Electronics (ICCE), pp. 7–12. IEEE (2018)
7. Park, Y., Baek, S.H., Kim, S.H., Tsui, K.L.: Statistical process control-based intrusion detection and monitoring. Qual. Reliab. Eng. Int. **30**(2), 257–273 (2014)
8. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural Comput. **13**(7), 1443–1471 (2001)
9. Sheather, S.J., Marron, J.S.: Kernel quantile estimators. J. Am. Stat. Assoc. **85**(410), 410–416 (1990)
10. Tavallaee, M., Bagheri, E., Lu, W., Ghorbani, A.A.: A detailed analysis of the KDD cup 99 data set. In: 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009, pp. 1–6. IEEE (2009)

11. Trinh, V.V., Tran, K.P., Huong, T.T.: Data driven hyperparameter optimization of one-class support vector machines for anomaly detection in wireless sensor networks. In: 2017 International Conference on Advanced Technologies for Communications (ATC), pp. 6–10, October 2017. https://doi.org/10.1109/ATC.2017.8167642
12. Wang, Y., Wong, J., Miner, A.: Anomaly intrusion detection using one class SVM. In: Information Assurance Workshop, pp. 358–364 (2004)