I.J. Wireless and Microwave Technologies, **2022, 1, 22-29**
Published Online February 2022 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2022.01.03

New Intrusion Detection Framework Using Cost Sensitive Classifier and Features

Phyo Thu Thu Khine
University of Computer Studies, Hpa-an, Myanmar
E-mail: phyothuthukhine@gmail.com

Htwe Pa Pa Win
University of Computer Studies, Hpa-an, Myanmar
E-mail: hppwucsy@gmail.com

Khin Nwe Ni Tun
University of Information Technology, Yangon, Myanmar
E-mail: knntun@gmail.com

Received: 30 September 2021; Accepted: 14 November 2021; Published: 08 February 2022

Abstract: The huge increase amount of Cyber Attacks in computer networks emerge essential requirements of intrusion detection system, IDS to monitors the cybercriminals. The inefficient or unreliable IDS can decrease the performance of security services and today world applications and make the ongoing challenges on the Cyber Security and Data mining fields. This paper proposed a new detection system for the cyber-attacks with the ensemble classification of efficient cost sensitive decision trees, CSForest classifier and the least numbers of most relevant features are selected as the additional mechanism to reduce the cost. The standard dataset, NSL-KDD, IDS is used to appraise the results and compare the previous existing systems and state-of-the-art methods. The proposed system outperforms the other existing systems and can be public a new benchmark record for the NSL-KDD datasets of intrusion detection system. The proposed combination of choosing the appropriate classifier and the selection of perfect features mechanism can produce the cost-efficient IDS system for the security world.

Index Terms: CSForest, Cyber Attacks, Cyber Security, Data mining, Feature Selection, Ensemble Classification, Intrusion Detection System, NSL-KDD

1. Introduction

Today world applications on internet are attacked with the more sophisticated manner from the cybercriminals due to the widespread use of technologies in every domain including every level of users. This evolution of malicious attacks poses serious challenges to the improvement of intrusion detection system, IDS. The foremost challenge to these IDS is in order to identify abnormal and obfuscated malware, because the malware authors are using new varies violation techniques for information hiding in detection process and they can create many suffers in a short period of time in every society. Therefore, the emergence of computer security environment has become essential for our daily lives [1, 2].

Every inappropriate action set attempts to destroy the confidentiality, integrity, or availability of various information resources means the intrusions, and the detection system in a network for those malicious actions is called a Network Intrusion Detection System (NIDS). The main target of the Network Intrusion Detection System can be simply described as monitoring and evaluating discovery, and reaction. The NIDS can be classified into two categories namely; Anomaly detection-based Network Intrusion Detection System (ADNIDS) and Signature-based Network Intrusion Detection System (SNIDS). SNIDS concentrates on the features of the information and ADNIDS emphasizes significant deviations of the users' activities by analyzing the normal traffic pattern. [3, 4]

Intrusion Detection System monitors the network traffic with multiple sensors to detect internal and external suspicious activities from network intrusions. The IDS may be hardware or software systems or the combination of both that can be capable of identifying the malicious activities or abnormal behavior in computer systems. After it has analyzed the information from the sensors, it sends feedback to the preventing systems. Although, different types of mechanism have been utilized to detect the anomalies in the efficient ways, there may be many challenges in developing of IDS to increase the performance in both complexity and accuracy [2, 5].

Many researchers have proposed different intrusion detection systems including misuse detection methods and anomaly detection methods using various ways. However, the misuse detection techniques determine known attacks with low false-positive rates and Anomaly systems determine unknown attacks with a high false-positive rate. To overcome these phenomenons, researchers turn to data mining approaches for intrusion detection systems by using classification, regression, clustering, outlier detection, and association rules. [6] They have been improving the IDS by proposing proficient features selection methods, the best performance classification mechanisms. The feature selection process reduces unnecessary features and can increase detection rate, accuracy and performance of the system [7].

Hence, this paper presents a new effective framework to detect the intrusion from the networks in cost sensitive ways. In this framework, the cost sensitive decision trees classifier, CSForest is used as the classifier and the most effective features are selected as the least number of input attributes to the system. The next sections in this paper are prepared as the following manner. Section 2 discusses the existing works that have achieved in intrusion detection techniques. Section 3 present the propose framework and section 4 evaluate the experiments. Conclusion of the work is described at section 5.

2. Previous Studies

There are a significant number of pervious existing works for IDS, intrusion detection systems by using single classifiers or ensemble classifiers with different feature reduction techniques or not. However, this section analyzed and discussed in the order of proposed year.

The group in [8] proposed a new dataset called NSL-KDD to make statistical analysis and calculate the performances of Intrusion Detection System. They also evaluated performance of their dataset with the state-of-the-art algorithms, such as Decision Tree, Naïve Bayes, NB Tree, Random Forest, Random Tree, Multilayer Perceptron and SVM. They described NB Tree has the highest performance of 82.02% and 66.15% for the KDDTest+ dataset and KDDTest-21 dataset respectively.

The authors of paper [9] presented a classification scheme using fuzzy logic for IDS system. They consider class label as the fuzzy set and use fuzzy search expression and genetic programming to classification processes. They achieve the 82.74% for NSL-KDD test sets.

The researchers in [10] proposed a new detection system using class-dependent features and the matrix are extracted in training phase and these matrixes are compared in the test phases. Their feature transformation methods, RCDFT achieved 80.141% in intrusion system.

Intrusion detection system utilizing SVM classifier is proposed in [11]. They evaluate various results with different number of features using the same SVM classification techniques and get 82.37% with all 41 features and 82.68 with 14 features. They planned to ensemble SVM with other techniques to improve accuracy.

The paper [12] is the proposal of IDS, intrusion detection system used adaptive search procedure using ensemble trees classification technique named GAR-Forest and feature selection technique to improve the accuracy. They compared their accuracy with C4.5, Naïve Byes and Multilayer perceptron and achieve 82.3989%, higher performance than the others. When they combine GAR-forest with InfoGain, CFS and Symmetrical Uncertainty features selection method their accuracy is more than all features used techniques and achieve 85.976 as the highest.

The work in [13] is the two-layer classification model using Certainty Factor modification of KNN and Naïve Bayes to detect suspicious behaviors of IoT-based attacks. They reduce the feature dimension using two layer of unsupervised method, PCA and supervised method, LDA. Their ensemble achieved 84.82%.

The research in [14] is the semi-supervised IDS applying Fuzzy logic. The feed-forward NN, SLFN with one hidden layer is trained in order to produce a fuzzy vector and the categorization is accomplished by using fuzzy quantity and clustering. They achieve 84.12%.

The model of two tiers to detect network anomaly is proposed in [15]. This is also the ensemble classification of certainty factor voting method of KNN classifiers, Naïve Bayes and Linear Discriminant Analysis, LDA for the reduction of dimension. The two-tier ensemble provides low complexity because of a dimension reduction method of feature selection, along with 83.240% of good classification rate.

The ensemble of bagging and Boosting with decision trees C4.5 and feature selection methodology is proposed in [16]. Two feature-selection methods FVBRM and Gain Ratio (GR) were tested to reduce the irrelevant features but GR produce the higher accuracy of 84.25% with 35 features and improve classification accuracy.

The authors in [17] proposed a detection system for anomaly intrusion using Decision Tree classifier named C5. Their C5 algorithm outperforms than Naïve Bayes, SVM and C4.5 and achieves 81% in accuracy for test set.

The research group of [18] presented an IDS, intrusion detection system that uses modification of binarization of grey wolf optimization technique called MBGWO feature selection method. They based on the ideas of data set separation and this data played import role in Feature selection method. The MBGWO achieve 81.58% just using 14 features.

The successful benchmark records are analyzed and produce a new record in the paper [19]. They proposed two stage classification technique in ensemble way and then also reduce the dimension level of features using hybrid feature selection method. They use Rotation Forest and Bagging strategies for classification and PSO, GA and ACO feature selection method are tried with REPT classifier and choose the best features. They can reduce to 37 features to produce 85.797% and 75.52% of accuracy and this record is the highest value at benchmark record for KDD test set.

The group of [3] recently proposed a new IDS which uses ensemble of decision tree, C5 and SVM classifier without reducing feature dimension. Their main contribution is the signature integration into abnormal intrusion detection system, which takes benefit from the strengths of this system that based on Anomaly called AIDS and signature-based intrusion detection system called SIDS. They achieve 83.24% for KDD test set.

3. Proposed Methodologies

The proposed framework consists of two parts: feature selection and classification.

3.1 Feature Selection

The primary objective of feature selection is for searching the accurate, compressed and significant subset of features in original feature dataset. Feature selection method can be categorized into filter, wrapper and ensemble approaches. To choose the features wrapper approach, calculate the precision on the features while the filter method uses the relevancy or correlation measures of the data features. Filter methods are not depending on the based classifiers and has less complexity, faster and more scalable than the wrappers. The ensembles are the hybrid methods of these two and or the integration with the learning processes to achieve the stability of feature selection standards. However, it is hard to make a change or modification in the model to get the higher performance of the classification system.

The standard intrusion detection datasets NSL-KDD include redundant and irrelevant attributes, which produce the lower performance of data classification algorithms and cause misclassification results [20-22]. The dimensional reduction of features before the classification can be more efficient and reduce cost and increase accuracy. Therefore, this paper finds the best or relevant attributes by selecting manual based on the knowledge of statistics of the data in attributes which can be missed classify and made the trial testing with the support of classifier CSForest.

3.2 Classification with Cost Sensitive Decision Forest

The Cost Sensitive Decision Forest so called CSForest is the ensemble decision tree developed by [23, 24] was originally designed to solve the prediction problems of imbalanced class in the sample datasets for software defect process in cost sensitive way. Cost-sensitivity may provide many benefits to the development processes when making the predictions to make optimization for monetary cost. It used the combination of the cost sensitive voting CSVoting when classification and the resampling techniques into forest building process to reduce the cost.

The CSForest algoritm takes the input dataset and class labels including user defined record of the number of trees, threadshold, pruning confidence factor and the minimum record number in a leaf. After getting the good attributes from the dataset, the CSForest calculate the ability of cost reducing for each attribute to get the best optimized splitting point. The cost sensitive trees are built using these splitting attributes as the root notes. If the build tree number reached required level, CSForest repeat the next level of the tree as the same manner. Then the root nodes relate to the started root node of origin. The possible numbers of trees that are built from an original root node can be computed form the attributes and the record numbers in the subset. When all the possible number of trees is built, the algorithm uses the same manner for other attributes in the dataset until it develops the trees as a forest. The algorithm is shown in Algorithm 1.

4. Experimental Evolution

4.1 Dataset Description

This paper emphasizes on the standard dataset NSL-KDD [4] for intrusion detection that are widely used by the researchers in the previous works for the benchmark records. It has no redundant instances, 42 attributes including class label. This system used 20% of the KDDTrain+ dataset in training model. It has the combination of 11743 instances for normal attacks and 13499 numbers for abnormal attacks, so the total 25192 samples in it. KDDTest+ and KDDTest-21 test sets are used as the two separate testing sets. They contain 22544 samples and 11850 samples respectively and are publicly provided to appraise the performance benchmark records of intrusion detection system.

4.2 Experimental Results

The experiment results are carried out to show the effectiveness of proposed mechanisms as the following manner: the first manner is for the proposed classifier; the second is for the selected features; the third is the combination of proposed mechanisms.

Step1: The initial step of the experiments is started to determine the proposed detection method Cost Sensitive Forest has the superior performance than other existing methods for the original datasets. KDDTrain+ is being used as the training and KDDTest+ is use as the test set similar to the existing works.

Algorithm 1. CSForest

Input: D_T, T, τ, ϵ, c, m_l
Output: a set of trees F

initialise a set of decision trees F to null;
initialise a set of good attributes A^g to null;
initialise a set of split points P^g to null;
set A^g, and P^g by calling **GetGoodAttributes**(D_T, τ, ϵ);
initialise i to 1;
while $|F| < T$ *and* $|F| < |A^g|$ **do**
  T_i = **BuildCSTree**(D_T, $A_i \in A^g$, $P_i \in P^g$, c, m_l);
  $F = F \cup T_i$;
  $i = i + 1$;
end
$i = 1$;
initialise K to $|F|$;
while $|F| < T$ *and* $i \leq K$ **do**
  /* divide the data set D_T. */
  if $A_i \in A^g$ *is categorical* **then**
    $D_T = \{D_1, D_2, ...D_{|A_i|}\}$;
  end
  if $A_i \in A^g$ *is numerical* **then**
    $D_T = \{D_1, D_2\}$;
    /* D_T is divided using A_i, P_i */
  end
  /* $|\overline{D_T}|$ = number of data segments in D_T, $|D_x|$ = number of records in the xth segment */
  for $j = 1$ *to* $|\overline{D_T}|$ **do**
    initialise A_j^g, and P_j^g to null;
    A_j^g, and P_j^g = **GetGoodAttributes**(D_j, τ, ϵ);
  end
  initialise number of possible trees, t_p to null;
  calculate, $t_p = \frac{\sum_{j=1}^{|\overline{D_T}|} |A_j^g| \times |D_j|}{\sum_{j=1}^{|\overline{D_T}|} |D_j|}$;
  initialise x to 1;
  while $|F| < T$ *and* $x \leq t_p$ **do**
    for $j = 1$ *to* $|\overline{D_T}|$ **do**
      if $|A_j^g| > x$ **then**
        t_j = **BuildCSTree**(D_j, $a_{x+1} \in A_j^g$, $p_{x+1} \in P_j^g$, c, m_l);
      end
      if $|A_j^g| \leq x$ **then**
        t_j = **BuildCSTree**(D_j, $a_1 \in A_j^g$, $p_1 \in P_j^g$, c, m_l);
      end
      Build a tree T_{new} by joining the root node of each t_j $(1 \leq j \leq |\overline{D_T}|)$ as a child node with the root node of T_i;
      $F = F \cup T_{new}$;
      $x = x + 1$;
    end
  end
  $i = i + 1$;
end
return F;

Table 1. Performance benchmark results comparison with some existing works of classification on KDDTest+

Works	Techniques	Accuracy (%)
Paper[4] , 2009	Naïve Bayes	76.560
Paper[4] , 2009	Decision Tree	81.050
Paper[4] , 2009	SVM	69.520
Paper[4] , 2009	Random Forest	80.670
Paper[4] , 2009	NB Tree	82.02
Paper[5], 2011	Fuzzy classifiers	82.74
Paper[7], 2014	SVM	82.37
Paper[10], 2017	FSSL Clustering	84.12
Paper[13], 2018	C4.5	81
Paper[3], 2020	C5+SVM	83.24
Proposed, 2020	Cost Sensitive Forest	84.5458

Table 2. Performance benchmark results comparison with some existing works of classification on KDDTest-21

Works	Techniques	Accuracy (%)
Paper[4] , 2009	NB Tree	66.16
Paper[10], 2017	FSSL Clustering	68.82
Proposed, 2020	Cost Sensitive Forest	70.6667

As can be viewed from the Table 1 and 2, FSSL clustering method proposed in 2017 has the superior performance for the original intrusion detection test dataset. But the proposed classification technique CSForest provide the better performance, nearly 0.42% increased for KDDTest+ and nearly 2% for KDDTEst-21 than that declared accuracy for the normal classification technique for all the features.

Step 2:

The experiment for step 2 is to discover the best features of the NSL-KDD datasets. The datasets are used as the previous step. Correlation-based feature selection (CFS) method is used with the various search methods such as Genetic Search, GreedyStepWise, PSOSearch, RankingSearch, Best First Search, and ReRankingSearch to discover the best attributes and calculate the accuracy for both test sets as displayed in Table 3.

Table 3. Accuracy results comparison from the generated features for KDD datasets.

Attribute Search Methods	**Total No: of Features**	**Accuracy (%)**	
		KDDTest+	**KDDTest-21**
PSOSearch	6	79.2184	60.962
Best First	8	81.973	66.1857
GreedyStepWise	8	81.973	66.1857
ReRankingSearch	8	81.973	66.1857
RankSearch	12	82.8336	67.7468
Genetic Algorithm	15	83.8139	69.6371

As can be viewed from Table 3, the features found from the Genetic Algorithm produce the best accuracy results among the described methods applied. The more features used to anticipate the instances; the higher accuracy results they produce as in general. However, the superior performance among the previous work declared 85.797 for Test+ and 72.52 for Test-21 respectively and it is lower than those measures. Therefore, the experiments are continued to discover the best features of the detection system by selecting the best attributes and then the results are showed in Table 4.

Table 4. Accuracy results comparison from the selected features for KDD datasets.

Selected Attribute Number	**Total No: of Features**	**Test Time(s)**		**Accuracy (%)**	
		KDDTest+	**KDDTest-21**	**KDDTest+**	**KDDTest-21**
All	41	0.44	0.23	84.5458	70.6667
2,3,4,5,6,23,24,29,30,32,33,34,35,36,37,40,41	17	0.39	0.19	81.9331	65.6624
2,3,4, 23,29, 32,33,34, 36, 40,41	11	0.47	0.22	84.1333	70.0422
2,3,4, 23,29, 32,33,34, 40,41	10	0.36	0.2	84.1288	70.0084
2,3,4,24,29,30,34,35,36,37	10	0.34	0.19	81.5028	65.4684
3,4,5,24,26,29,30,33,35,37	10	0.36	0.2	83.5877	68.8861
3,4,5,6,14,26,29,30,32,37	10	0.41	0.16	84.3905	70.692
3,4,5,6,12,14,26,29,30,32,37	11	0.38	0.16	84.4349	70.7764
3,4,5,6,12,26,29,30,32,37	10	0.39	0.17	84.395	70.7004
3,4,5,6,26,29,30,32,37	9	0.41	0.17	85.3797	72.5992
3,4,5,6,14,26,29,30,32,37,39	11	0.41	0.17	84.1732	70.2869
3,4,5,6, 29,30,32,37	8	0.33	0.17	85.9475	73.7046
3,4,5,6, 29,30,37	7	0.48	0.23	86.635	75.0127
3,4,5,6, 29,30	6	0.34	0.17	85.9697	73.7806
3,5,6, 29,30,32,37	7	0.25	0.16	87.5887	76.8017

Table 5. Best features for NSL-KDD dataset

No	S. No	Features Name	Description
1	3	Service	network service on the destination, e.g., HTTP, telnet, etc.
2	5	Dst_bytes	number of data bytes from destination to source
3	6	Flag	normal or error status of the connection
4	29	srv_serror_rate	% of connections that have „SYN‟ errors
5	30	srv_rerror_rate	% of connections that have „REJ‟ errors
6	32	dst_host_count	No. of connections to the same host as the current connection in the past two seconds
7	37	dst_host_srv_count	No. of connections to the same service as the current connection in the past two seconds

As displayed in Table 4, the various exhausted experiments have been done by selecting the efficient attributes for both test sets and finally the best features of the system have evolved. The reduced selected features not only improve the accuracy, but also decrease the time taken amount. It takes the less time and produce the minimum features number to test the instances for both test sets. The noticeable points found that, the accuracy can be improved regardless the number of attributes. In addition, the time consumption is not depending on the number of attributes. Therefore, the best attribute, the most relevant features for intrusion detection system for NSL-KDD data set are described in Table 5. Now the benchmark results from the some of the previous existing works are indicated in Table 6 and 7 for KDDTest+ and KDDTest-21 respectively. Those systems are proposed by using different feature selection methodologies and different combination of detection methods to improve the accuracy.

Table 6. Performance benchmark comparison results for KDDTest+.

Works	Classification Method	Feature Selection Method	Total No: of Features	Accuracy (%)
Paper [6], 2012	Decision Tree	RCDFT		80.141
Paper[7], 2014	SVM	-	14	82.68
Paper[8], 2016	GAR Forest	SU	32	85.056
Paper[8], 2016	GAR Forest	CFS	19	82.976
Paper[8], 2016	GAR Forest	InfoGain	8	83.641
Paper[9], 2016		Two-layer	35	84.82
Paper[11], 2017		LDA	-	83.240
Paper [12], 2018		Gain Ratio	35	84.25
Paper[14], 2019		MBGWO	14	81.58
Paper[15], 2019	Two stage Ensemble	Hybrid	37	85.797
Proposed, 2020	Cost Sensitive Forest	Select Attributes	7	87.5887

Table 7. Performance benchmark comparison results for KDDTest-21.

Works	Classification Method	Feature Selection Method	Total No: of Features	Accuracy (%)
Paper [6], 2012	Decision Tree	RCDFT		58.80
Paper[15], 2019	Two stage Ensemble	Hybrid	37	75.52
Proposed, 2020	Cost Sensitive Forest	Select Attributes	7	76.8017

The comparative benchmark results presented in table 6 and 7 proved that the proposed mechanisms are very effective approach for the IDS, intrusion detection system. Then the achieved results can be comparative against the state-of-the-art mechanisms and some of the well-known existing approaches. The weighted average measurement for the test sets are showed in Table 8.

Table 8. The benchmark records for the NSL-KDD datasets.

Dataset	TP Rate	FP Rate	Precision	Recall	F-Measure	Accuracy (%)
KDDTest+	0.876	0.116	0.881	0.876	0.876	87.5887
KDDTest-21	0.768	0.351	0.814	0.768	0.785	76.8017

5. Conclusion

This paper proposed a new framework to detect the cyber-attacks by using the Cost Sensitive Decision Trees classifier and the best relevant features. The experiments are done by using the standard dataset to compare the benchmarks results and the proposed approach outperform than some of the other existing approaches described above. The benchmark results for NSL-KDD dataset can be announced with reduce cost in both classification and selecting features from this analysis. The best classifier can improve the detection task and the most relevant features and the less attribute numbers may provide significantly benefits for the online immediate security system. However, finding the more effectiveness and the better performance for the data mining mechanism and data security methods is the essential research for the security world and may be the ongoing works.

References

[1] Khraisat, A., Gondal, I., Vamplew, P. *et al.* Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecur* 2, 20 (2019). https://doi.org/10.1186/s42400-019-0038-7

[2] Faria M.M., Monteiro A.M. (2019) Intrusion Detection in Computer Networks Based on KNN, K-Means++ and J48. In: Arai K., Kapoor S., Bhatia R. (eds) Intelligent Systems and Applications. IntelliSys 2018. Advances in Intelligent Systems and Computing, vol 868. Springer, Cham.https://doi.org/10.1007/978-3-030-01054-6_19

[3] Sandeep Gurung, Mirnal Kanti Ghose, Aroj Subedi,"Deep Learning Approach on Network Intrusion Detection System using NSL-KDD Dataset", International Journal of Computer Network and Information Security(IJCNIS), Vol.11, No.3, pp.8-14, 2019.DOI: 10.5815/ijcnis.2019.03.02

[4] Yaser Ghaderipour, Hamed Dinari. " A Flow-Based Technique to Detect Network Intrusions Using Support Vector Regression (SVR) over Some Distinguished Graph Features ", International Journal of Mathematical Sciences and Computing (IJMSC), Vol.6, No.4, pp.1-11, 2020. DOI: 10.5815/ijMSC.2020.04.01

[5] Khraisat, A.; Gondal, I.; Vamplew, P.; Kamruzzaman, J.; Alazab, A. Hybrid Intrusion Detection System Based on the Stacking Ensemble of C5 Decision Tree Classifier and One Class Support Vector Machine. *Electronics* 2020, *9*(1), 173; https://doi.org/10.3390/electronics9010173

[6] D P Gaikwad, "Intrusion Detection System Using Ensemble of Rule Learners and First Search Algorithm as Feature Selectors", International Journal of Computer Network and Information Security(IJCNIS), Vol.13, No.4, pp.26-34, 2021. DOI: 10.5815/ijcnis.2021.04.03

[7] B.A. Manjunatha, Prasanta Gogoi, M. T. Akkalappa, "Data Mining based Framework for Effective Intrusion Detection using Hybrid Feature Selection Approach", International Journal of Computer Network and Information Security(IJCNIS), Vol.11, No.8, pp.1-12, 2019.DOI: 10.5815/ijcnis.2019.08.01

[8] Tavallaee, M., Bagheri, E., Lu, W., Ghorbani, A.A., 2009. A detailed analysis of the kdd cup 99 data set, in: 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, IEEE. pp. 1–6. doi:10.1109/CISDA.2009.5356528.

[9] Krömer, P., Platoš, J., Snášel, V., Abraham, A., 2011. Fuzzy classification by evolutionary algorithms, in: 2011 IEEE International Conference on Systems, Man, and Cybernetics, IEEE. pp. 313–318. doi:10.1109/ICSMC.2011.6083684.

[10] M. Mohammadi, B. Raahemi, A. Akbari, and B. Nassersharif, "New classdependent feature transformation for intrusion detection systems", Secur. Commun. Netw., vol. 5, no. 12, pp. 1296_1311, 2012. https://doi.org/10.1002/sec.403

[11] Pervez, M.S., Farid, D.M., 2014. Feature selection and intrusion classification in nsl-kdd cup 99 dataset employing svms, in: The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014), IEEE. pp. 1–6. doi:10.1109/SKIMA.2014.7083539.

[12] Kanakarajan N.K., Muniasamy K. (2016) Improving the Accuracy of Intrusion Detection Using GAR-Forest with Feature Selection. In: Das S., Pal T., Kar S., Satapathy S., Mandal J. (eds) Proceedings of the 4th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA) 2015. Advances in Intelligent Systems and Computing, vol 404. Springer, New Delhi. https://doi.org/10.1007/978-81-322-2695-6_45

[13] H. H. Pajouh, R. Javidan, R. Khayami, D. Ali, and K.-K. R. Choo, "A two layer dimension reduction and two-tier classification model for anomaly based intrusion detection in IoT backbone networks", 2016 IEEE Transactions on Emerging Topics in Computing (Volume: 7 , Issue: 2 , April-June 1 2019). DOI: 10.1109/TETC.2016.2633228

[14] Ashfaq, R.A.R.,Wang, X.Z., Huang, J.Z., Abbas, H., He, Y.L., 2017. Fuzziness based semi-supervised learning approach for intrusion detection system. Information Sciences 378, 484–497. doi:10.1016/j.ins.2016.04.019.

[15] Pajouh, H.H., Dastghaibyfard, G. & Hashemi, S. Two-tier network anomaly detection model: a machine learning approach. J Intell Inf Syst 48, 61–74 (2017). https://doi.org/10.1007/s10844-015-0388-x

[16] Pham, N.T., Foo, E., Suriadi, S., Jeffrey, H., Lahza, H.F.M., 2018. Improving performance of intrusion detection system using ensemble methods and feature selection, in: Proceedings of the Australasian Computer Science Week Multiconference, ACM. p. 2. doi:10.1145/3167918.3167951.

[17] Khraisat A., Gondal I., Vamplew P. (2018) An Anomaly Intrusion Detection System Using C5 Decision Tree Classifier. In: Ganji M., Rashidi L., Fung B., Wang C. (eds) Trends and Applications in Knowledge Discovery and Data Mining. PAKDD 2018. Lecture Notes in Computer Science, vol 11154. Springer, Cham. https://doi.org/10.1007/978-3-030-04503-6_14

[18] Alzubi, Q.M., Anbar, M., Alqattan, Z.N.M. et al. Intrusion detection system based on a modified binary grey wolf optimisation. Neural Comput & Applic (2019). https://doi.org/10.1007/s00521-019-04103-1

[19] "TSE-IDS: A Two-Stage Classifier Ensemble for Intelligent Anomaly-Based Intrusion Detection System", IEEE Access, Volume 7, 94497 – 94507. DOI: 10.1109/ACCESS.2019.2928048

[20] D. Selvamani and V. Selvi, "A Comparative Study on the Feature Selection Techniques for Intrusion Detection System", Asian Journal of Computer Science and Technology, ISSN: 2249-0701 Vol.8 No.1, January-March 2019, pp. 42-47.

[21] Balasaraswathi, V.R., Sugumaran, M. & Hamid, Y., "Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms", J. Commun. Inf. Netw. 2, 107–119 (2017). https://doi.org/10.1007/s41650-017-0033-7

[22] Zhou, Y., Cheng, G., Jiang, S., Dai, M., "Building an Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier", Cryptography and Security (cs.CR); Machine Learning (cs.LG). 2 April 2020, 107247, doi:10.1016/j.comnet.2020.107247, doi: arXiv:1904.01352v4

[23] Siers M.J., Islam M.Z. (2014) Cost Sensitive Decision Forest and Voting for Software Defect Prediction. In: Pham DN., Park SB. (eds) PRICAI 2014: Trends in Artificial Intelligence. PRICAI 2014. Lecture Notes in Computer Science, vol 8862. Springer, Cham. https://doi.org/10.1007/978-3-319-13560-1_80

[24] M.J.Siers,M.Z.Islam,Software defect prediction using a cost sensitive decision forest and voting and a potential solution to the class imbalance problem, nformation Systems(2015), HYPERLINK "http://dx.doi.org/10.1016/j.is.2015.02.006i" http://dx.doi.org/10.1016/j.is.2015.02.006i

Authors' Profiles

Phyo Thu Thu Khine received her Ph.D (IT) from University of Computer Studies, Yangon, Myanmar in 2012. She is currently working as a Lecturer at the University of Computer Studies, Hpa-an, Myanmar. Her research interests include Image Processing, Speech processing, Digital Signal processing, Database Management System and Big Data.

Htwe Pa Pa Win received her Ph.D (IT) from University of Computer Studies, Yangon, Myanmar in 2012. She is currently working as a Lecturer at the University of Computer Studies, Hpa-an, Myanmar. Her research interests include Image Processing, Speech processing and Digital Signal processing.

Khin Nwe Ni Tun is a professor of Information Science Department, University of Information Technology, Yangon, Myanmar. Her research interests include Image Processing, Database Management System, Big Data, Data Mining and Web Mining.

How to cite this paper: Phyo Thu Thu Khine, Htwe Pa Pa Win, Khin Nwe Ni Tun, "New Intrusion Detection Framework Using Cost Sensitive Classifier and Features", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.12, No.1, pp. 22-29, 2022.DOI: 10.5815/ijwmt.2022.01.03

