Journal of Engineering and Applied Sciences 14 (2): 415-422, 2019
ISSN: 1816-949X

Host-Based Intrusion Detection Architecture Based on Rough Set Theory and Machine Learning

[1]Hayri Sever and [1, 2]Ahmed Nasser
[1]Department of Computer Engineering, Hacettepe University, Ankara, Turkey
[2]Department of Control and Systems Engineering, University of Technology, Baghdad, Iraq

Abstract: Intrusion detection is considered as a remarkable approach used in network and computer security. In this study, we proposed a host based IDS architecture that exploits the adaptive aspect of machine learning mechanisms and rough set theory. The proposed IDS architecture involves using new feature extraction method based on statistical measures which generate a training dataset with less feature space compared to the ones generated by traditional methods used in literature. The proposed IDS architecture also utilizes the principles of rough set theory in term of attribute reduction techniques. Two variations of rough set attribute reduction (Crisp and fuzzy) are considered to reduce the feature space by removing redundant and irrelative attributes which leads to improving the system performance. Rough Set Classification (RSC) approach is used to generate the IDS decision model by taking the form of "IF-THEN" rules using MODLEM rule induction algorithm. Our test and comparison of RSC with four standard classification methods showed that the RSC yielded highly accurate results in the term of F-score. The test experiments also show the impact of the attribute reduction method on increasing the classification accuracy.

Key words: Host-based IDS, computer security, machine learning, rough set theory, feature selection, novelty detection

INTRODUCTION

Host-based Intrusion Detection Systems (IDS) also called Anomaly Detection (AD), collects information regarding the application and the processes runs on a specific system. IDS agents are installed on host machines and act as sensors to collect information about system events corresponding to a specific process execution. System calls are commonly used to monitor the system events and are recorded by the operating system via audit trails (Garcia-Teodoro *et al*., 2009). Data recorded from normal system applications (non-malicious) can be used as a reference for learning the decision engine model of the IDS under normal system behavior. When a new data become available, the learned system model can be used to determine whether they constitute normal or abnormal system behavior. In the case of abnormal system behavior, the IDS either raises an alarm to warn the user or reacts by closing the network connection or terminating the process execution.

The IDS agents monitor the system in real-time and if the detection approach that IDS relies on is not optimized, it will consume the resources of the host system that installed on and results in a reduction in the performance of both the host and IDS (Warrender *et al*., 1999).

From this regard, we consider designing a host-based IDS architecture based on rough set theory and machine learning algorithms by taking into our consideration the data analyzing and processing complexity reduction.

Our architecture involves using a new different statistical based feature extraction technique rather than the traditional N-gram model and frequency based pattern recognition methods which were widely used in previous IDS studies (Creech and Hu, 2013, 2014, Xie *et al*., 2014a, b; Xie and Hu, 2013). These traditional machine learning pattern recognition methods tend to generate data with a large dimensional feature space which can increase the complexity of the machine learning model used in the IDS and ladies to decrease the system efficiency (Warrender *et al*., 1999).

The proposed IDS architecture also utilizes the rough set theory which has been widely used for attribute reduction with much success to reduce the number of features by selecting only relevant features with class label of the dataset (Li *et al*., 2016). This will enhance the performance of IDS machine learning system by reducing the time and memory required to make the decision.

We also consider using a rough set based machine learning classification methods to build our host-based IDS detection model and comparing its performance against different standard machine learning classifiers.

Corresponding Author: Hayri Sever, Department of Computer Engineering, Hacettepe University, Ankara, Turkey

Literature review: Creech and Hu (2013) introduced a new dataset for modern Linux IDS called ADFA-LD. This dataset is based on the latest exploits and attacks in modern Linux. Different IDS decision-engine algorithms were used to evaluate the performance of the new dataset. Based on the results given in their research, the research discovered that the KDD dataset proposed by Lee *et al.* (1998) does not provides a suitable performance against modern attacks and the features currently used to build IDS decision-engine models are not sufficient to achieve good results.

These findings which are mentioned above motivated us to propose and use a different statistical feature to these commonly used in IDS decision-engine classification methods. These conclusive findings have also led to the preference for ADFA-LD over KDD dataset in building a host-based AD decision engine (Creench and Hu, 2014).

In research presented by Xie and Hu (2013) Host-based Anomaly Detection Systems (HADSs) were evaluated by analyzing ADFA-LD using feature extraction techniques based on length, common patterns and frequency. Frequency-based K-Nearest Neighbor (KNN) classification was applied to ADFA-LD to build a machine-learning-based HADS. The evaluation results show that normal behavior can be efficiently detected using frequency-based features. However, distinguishing between normal and anomalous behavior requires a further improvement.

Xie *et al.* (2014a) proposed a HADS using one-class Support Vector Machine (SVM) algorithm and short sequence model based on ADFA-LD. A short sequence model is built using the sub-sequences form the normal traces and the test instances that does not fit with this model will be considered as abnormal. To build the short sequence matrix, the training traces were continuously transformed into fixed-length vectors and duplicate vectors were removed from the matrix to decrease the complexity. The short sequences obtained from the trace files were weighted with respect to frequency and used to train one-class SVM classifier. Receiver Operating Characteristic (ROC) curves were used to validate the one-class SVM trained on the short sequence model. The validation results show a reasonable performance along with a low computational cost.

Xie *et al.* (2014b) presented a frequency-based algorithm that applied to ADFA-LD. First, Principal Component Analysis (PCA) was used to map the original high-dimensional frequency vectors into a lower-dimensional space and a variety of distance functions were examined to validate the effectiveness of these new vectors. Frequency-based algorithms such as k-Means Clustering (kMC) and KNN were tested in a different setting with performance metrics such as the false positive rate and accuracy th at applied in ROC form. The results obtained by Xie *et al.* (2014b) were evaluated according to the performance of the KNN and kMC frequency-based algorithms using ADFA-LD. It was found that the KNN algorithm was less effective in detecting attacks with kMC giving a higher detection rate for most attack types.

Khreich *et al.* (2017) presented a novel feature extraction technique, a new anomaly detection system to reduce the false alarm rate. The suggested feature extraction approach starts by segmenting the system call traces into multiple N-grams of variable length and mapping them to a fixed-size sparse feature vectors which were used to train OC-SVM detectors and then a performance evaluation was conducted on ADFA-LD dataset.

Vijayanand *et al.* (2017) developed a multi-SVM based intrusion detection system in which each classifier detects specific attack only and used to detect the cyber-attacks occurring in Advanced Metering Infrastructure (AMI) communication network of smart grid using a mutual information technique that selects the input features of classifier by analyzing the relation between different features with attacks. The performance of developed intrusion detection system was evaluated by training and testing the classifier with ADFA-LD dataset.

MATERIALS AND METHODS

Dataset: A dataset that consists of class label information regarding normal and abnormal (attack) system behaviors can be used to build a machine learning classification model that can be subsequently used to analyze the current system behaviors and decide whether there is an attack present or not. In this research, we used Australian Defense Force Academy Linux Dataset (ADFA-LD) dataset to build our IDS architecture.

ADFA-LD by Creech and Hu (2013) is generated by the Cyber Security Lab., at the University of New South Wales, Canberra, Australia and has been available online, since, 2013. The ADFA-LD data are collected on a modern Linux local server that offers remote access, database, web server and file sharing services. ADFA-LD dataset is consisting of multiple trace files that generated by recording the system calls corresponding to a specific process in the form of their syscall identity number. ADFA-LD dataset contains six main attack categories (Hydra FTP, Hydra SSH, Adduser, Webshell, Linux and Java Meterpreter) (Creech and Hu, 2014) as well as the normal system behavior.

The exploits used in ADFA-LD represent a complete system compromise from initial penetration through to privilege escalation. A comparison has shown that ADFA-LD offers more complexity than its competitor KDD99 intrusion detection dataset (Lee *et al*., 1998). Thus, it can represent the current cyber-attacks more realistically and can be used as a more relevant metric to evaluate the performance of IDSs (Xie and Hu, 2013). In the following study, we used novelty detection approach to compare the ADFA-LD dataset with the enhanced version of the KDD99 (NSL-KDD) that presented by Tavallaee *et al*. (2009).

Novelty detection: Novelty or anomaly can be considered as events or patterns in the data which doesn't match an expected behavior that the data should produce. These novel events may occur in the system infrequently and can cause a malfunction in the system operation. Novelty detection is referring to identify the system abnormal behaviorss that doesn't fit the normal system state by identifying the outliers that differ from the ordinary data distribution. There are two type techniques used for novelty detecting, parametric methods such as Gaussian mixture models and nonparametric such kernel density estimators. Non-parametric techniques have the advantage over the parametric because it doesn't requiring any knowledge or assumptions about the underlying distribution of the data (Miljkovic, 2010).

Parzen window is non-parametric density estimation method which is widely used in novelty detection. For a data distribution $D = \{x_1, x_2, ..., x_n\}$, Parzen window used generates a Probability Density Function (PDF) estimator $\hat{P}(x)$ such as:

$$\hat{P}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right) \quad (1)$$

Where:

k = The kernel function
h = A smoothing parameter

This method can be used for novelty detection by generating a density estimator $\hat{P}(x)$ from the normal training set and if probability density of a given test pattern is below some predefined threshold then it considered to be novel. We compare ADFA-LD dataset with it competitor NSL-KDD (Tavallaee *et al*., 2009) with the regard to novelty detection.

The NSL-KDD dataset is an enhanced version of the KDD99 (Lee *et al*., 1998) intrusion dataset which is suggested to solve some of the inherent drawbacks in the KDD99. It has many advantages over the original KDD dataset such as not include redundant records has a sufficient number of records for each train and test subsets, the number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD dataset (Tavallaee *et al*., 2009).

Parzen window novelty detection model built by using the class of the normal system behavior in both of ADFA-LD and NSL-KDD datasets and evaluated using Area Under the Receiver Operating Characteristic Curve ROC (AUC) (Fawcett, 2006) as shown in Fig. 1a. The true positive axis on ROC is corresponding to the test instances classified as novel correctly, where false positives axis corresponds to instances classified as novel incorrectly. The maximum value for the AUC is 1.0, which denotes an excellent classifier.

The result in Fig. 1a shows that both the ADFA-LD and NSL-KDD datasets have similar performance in term of detecting the novel behaviors, although, the ADFA-LD is more complex than the NSL-KDD as it shown by the comparison between the two datasets that done in (Creech and Hu, 2013).

Methods for host-based IDS: The proposed IDS architecture shown in Fig. 1b, utilizes both of rough set theory and machine learning approaches such as feature extraction, feature reduction and classification. Statistical feature extraction approach is considered to generate a training dataset with a minimum number of attributes. Two variations of rough set attribute reduction method (Crisp and fuzzy) are used to reduce the feature space by selecting only relevant features related to the class labels of the dataset. A rough set based classification approach is considered for building the IDS decision model based on rule induction algorithm. Each part of the proposed host-based IDS architecture is explained as following (Fig. 2):

Prepressing: Data preprocessing is the transformation of data into another format, so, it will be more effectively and easily processed. There are different methods used for preprocessing. Sampling is the process of choosing the ideal subset from a big chunk of information, feature extraction is the process that extract the important information form the data, de-noising is the process that eliminates the noise from data and finally, the normalization process of rearranging and reorganizing the data (Dua and Du, 2016).

Feature extraction and weighting: we propose a new feature extraction technique based on the statistical analysis of the ADFA-LD intrusion detection dataset. This statistical feature extraction technique emphasizes different statistical measures such as minimum, maximum, standard deviation, variance, most

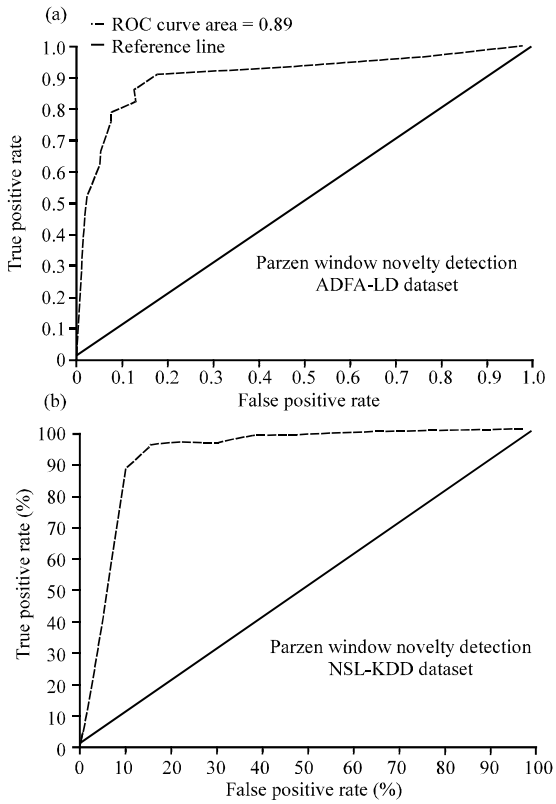

Fig. 1: A comparison between the performance of ADFA-D and NSL-KDD datasets, by using Perzen window novelty detection method; a) ROC, AUC = 0.89 and b) ROC, AUC = 0.90

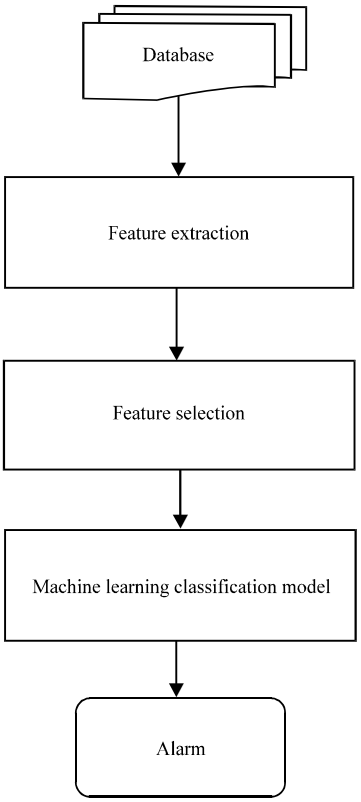

Fig. 2: The proposed host-baased IDS architecture

Frequent syscall, second most frequent syscall, median, skewness, harmonic mean, kurtosis) which can induce a training data with lower dimensional feature space compared to that one generated by traditional pattern recognition methods used in the previous related studies. These lower dimensional training data can be used to build less complex machine learning detection model for the IDS and can reduce the data, analyzing and processing complexity which leads to increase the performance of both IDS and the host system.

Feature selection: Feature selection or attribute reduction can be considered as the problem of finding the optimal feature subset that has most relevance to the class labels. By removing redundant features, feature selection algorithms reduce both the system complexity and processing time and enhance the recognition accuracy. Among several feature selection methods, "Sequential Feature Selection" refers to iterative algorithms that search in a sequential deterministic manner for the best (suboptimal) feature subset (Dash and Liu, 1997). In this study, we investigate two different feature selection methods, Crisp and fuzzy rough sets to reduce the feature space of the dataset.

Rough Sets Theory (RST) is a mathematical tool that had been used successfully to discover data dependencies and reduce the number of attributes contained in a dataset by purely structural methods (Pawlak, 2012). The rough sets attribute reduction method relies on the RST. One of the major advantages of RST is that it reduces the number of features in a specific dataset without the need for additional information (Wang *et al*., 2016a).

In rough set based attribute reduction method, rough sets are used to define equivalence classes approximately. A specific class label C can be defined or approximated by two rough sets. The first contains elements that definitely belong to that class, called the lower approximation, whereas the other contains elements that possibly belong to class C, called the upper approximation (Wang *et al*., 2016b). The P-lower and upper approximations of X can be defined as:

$$\_PX = \{x[x]_P \subseteq X \quad (2)$$

$$\overline{P}X = \{x|[x]_P \bigcap X \neq \phi\} \quad (3)$$

The positive region of P of Q can be found as bellow if P and Q be equivalence relations over U:

$$POS_p(Q) = \bigcup_{x \in U_{/Q}} \underline{P}X \quad (4)$$

There are two main approaches for finding the rough set reduct subsets of features (i.e., the selected features). The first considers the degree of dependency and the second is concerned with the discernibility matrix. The degree of dependency of attributes set P with respect to class labels set Q can be defined as:

$$\gamma_p(Q) = \frac{|POS_p(Q)|}{|U|} \quad (5)$$

where, the discernibility matrix of a decision table D = (U, C d) is defined as:

$$c_{ij} = \{a \in C | a(x_i) \neq a(x_j)\} i, j = 1, ..., |U| \quad (6)$$

The attributes reduction process is done by comparing equivalence relations which generated by the attributes sets. A reduct R is defined as a subset of least cardinality of the conditional attribute set A such that $\gamma_R(C) = \gamma_A(C)$ where, C is the decision attribute (class).

Fuzzy rough set feature reduction uses a similar concept to the Crisp approach described above. However, in the fuzzy-based approach, fuzzy relations are used to define both the generalized upper and lower approximations (Jia *et al*., 2016).

In the neighborhood rough set method, the radius of the neighborhood (neighborhood mean) influences the reduction performance. This is because the classification granularity determines the number of training samples within the classification boundary region. In case of the size of the neighborhood relation is equal to zero, neighborhood rough sets are called as generalization rough set (Wang *et al*., 2016a, b). We used both the Crisp and fuzzy-based neighborhood rough set feature reduction algorithms and size of the neighborhood is set to "0.1".

Machine learning classifiers: A classifier is a machine learning approach that places data items into one of C classes based on previous knowledge. The major goal of a classification algorithm is to maximize the classification accuracy with instances that are not included in the training set (Witten *et al*., 2016).

Rough Set Classification (RSC) uses both of attribute reduction and rule generation method for generating intrusion detection decision models. RSC performs attribute reduction before generating classification rules. Attribute reduction can be done by finding the reduct which can be defined as a minimal subset of attributes that has same classification power as the original set of attributes. RSC generated classification models in a form of "IF-THEN" rules, based on reduct calculated earlier (Zhang *et al*., 2004). Rule induction method such an MODLEM (Stefanowski, 1998) which is a modified version of LEM2 rule induction algorithm can be used to generate the optimal rule set that used in the RSC method.

Based on sequential covering, MODLEM used generates a minimal set of decision rules for every decision class or its rough approximation which attempts to cover only all positive examples of the given decision class.

MODLEM rule induction algorithm starts by generating the first rule that satisfies the best conditions criteria such as class entropy measure or Laplacian accuracy. After adding the rule to the final rule list, then all learning positive examples that match this rule are removed from consideration. The process is repeated while some positive examples of the decision class remain still uncovered. Then, the procedure is sequentially repeated for each set of examples from the other decision classes (Stefanowski, 2007). The minimal set of rules generated by MODLEM algorithm is used as machine learning classifier.

In the next study, we will compare the performance of our proposed rough set classification approach against to four standard classifiers (SVM with a linear kernel (Mulay *et al*., 2010) Naive Bayes (NB) (Mukherjee and Sharma, 2012) kNN with a cosine-based distance and (k = 5) (Liao and Vemuri, 2002) and a decision tree (Mulay *et al*., 2010).

RESULTS AND DISCUSSION

Performance measure and system evaluation: To evaluate the performance of the built IDS decision model, we considered the most commonly used classification model evaluation metric in the literature which is the F-score. F-score represent the harmonic means between precision and recall and can be calculate as following (Dua and Du, 2016).

- Precision = TP/(TP+FP)
- Recall = TP/(TP+FN)
- F-score = 2×Precision×Recall/(Precision+Recall)

Table 1: The number and indexes of the features selected by both of Crisp and fuzzy rough set methods

Methods	No. of selected features	Selected feature indexes
Crisp rough set	7	[7, 5, 9, 4, 6, 10, 2]
Fuzzy rough set	6	[7, 9, 4, 10, 6, 2]

Table 2: The classification performance in term of F-score for each classification method with respect to the number of features selected by Crisp and fuzzy rough set methods

		Classifier				
Methods	No. selected features	RSC	KNN	SVM	NB	D-tree
Crisp rough set	7	94.2	88.3	82.9	78.2	88.0
Fuzzy rough set	6	93.7	88.0	82.1	76.8	87.5
Full dataset	10	88.9	86.2	82.0	74.8	86.1

Where True Positive (TP) refers to number of attacks that detected as attack, False Positive (FP) refers to number of attacks that detected as normal behavers, True Negative (TN) refers to number of normal behavers detected as normal behavers and False Negative (FN) refers to number of normal behavers detected as attack. The process of evaluation our system involves the following steps:

- Apply the feature selection method over the dataset
- Generate a new dataset with only the features selected by feature selection method
- Use k-fold evaluation method to divide the dataset into training and testing subset
- Use training subset to build a classification model using different classification methods
- Use the testing subset to evaluate the performance of the classification models with regards to F-score

We used both the Crisp and fuzzy-based rough set feature reduction algorithms to find the minimal attribute subset that provides the maximum classification performance. The indexes of the reduct (selected attributes) generated by using the Crisp and the fuzzy rough set feature reduction method are shown in Table 1 as:

The Crisp rough set attribute reduction generates a reduct set with seven attributes while fuzzy rough set method attained to generate a reduct with only 6 features out of total ten attributes.

The dataset with the attributes that redacted by using both Crisp and fuzzy rough set attribute reduction method is then used to build a classification model based on RSC method discussed in previous study as well as another four standard classification methods which are SVM with a linear kernel, Naive Bayes (NB) kNN with a cosine-based distance and (k = 5) and a decision tree.

The classification F-score with regard to each dataset with attributes selected by Crisp and fuzzy rough set methods for each classification method using 10-fold cross-validation is shown in Table 2 as:

Regarding to the F-score results obtained in Table 2, we can clearly see that the RSC classifier has the best performance in terms of accuracy over the other classifiers. RSC classifier obtained 94.8% F-score by generating 81 decision rules from the dataset with seven features, selected by Crisp rough set attribute reduction method. The other dataset with 6 features obtained by fuzzy rough set attribute reduction method generates 77 decision rules and provided 93.7% F-score for the RSC classifier. For the full dataset with 10 features, the RSC provided 88.9% F-score by generating 210 decision rules.

CONCLUSION

In this research, a host-based IDS architecture was developed using machine learning approaches. This architecture is based on the generation of new datasets with fewer attributes than the original ADFA-LD, using 10 different statistical analysis measures. The architecture also utilizes Crisp and fuzzy rough sets based feature reduction approaches to identify the minimum feature subset that provides the best classification performance (by removing the redundant and irrelevant attributes) while reducing the time required to classify new instances.

We also used Parzen window novelty detection approach to compare the performance of ADFA-LD dataset with its competitor NSL-KDD dataset and the results shows that the two datasets have a comparable performance in term of detecting the novel behaviors.

The Crisp rough sets based feature reduction approach provides a 30% reduction in the feature space by selecting 7 features out of 10 features from the generated new dataset. While by selecting only 6 features the fuzzy rough set approach provides a 40% reduction in the feature space.

Rough Set Classification (RSC) approach is considered for building the IDS decision model. RSC uses the datasets with the features that reduced by Crisp and fuzzy rough set approaches to generate classification rules using MODLEM rule induction algorithm.

The generated RSC Model also compared with four standard classification models (KNN, SVM, NB and D-tree) in term of classification F-score. The experiment results show that the IDS decision model built using RSC classification method has a superior performance in the regards to the classification accuracy, over the other four classification methods.

The experiment also shows the impact of the attribute reduction method on the classification accuracy. For instance, the RSC classification model obtained an F-score value of 88.9% over the full ten features while using the 6 features that selected by the fuzzy rough set method increase the F-score value by about 6%.

REFERENCES

Creech, G. and J. Hu, 2013. Generation of a new IDS test dataset: Time to retire the KDD collection. Proceedings of the 2013 IEEE International Conference on Wireless Communications and Networking (WCNC'13), April 7-10, 2013, IEEE, Shanghai, China, ISBN:978-1-4673-5938-2, pp: 4487-4492.

Creech, G. and J. Hu, 2014. A semantic approach to host-based intrusion detection systems using contiguousand discontiguous system call patterns. IEEE. Trans. Comput., 63: 807-819.

Dash, M. and H. Liu, 1997. Feature selection for classification. Intell. Data Anal., 1: 131-156.

Dua, S. and X. Du, 2016. Data Mining and Machine Learning in Cybersecurity. CRC Press, Boca Raton, Florida, USA., ISBN-13:978-1-4398-3943-0, Pages: 224.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett., 27: 861-874.

Garcia-Teodoro, P., J. Diaz-Verdejo, G. Macia-Fernandez and E. Vazquez, 2009. Anomaly-based network intrusion detection: Techniques, systems and challenges. Comput. Secur., 28: 18-28.

Jia, X., L. Shang, B. Zhou and Y. Yao, 2016. Generalized attribute reduct in rough set theory. Knowl. Based Syst., 91: 204-218.

Khreich, W., B. Khosravifar, A. Hamou-Lhadj and C. Talhi, 2017. An anomaly detection system based on variable N-gram features and one-class SVM. Inf. Software Technol., 91: 186-197.

Lee, W., S.J. Stolfo and K.W. Mok, 1998. Mining Audit Data to Build Intrusion Detection Models. Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD'98), August 27-31, 1998, Association for the Advancement of Artificial Intelligence (AAAI), Palo Alto, California, USA., pp: 66-72.

Li, H., D. Li, Y. Zhai, S. Wang and J. Zhang, 2016. A novel attribute reduction approach for multi-label data based on rough set theory. Inf. Sci., 367: 827-847.

Liao, Y. and V.R. Vemuri, 2002. Use of k-nearest neighbor classifier for intrusion detection. Comput. Secur., 21: 439-448.

Miljkovic, D., 2010. Review of novelty detection methods. Proceedings of the 33rd International Conference on Convention MIPRO, May 24-28, 2010, IEEE, Opatija, Croatia, ISBN:978-1-4244-7763-0, pp: 593-598.

Mukherjee, S. and N. Sharma, 2012. Intrusion detection using naive Bayes classifier with feature reduction. Procedia Technol., 4: 119-128.

Mulay, S.A., P.R. Devale and G.V. Garje, 2010. Intrusion detection system using support vector machine and decision tree. Intl. J. Comput. Appl., 3: 40-43.

Pawlak, Z., 2012. Rough Sets: Theoretical Aspects of Reasoning about Data. Springer, Berlin, Germany, ISBN:978-94-010-5564-2, Pages: 229.

Stefanowski, J., 1998. The rough set based rule induction technique for classification problems. Proceedings of 6th European Conference on Intelligent Techniques and Soft Computing EUFIT Vol. 98, September 7-10, 1998, Aachen Publisher, Aachen, Germany, pp: 109-113.

Stefanowski, J., 2007. On Combined Classifiers, Rule Induction and Rough Sets. In: Transactions on Rough Sets VI., Peters, J.F., A. Skowron, I. Duntsch, J. Grzymala-Busse and E. Orlowska et al. (Eds.). Springer, Berlin, Germany, ISBN:978-3-540-71198-8, pp: 329-350.

Tavallaee, M., E. Bagheri, W. Lu and A.A. Ghorbani, 2009. A detailed analysis of the KDD CUP 99 data set. Proceedings of the IEEE International Symposium on Computational Intelligence for Security and Defense Applications (CISDA'09), July 8-10, 2009, IEEE, Ottawa, Canada, ISBN:978-1-4244-3763-4, pp: 1-6.

Vijayanand, R., D. Devaraj and B. Kannapiran, 2017. Support vector machine based intrusion detection system with reduced input features for advanced metering infrastructure of smart grid. Proceedings of the 4th International Conference on Advanced Computing and Communication Systems (ICACCS), January 6-7, 2017, IEEE, Coimbatore, India, ISBN:978-1-5090-4559-4, pp: 1-7.

Wang, C., M. Shao, Q. He, Y. Qian and Y. Qi, 2016a. Feature subset selection based on fuzzy neighborhood rough sets. Knowl. Based Syst., 111: 173-179.

Wang, W., Z. Yang and M. Zhang, 2016b. Intrusion Detection Technology based on Rough Set Attribute Reduction Theory. In: Human Centered Computing, Zu, Q. and B. Hu (Eds.). Springer, Cham, Switzerland, ISBN:978-3-319-31853-0, pp: 779-786.

Warrender, C., S. Forrest and B. Pearlmutter, 1999. Detecting intrusions using system calls: Alternative data models. Proceedings of the 1999 IEEE International Symposium on Security and Privacy, May 14, 1999, IEEE, Oakland, California, USA., pp: 133-145.

Witten, I.H., E. Frank, M.A. Hall and C.J. Pal, 2016. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers Company, New York, USA.,.

Xie, M. and J. Hu, 2013. Evaluating host-based anomaly detection systems: A preliminary analysis of ADFA-LD. Proceedings of the 6th International Congress on Image and Signal Processing (CISP) Vol. 3, December 16-18, 2013, IEEE, Hangzhou, China, ISBN:978-1-4799-2764-7, pp: 1711-1716.

Xie, M., J. Hu and J. Slay, 2014c. Evaluating host-based anomaly detection systems: Application of the one-class SVM algorithm to ADFA-LD. Proceedings of the 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), August, 19-21, 2014, IEEE, Xiamen, China, ISBN:978-1-4799-5148-2, pp: 978-982.

Xie, M., J. Hu, X. Yu and E. Chang, 2014a. Evaluating host-based anomaly detection systems: Application of the frequency-based algorithms to ADFA-LD. In: Proceedings of the International Conference on Network and System Security, Au, M.H., B. Carminati and C.C.J. Kuo (Eds.). Springer, Cham, Switzerland, ISBN:978-3-319-11697-6, pp: 542-549.

Zhang, L.H., G.H. Zhang, L. Yu, J. Zhang and Y.C. Bai, 2004. Intrusion detection using rough set classification. J. Zhejiang Univ. Sci. A., 5: 1076-1086.