Hindawi
Journal of Computer Networks and Communications
Volume 2021, Article ID 8845540, 9 pages
https://doi.org/10.1155/2021/8845540

Research Article

Towards a Scalable and Adaptive Learning Approach for Network Intrusion Detection

Alebachew Chiche[1,2] **and Million Meshesha**[1,2]

[1]*Department of Information Systems, College of Computing, Debre Berhan University, Debre Birhan, Ethiopia*
[2]*School of Information Science, Addis Ababa University, Addis Ababa, Ethiopia*

Correspondence should be addressed to Alebachew Chiche; alebachew.chz@dbu.edu.et

Received 10 April 2020; Revised 10 November 2020; Accepted 28 December 2020; Published 19 January 2021

Academic Editor: Rui Zhang

This paper introduces a new integrated learning approach towards developing a new network intrusion detection model that is scalable and adaptive nature of learning. The approach can improve the existing trends and difficulties in intrusion detection. An integrated approach of machine learning with knowledge-based system is proposed for intrusion detection. While machine learning algorithm is used to construct a classifier model, knowledge-based system makes the model scalable and adaptive. It is empirically tested with NSL-KDD dataset of 40,558 total instances, by using ten-fold cross validation. Experimental result shows that 99.91% performance is registered after connection. Interestingly, significant knowledge rich learning for intrusion detection differs as a fundamental feature of intrusion detection and prevention techniques. Therefore, security experts are recommended to integrate intrusion detection in their network and computer systems, not only for well-being of their computer systems but also for the sake of improving their working process.

1. Introduction

Nowadays, network-based computer systems are becoming a common place for modern society, because of this a network intruder has focused on them. Therefore, we need to have a new protection approach for computer networks. From different literatures, we understood that the concept of intrusion detection system (IDS) was introduced by Anderson for the first time in 1980 [1] and later dignified by Dorothy Denning [2].

According to Farid et al. [3], in the beginning, host-based intrusion detection system (HIDS) was implemented as intrusion detection system that had located in different end systems, but the attention of researchers has been gradually shifted towards a network-based intrusion detection system as the use of network systems grow rapidly.

As illustrated in different scientific works, both internal and external attacks are increasing in institutions with the fast growth of Internet and network services [3]. According to Heady et al. [4], an intrusion is defined as a kind of attempt that tries to negatively affect the normal functioning of network and computer systems such as illegal use of super user account for gaining access, repudiating services on computer systems.

Talwar and Goyal [5] defined intrusion detection system as "a phenomenon or device that analyses system and network activity for unauthorized activity." As defined by Talwar and Goyal, intrusion detection system is any process or software that monitors a system or network of systems against any intrusion activity. So, an attempt to catch the abnormal action before they do damage on the computer system is the final goal of any intrusion detection system. So, an IDS safeguards a system from attack, misuse, and any nasty activity.

There are mainly two types of intrusion detection techniques based on the approach followed for detecting network intrusion: signature and anomaly-based intrusion detection model [6, 7]. On the basis of a computer system, anomaly-based intrusion detection approach identifies abnormal behavior of the network traffic by creating baseline

on the normal behavior of network traffics. The signature-based intrusion detection approach uses a knowledge base that stores a signature of known network intrusions and performs a comparison of knowledge base with incoming network traffics to identify known attack only.

On the other side, intrusion detection can be categorized as network and host-based intrusion detection on the basis of intrusion audit data analysis [7]. Network-based intrusion detection system (NIDS) monitors the network traffic that crosses the entire network. To make effective detection of network intrusions, it should have the ability of standing against large amount of network traffic. It must collect all the network traffic and analyze it quickly while the volume of network traffic increases. Host-based intrusion detection system (HIDS), on the other hand, collects record of the audit data that is tracked within the single host.

Classical intrusion prevention systems such as information protection using encryption and authentication have been used as a first line of defense. There are exploitable weaknesses in every system because of the complexity of systems, configuration, and design errors [8]. Moreover, most intrusion detection systems have been constructed and implemented based on the knowledge and understanding of the systems designer and developers about known intrusions. Thus, the successes of the intrusion detection systems are limited to the novel attacks.

Previous research suggests that intelligence is very likely to help security experts to detect and prevent them easily [7]. An extensive reading of various literatures from different leading electronic journal databases suggests that no academic research has examined how to make training data scalable, how to train the machine learning algorithms adaptively from past experience, and how to provide corrective actions for detected attack. Existing research has addressed several aspects of intrusion detection, such as modeling intrusion detection using machine learning techniques [9–11], optimal attribute selection and classification [12], and adaptive intrusion detection [8, 13, 14]. But research on intrusion detection has concentrated only on constructing a predictive model using machine learning algorithms with a static data. Considering the issues of a detection model with scalable and adaptive learning features in particular, the literature is almost silent on the details of investigating intrusion detection systems with a scalable data, adaptive learning, and knowledge base collaboratively. Thus, there is no complete picture of the way adaptive and scalable intrusion detection systems are developed.

Although extensive research has explored the characteristics and dynamics of intrusion detection systems using different methods and techniques [3, 7, 11, 15], much less research has investigated intrusion detection system with a scalable data, classifier pattern, and adaptive learning approach. The explosive growth of network-based economy conveys the need for the research that extends the traditional intrusion detection trained on a stationary data for constructing detection model.

Several network intrusion detection systems have been built manually [16]. So, these systems have been dependent on the understanding and knowledge of the experts who designed them. Consequently, the performance of previous intrusion detection systems depends on the knowledge and skills of those experts about the computer systems and characteristics of network intrusions. They are also limited in identifying novel attacks that come from different network environment. So far, scalability has received relatively little attention in intrusion detection research and can be broadly categorized in terms of network and temporal and traffic scalability [17].

The main purpose of this research is therefore to approach a model for analyzing large volume of data to get hidden patterns, constructing a scalable and adaptive classifier for intrusion detection; that is, the study explored the effect of combining machine learning- and knowledge-based systems to address the problem of static data and detection model for network intrusion detection. In this approach, classifiers are inductively trained on the selected attributes using the prepared and preprocessed training data. So, the classifiers can construct a network intrusion detection model for identifying whether the given instance is "normal" or "abnormal." In the meantime, the knowledge-based system plays a vital role in updating the predicted instances to original training data and suggesting corrective action for predicted attack. Because previous researches only apply machine learning algorithms on the given data to come up with a predictive model, this approach has no way to update training data and predictive model as well. This approach is significantly different from the traditional signature-based approaches. Due to this, previous works are not scalable and adaptive in their learning approach on a given data. Moreover, previous works have no capability to use new audit data tried on the model for next learning. So, the model can update the pattern based on the updated dataset. This research is experimented on the offline data collected from NSL-KDD [18] intrusion detection dataset.

The rest of the article is structured as follows. Section 2 provides reviews of related works. Section 3 presents methods and algorithms with experimental analysis. Section 4 shows NSL-KDD intrusion dataset. Sections 5 and 6 describe the methodology, result, and discussion of scalable and adaptive learning approach. Finally, Section 7 provides concluding remarks of the work.

2. Related Works

Though researchers have contributed and lay down a base towards developing an intrusion detection system using different techniques, much of the previous work in network intrusion detection focuses on constructing a predictive model [19] for detecting the network traffic either as normal or abnormal. Omar et al. [13] proposed a hybrid machine learning model by combining the unsupervised and supervised classification algorithms for intrusion detection which uses a combination of K-means, fuzzy C-means, and GSA clustering algorithms to obtain similar patterns of a user's activity. Then, a combination of support vector machine and gravitational search algorithm are implemented as a hybrid classification to improve the detection accuracy of the proposed method. Farid et al. [14, 16] proposed an

adaptive learning approach for network intrusion detection which calculates and identifies best attributes from dataset using a combined ID3 decision tree algorithm and naïve Bayesian classifier. Their experimental results showed that the proposed approach achieves high classification accuracy and reduces false positive rate using KDD99 benchmark intrusion detection dataset. Ye and Li [20] presented a slightly different approach called scalable clustering technique for intrusion signature recognition. In this paper, a combination of supervised machine learning algorithms, namely, clustering and classification algorithms, were implemented for predicting network intrusions. Xu and Shelton [21] presented a general intrusion detection technique for both host-based and network-based intrusion detection systems. The paper presents a hierarchical CTBN model for the network packet traces which was constructed and used Rao-Blackwellized particle filtering to learn the parameters. At the same time, they developed a novel learning method to deal with the finite resolution of system log file time stamps for host-based intrusion detection system.

In literature, we understood that using different machine learning techniques, a number of intrusion detection systems are developed. For instance, some research studies apply single learning techniques, such as self-organizing map [22], neural networks [23], genetic algorithms [24], decision tree [25, 26], and pattern matching algorithms [27] to develop intrusion detection model. On the other hand, some intrusion detection systems such as hybrid approach or ensemble techniques are [28] developed by combining different machine learning techniques and ensemble classifiers by combining multiple weak learners [29]. These all techniques aforementioned are constructed as a predictive model in particular, tangibly to detect or classify whether an incoming network traffic is intrusion or normal access. However, there is no attempt to design scalable and adaptive learning approach for intrusion detection.

3. NSL-KDD Dataset

Nowadays, knowledge discovery in database (KDD) is a standard intrusion dataset used in various intrusions detection design and implementation research works for mitigating network intrusions [30]. NSL-KDD dataset is an improvement of KDD'99 with a fundamental change made to solve the doubt and problems found in previous KDD'99; however, still there is a problem in the new version of KDD but with great advantages over KDD'99. As stated by Talwar and Goyal [5], this version of the dataset has been more applicable for real networks as well. As claimed by Aggarwal and Sharma [30], the new version of KDD dataset (NSL-KDD) modified and developed from the fundamental problem existed in the old KDD'99 benchmark intrusion dataset. The problem of redundancy and missing values existed in KDD'99 are alleviated in new NSL-KDD benchmark intrusion dataset. In this empirical study, NSL-KDD intrusion dataset which is similar with KDD'99 dataset with 42 attributes is used.

As stated in various literatures [5, 30–32] similar to KDD'99 dataset, the classes in NSL-KDD are categorized into five main classes, namely, 4 main intrusion classes (such as DOS, probe, R2L, and U2R) and 1 normal class.

(1) Denial of service (DoS) attack which blocks legitimate user requests unreasonably depletes the computing resource such as power or memory of a victim machine to make it too busy or too full to handle legitimate requests.

(2) Remote-to-user (R2L) is an attack that unauthorized access from a remote machine to a local account by sending a kind of packets to gain a local access of a victim machine through a network.

(3) User-to-root (U2R) is an attack that an intruder uses a normal login account and tries to gain an administrator account by using a vulnerability of the victim system.

(4) Probing (probe) is an attack that scans and gathers information from remote victim machine through network with the objective to gain information and find the vulnerabilities for exploits.

So, for this work, we downloaded the NSL-KDD dataset in CSV format and converted it to ARFF format for experimental analysis. Following preprocessing step, the data are cleaned to get correct input to feed to classification algorithm so as to construct a predictive model.

4. Methodology

In this paper, a scalable and adaptive learning network intrusion detection system is presented. The system is designed by integrating machine learning model with knowledge base. Approach and procedures followed in this study are described as follows.

4.1. A Scalable and Adaptive Learning Approach. Literature on constructing a predictive model for network intrusion detection (NID) is rich, but the state-of-the-art NID does not cover the scalability and adaptive characteristics of the intrusion detection model. Scalability is becoming increasingly required for today's network intrusion detection [17]. This is because of the rapid growth of the large volumes of modern network traffic that needs fast monitoring with a continually changing attack activity. In the meantime, the new approach adjusts and adapts itself with the newly updated network connections. Accordingly, the machine learning automatically learns the new problem when there is change in network connection properties.

The implementation for the proposed approach is conducted with the help of Prolog programming language and WEKA 3.8 machine learning tool, and WEKA library functions are used for feature selection and classifier construction techniques.

The proposed approach for scalable and adaptive network intrusion detection is presented in Figure 1. It consists of two major modules. Therefore, in this section, we tried to discuss the details of the proposed approaches.

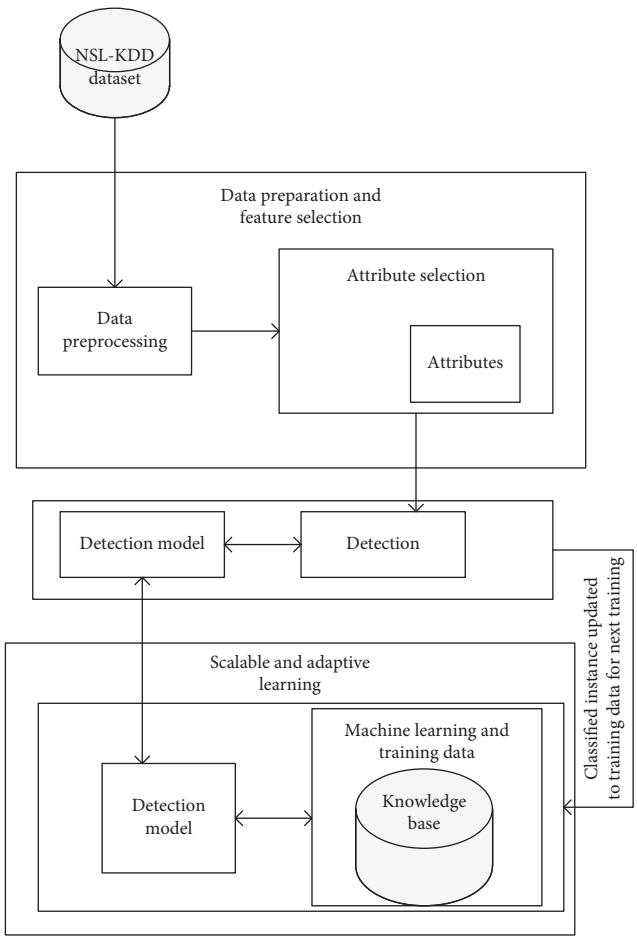

Figure 1: Scalable and adaptive learning approach for NID.

Figure 1 shows the architecture of new network intrusion detection model representing its main modules and subsystems. As described in Figure 1, the new approach is the constitute of two major subsystems: the supervised learning (SL), connector, and the knowledge-based system (KBS). In fact, the learning subsystem is a collective result of database, pattern extraction, and update detection modules. The learning subsystem is mainly responsible for learning from the dataset incrementally and adaptively using machine learning algorithm. On the other hand, the knowledge-based system represents the machine learning result to detect the type of incoming network connection, and it automatically updates the new network connection as an instance in original training dataset.

To implement the above proposed approach (see Figure 1), we design the algorithm depicted in Algorithm 1 that incorporates machine learning and knowledge base for detecting network intrusion. For the experiment, NSL-KDD dataset is downloaded from "KDD Cup 1999 Data," http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (accessed on March 12, 2018).

So, the following tasks are performed to develop a scalable and adaptive intrusion detection system using an integrated approach to develop machine learning-based knowledge-based system.

4.2. Data Preprocessing Section. As stated by Aggarwal and Sharma [30], NSL-KDD benchmark intrusion detection dataset is a refined version of KDD′99 in which there are 494021 instances in the 10% training dataset. In NSL-KDD intrusion dataset, four classes of attacks are incorporated, such as remote-to-user (R2L), user-to-root (U2R), denial of service (DoS), and probe in which 22 different attacks are included specifically. In NSL-KDD dataset, 41 total attributes are identified and incorporated. For the dataset to be suitable for experimentation using machine learning algorithms, the data need to undergo data preprocessing step, where data cleaning, data size balancing, data size reduction, and dimensionality reduction (feature reduction) are performed.

During data cleaning activities such as handling missing values, avoiding duplications and handling outliers are performed. Moreover, sampling and feature selection techniques are applied on the NSL-KDD intrusion dataset to produce manageable NSL-KDD dataset appropriate for the experiment. Finally, based on the aforementioned activities such as sampling methods, a total of 40,558 instances are prepared for the experiment.

4.3. Attribute Selection. In constructing high performance intrusion detection systems, one of the important research challenges is effective attributes selection from intrusion detection datasets. Accuracy of intrusion detection model has been greatly affected by the presence of irrelevant and redundant attributes in the intrusion detection dataset. As described by Lee et al. [8], 41 attributes were constructed for each network connection on NSL-KDD intrusion detection dataset. To filter best attributes used in constructing intrusion detection model to identify abnormal network connections from a given dataset, attribute selection methods have been applied.

By applying forward attribute selection, the best attributes among the candidate subsets have been identified. Therefore, building any intrusion detection system based on all attributes is not cost effective and requires more computational resources [33–35]. Hence, it becomes very important to strategically sample data that may work well for intrusion detection system. In view of that, best performing 19 representative sample attributes (see Table 1) have been selected from a total of 41 attributes in NSL-KDD benchmark intrusion dataset. Therefore, the performance of intrusion detection systems can be improved by using attribute selection methods.

4.4. Classification Modeling. According to Neethu [16], constructing classification model is one of the main challenges for intrusion detection system, which is to construct effective models to identify normal behaviors from abnormal behaviors of network connection by observing collected audit data. In addition, one of the main challenges in intrusion detection systems is learning from static intrusion data to construct a classifier. To solve this problem, various researchers are working continually but former intrusion detection systems are analyzed and constructed by human

```
Input: original training dataset D
Output: classification instance as attack or normal
Use features selection and extract best features
Train machine learning algorithms ML, where ML is machine learning
Select best classifiers such as random forest (RF)
Incorporate RF with D as KB, where KB is the knowledge base
While (new instance == true)
  {
    Apply classifier RF,
    Get class of instance I, as attack or normal, where I is the classified instance
      For (I == true)
      {
        KB fetch classified instance I, where KB consists ML and D
        string comp=compare I with D
        If (comp is not true)
          {
            new instance is not added to D, where D is training dataset
            training dataset not updated
            }
      Else
              {
            new instance is not added to D, where D is training dataset
            training dataset is updated and ready for next training
            New pattern P is generated
            Applied for next classification
          }
      }
  }
```

Algorithm 1: Scalable and adaptive learning approach for network intrusion detection.

experts manually on network audit data [8, 16]. Analyzing and drawing intrusion detection rules from a large and growing volume of audit data using human security experts are very tedious and boring. Also, it may be possible to identify known attacks by using human experts, but it is totally difficult for human security experts to identify novel attacks from dynamic and large size of intrusion data [16].

Nowadays, many machine learning algorithms have become very common and attracted more and more interests in recent years for classifying network connections into normal and abnormal [16]. Some of the popular machine learning algorithms used for classifying a given intrusion audit data include decision tree, support vector machine, neural network, genetic algorithm, Naïve Bayesian, and Fuzzy logic. Since the attackers and behavior of network attacks are becoming complicated and continuously changing their way of attacking and patterns, it is very difficult to detect several new attacks that come through the network. Therefore, Neethu [16] acclaims that machine learning algorithms applied in different intrusion detection researches need an improvement in their classification accuracy.

In this paper, more algorithms are experimented with a NSL-KDD intrusion dataset for intrusion detection to get the best classifier. Accordingly, Bayes Net, random forest, and SMO classification algorithms were experimented to construct and select the best classifier for the next steps. The algorithms are experimented using 40,558 instances with 19 attributes. 10-fold cross-validation is selected as test mode for classification. To evaluate the algorithms, various performance measures were used. The results of the experiments are compared with different evaluation criteria. The comparison results are given in Table 2.

The comparative analysis of the classifiers in Table 2 shows that the random forest classifier registered the best performance of 99.91 %, 99.9%, and 0.1% classification accuracy, true positive rate, and false positive rate, respectively. The SMO came second best with 98.12 %, 99.9, and 3.8% classification accuracy, true positive rate, and false positive rate, respectively. Bayes Net however came out last with 97.57%, 99%, and 3.5% classification accuracy, true positive rate, and false positive rate. Empirical result therefore shows that the random forest gives better performance for detecting attacks than the two other classifiers.

The performance of the random forest (RF) is illustrated in Tables 1 and 2. Random forest (RF) works better than both SMO and Bayes Net for normal, DoS, probe, and R2L classes. For R2L classes, it performed less than Bayes and SMO classifiers.

Our experiments show that the random forest (RF) gives better accuracy for normal, DoS, probe, and R2L classes compared to SMO and Bayes Net and it gives the worst accuracy for detecting U2R class of attacks. For U2R class, both SMO and Bayes Net methods give the same performance. There is only a small difference in the accuracy for

Table 1: The list of selected attributes.

SNO	Attributes	Data type	Description
1	num_failed_logins	Continuous	Number of failed login attempts
2	logged_in	Discrete	1 if successfully logged in, 0 otherwise
3	Urgent	Continuous	Number of urgent packets
4	dst_bytes	Continuous	No. of data bytes from destination to source
5	root_shell	Discrete	1 if root shell is received, 0 otherwise
6	dst_host_srv_diff_host_rate	Continuous	% of connections to different destination machines, among the connections aggregated in dst_host_srv_count
7	Service	Discrete	Network service on destination like http and telnet
8	serror_rate	Continuous	% of connection with SYN errors
9	srv_serror_rate	Continuous	% of same connection with SYN errors
10	same_srv_rate	Continuous	% of connection with same services
11	rerror_rate	Continuous	% of connection with REJ errors
12	Count	Continuous	No. of cons to same host as the current con in past 2 sec
13	protocol_type	Discrete	Type of protocol like tcp and udp
14	num_file_creations	Continuous	No. of file creations
15	srv_diff_host_rate	Continuous	% of con to diff. host
16	Duration	Continuous	Length of connections in seconds
17	is_guest_login	Discrete	1 if guest is logged in, 0 otherwise
18	wrong_fragment	Continuous	No. of wrong fragments
19	is_host_login	Discrete	1 if host is logged in, 0 otherwise

DoS classes for SMO and Bayes Net but there is a significant difference for probe classes. Since U2R and R2L classes have small training data compared to other classes, it seems that SMO and Bayes Net classifiers give good accuracy with small training datasets. The R2L class for RF is better for the RF compared to both SMO and Bayes Net.

As evident from Tables 2 and 3, all the classifiers considered so far could not perform well for detecting all the attacks. To take advantage of the performance of the three classifiers, a random forest (RF) is selected for next integration with knowledge base to come up with a scalable and adaptive learning approach for intrusion detection.

Based on experimental results depicted in Table 2, random forest classifier gives better performance with a prediction accuracy of 99.91% for detecting attacks than SMO and Bayes Net algorithms. After all, random forest (RF) classifier is selected as a best performed classifier to integrate with knowledge base. From the empirical results, we can understand that random forest has scored better performance in terms of classification accuracy in which RF is better in simple and linearity structure of the dataset.

5. Connecting Supervised Learning with Knowledge Base

To connect machine learning- and knowledge-based system, different programming languages, libraries, and tools are applied. So, WEKA class libraries, SWI_WEKA package, Java Prolog Interface library (JPL), Java programming language, and Prolog logical programming have been used in the integration process. To come with scalable and adaptive intrusion detection model, the following modules have been implemented. All modules in the approach are affected when new network connection was initiated. The modules are described as follows:

We further analyze confusion matrix to assess the effectiveness of the proposed approach. Table 4 represents the confusion matrix for the proposed model.

According to the result in Table 4, the confusion matrix shows our scalable and adaptive detection model can perform well on all classes. From the confusion matrix, one can understand that random forest classifies 40,521 instances out of 40,558 correctly and 27 instances incorrectly. So, the confusion matrix shows that random forest achieves better with 99.91% of classification accuracy.

6. Discussion of Result

This study investigates that scalable and adaptive learning approach for intrusion detection is possible through combination of machine learning- and knowledge-based system. To our knowledge, this study is the first study which gives practical demonstration on the possibilities of scalable and adaptive learning approach for improving intrusion detection. The average accuracy of the three algorithms across the datasets is shown in Figure 2. Firstly, as presented in Figure 2, SMO, Bayes Net, and random forest classifiers have the best average accuracy, i.e., 98.12%, 97.57%, and 99.91%, respectively, when using supervised learning. According to the experiment result shown in Table 2, our approach achieves a better prediction accuracy for all classes of network connection categories. In the meantime, the performance of the classifiers gets an acceptable TP, precision, and sensitivity ratio as well. These results further prove that among the three classifiers, random forest has a weighted comparative advantage over others for intrusion detection. In general, better detection accuracy has been registered on the NSL-KDD datasets; on the average, 99.91% accuracy is obtained. This is because the linearity and qualities of dataset were reasonably good. But we faced problem in the approach, which is latency

Table 2: Performance comparison of algorithms with different evaluation metrics.

Performance metrics	SMO (%)	Bayes Net (%)	Random forest (%)
TP rate	98.1	97.6	99.9
FP rate	2.0	2.2	0.1
Precision	98.2	97.6	99.9
Recall	98.1	97.6	99.9
F-measure	96.4	97.6	99.9
Accuracy	98.1	97.6	99.91

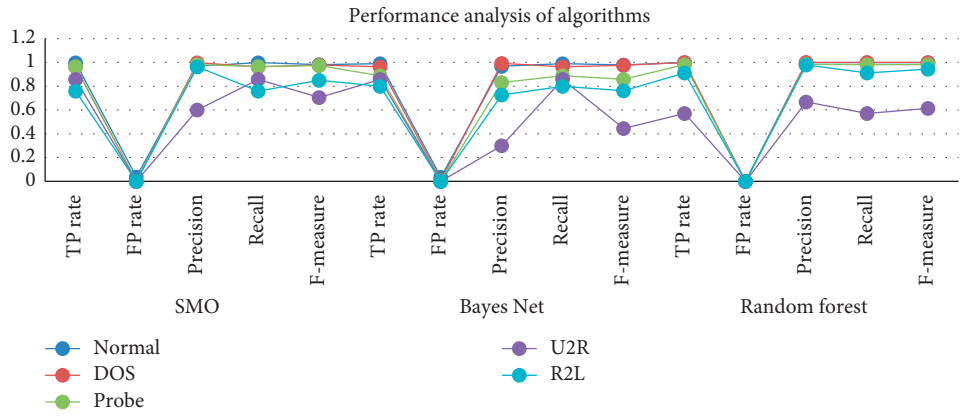

Table 3: Performance comparison of the three classifiers.

Attack type	SMO (%)	Bayes Net (%)	Random forest (%)
Normal	99.90	98.98	99.94
DoS	96.2	96.3	99.98
Probe	96.6	88.66	98.08
U2R	85.7	85.7	85.14
R2L	75.96	79.8	91.34

Table 4: Confusion matrix for random forest.

	Predicted classes					
	Normal	DoS	Probe	U2R	R2L	—
Actual classes	21355	3	4	1	2	Normal
	7	18464	1	0	0	DoS
	3	3	614	0	0	Probe
	3	0	0	4	0	U2R
	8	0	0	2	84	R2L

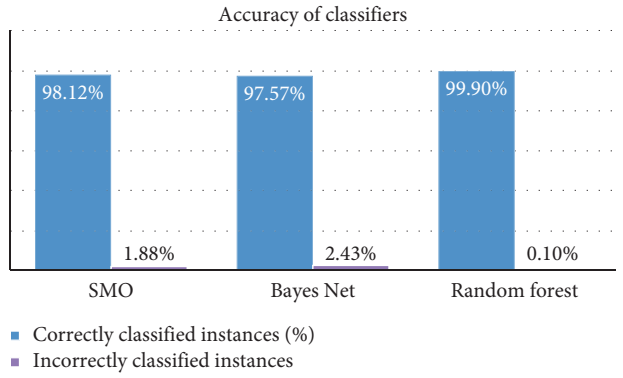

Figure 2: Classification accuracy of classifiers.

after connecting the machine learning result with the knowledge-based system. After connecting the machine learning result and knowledge-based system, the time taken to detect the network connection is slightly reduced. The other challenge we faced in this work is unavailability of instant data to test the approach. So, the approach is tested on offline data publicly available online. Generally, the study empirically proof the possibility of incorporating machine learning- and knowledge-based system for the sake of developing scalable and adaptive learning approach for intrusion detection at the same time. We observed that machine learning- and knowledge-based systems are essential to each other. So, our experiment result shows that after integration of machine learning and knowledge base, 99.89% classification accuracy is achieved on the preprocessed NSL-KDD intrusion dataset.

7. Conclusion

While the impact of intrusion detection and prevention techniques has typically been studied in terms of the benefit it brings to organization to protect their systems from network intrusions, various previous studies have been studied to implement intrusion detection system. This study provides knowledge on implementing scalable and adaptive intrusion detection that is not emphasized by former researchers. As we presented, what makes our system scalable and adaptive is that whenever the dataset is updated, the system is also automatically updating the model such that the new pattern is also taken into account during next prediction.

The empirical result shows that the proposed approach achieves 99.91% classification accuracy using random forest machine learning algorithm as classifier, and the modified version of intrusion dataset, that is, NSL-KDD was suitable for the experiment. Consequently, this work reveals a new benefit of combining machine learning and knowledge base for implementing intrusion detection system, strengthening the security of organizational computer systems in that intrusion detection system is becoming an important practice for organizational success. Thus, such kinds of security appliance should be added to the network infrastructure of organizations to improve organizational work flow and performance and to secure their computer systems. Improving an efficiency of the approach is one of our future works, parallelly improving the machine learning algorithms towards the detection of other instant datasets. This needs instantly capturing any network connection data for extracting patterns and knowledge for updating the knowledge-based system.

Data Availability

The dataset used in this work is publicly available as a benchmark for research purposes, https://www.unb.ca/cic/datasets/nsl.html. So, the preprocessed data obtained to support the findings of this work are available from the authors upon request. All the supporting open-source codes for integration activities are available to the research community under an open-source license for the researchers.

Conflicts of Interest

The authors hereby declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] J. P. Anderson, *Computer Security Threat Monitoring and Surveillance*, James P. Anderson Company, Fort Washington, MD, USA, 1980.

[2] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, vol. 13, no. 2, pp. 222–232, 1987.

[3] D. M. Farid, J. Darmont, N. Harbi, H. N. Huu, and M. Z. Rahman, "Adaptive network intrusion detection learning: attribute selection and classification," *International Journal of Computer and Information Engineering*, vol. 3, no. 12, pp. 2762–2766, 2009.

[4] R. Heady, G. F. Luger, A. B. Maccable, and M. Servilla, "The architecture of a network level intrusion detection system," *Osti.Gov*, 1990.

[5] S. Talwar and D. Goyal, "Data mining based classification technique for adaptive intrusion detection system using machine learning," *International Journal of Advances in Engineering Sciences*, vol. 5, no. 3, pp. 16–19, 2015.

[6] H. P. S. Sasan and M. Sharma, "Intrusion detection using feature selection and machine learning algorithm with misuse detection," *International Journal of Computer Science and Information Technology*, vol. 8, no. 1, pp. 17–25, 2016.

[7] T. Dagne, "*Constructing predictive model for network intrusion detection: network intrusion detection model*," M.S. thesis, Addis Ababa University, Addis Ababa, Ethiopia, 2012.

[8] W. Lee, S. J. Stolfo, and K. W. Mok, "Adaptive intrusion detection: a data mining approach," *Artificial Intelligence Review*, vol. 14, no. 6, pp. 533–567, 2000.

[9] S. Sivaranjani, "Network intrusion detection using data mining technique," *International Journal of Advanced Research in Computer Engineering & Technology*, vol. 3, no. 6, pp. 2219–2224, 2018.

[10] A. Chalak, "Data mining techniques for intrusion detection and prevention system," *International Journal of Computer Science and Network Security*, vol. 11, no. 8, pp. 200–203, 2011.

[11] G. V. Nadiammai and M. Hemalatha, "Effective approach toward intrusion detection system using data mining techniques," *Egyptian Informatics Journal*, vol. 15, no. 1, pp. 37–50, 2014.

[12] N. Gupta, N. Singh, V. Sharma, T. Sharma, and A. S. Bhandari, "Feature selection and classification of intrusion detection system using rough set," *International Journal of Communication Network Security*, vol. 2, no. 2, pp. 20–23, 2013.

[13] S. Omar, H. H. Jebur, and S. Benqdara, "An adaptive intrusion detection model based on machine learning techniques," *International Journal of Computer Applications*, vol. 70, no. 7, pp. 1–5, 2017.

[14] D. M. Farid, H. Nouria, and M. Z. Rahman, "Combining naive Bayes and decision tree for adaptive intrusion detection," *International Journal of Network Security & Its Applications*, vol. 2, no. 2, pp. 12–25, 2010.

[15] N. Farnaaz and M. A. Jabba, "Random forest modeling for network intrusion detection system," in *Proceedings of the Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)*, Hyderabad, India, January 2016.

[16] B. Neethu, "Adaptive intrusion detection using machine learning," *IJCSNS International Journal of Computer Science and Network Security*, vol. 13, no. 3, pp. 118–124, 2013.
[17] S. A. Shaikh, H. Chivers, P. Nobles, J. A. Clark, and H. Chen, "Towards scalable intrusion detection," *Network Security*, vol. 2009, no. 6, pp. 12–16, 2009.
[18] University of New Brunswick, *NSL-KDD Dataset*, University of New Brunswick, Fredericton, Canada, 2018, https://github.com/defcom17/NSL_KDD.
[19] KDD Cup 1999. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.
[20] N. Ye and X. Li, "A scalable clustering technique for intrusion signature recognition," in *Proceedings of the 2001 IEEE Workshop on Information Assurance and Security*, West Point, NY, USA, June 2001.
[21] J. Xu and C. R. Shelton, "Intrusion detection using continuous time bayesian networks," *Journal of Artificial Intelligence Researchficial Intelligence Research*, vol. 39, pp. 745–774, 2010.
[22] T. Y. Christyawan, A. A. Supianto, and W. F. Mahmudy, "Anomaly-based intrusion detector system using restricted growing self organizing map," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 3, pp. 919–926, 2019.
[23] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, New Jersey, NJ, USA, 2nd edition, 1999.
[24] H. Suhaimi, S. Izwan Suliman, I. Musirin et al., "Network intrusion detection system using immune-genetic algorithm (IGA)," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 2, pp. 1060–1065, 2019.
[25] T. Mitchell, *Machine Learning*, McGraw-Hill, New York, NY, USA, 1997.
[26] L. Breiman, *Classification and Regressing Trees*, Wadsworth International Group, Wadsworth, OH, USA, 1984.
[27] I. Obeidat and M. AlZubi, "Developing A faster pattern matching algorithms for intrusion detection system," *International Journal of Computing*, vol. 18, no. 3, pp. 278–284, 2019.
[28] J. S. R. Jang, E. Mizutani, and C. T. Sun, *Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice-Hall, New Jersey, NJ, USA, 1996.
[29] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiersfiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.
[30] P. Aggarwal and S. K. Sharma, "Analysis of KDD dataset attributes - class wise for intrusion detection," in *Proceedings of the 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015)*, Ghaziabad, India, MAR 2015.
[31] H. Suhaimi, S. I. Suliman, I. Musirin, A. F. Harun, and R. Mohamad, "Network intrusion detection system by using genetic algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 3, p. 1593, 2019.
[32] S. Revathi and A. Malathi, "A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection," *International Journal of Engineering Research & Technology*, vol. 2, no. 12, pp. 1848–1853, 2018.
[33] P. Soni and P. Sharma, "An intrusion detection system based on KDD-99: data using data mining techniques and feature selection," *International Journal of Soft Computing and Engineering*, vol. 4, no. 3, pp. 112–118, 2014.
[34] F. Salo, M. Injadat, A. B. Nassif, A. Shami, and A. Essex, "Data mining techniques in intrusion detection systems: a systematic literature review," *IEEE Access*, vol. 6, pp. 56046–56058, 2018.
[35] N. Kumar, B. Sivarama Bhadri Raju, M. S. V. Vardhan, and B. Vishnu, "A novel approach for selective feature mechanism for two-phase intrusion detection system," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 1, pp. 105–116, 2019.