International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-3, Issue-12, December 2015

A Novel Ensemble Learner Algorithm for Anomaly Based Detection in Intrusion Prevention System

Samah Osama M. Kamel, Hany M. Harb, Nadia Hegazi, Adly S. Tag El Dein, Hala M.Abd El Kader

***Abstract*— With the rapid growth of information technology and its application, the virus has become one of the major threats to information security. This research focuses on anomaly based detection and prediction in the network based intrusion prevention system (IPS) to detect the novel attack with identifying the type of the novel attack to deal with such attack and will prevent it. The use of datamining can successfully detect novel attacks, but the challenge of this technique is its high false alarm and classification time. The continuous network traffic changes over time due to alter the original data. An automated mechanism is used to detect the change of the traffic pattern to protect the performance of IPS. So we need to build a model to learn new events that it accommodates the new data without destroying the original data. But the main problem in this subject is concept drift because of The original data may be conflicting with a new data. This paper will present a novel ensemble learner algorithm which based on grading ensemble learner to deal with the concept drift problem and achieve the detection of novel threats with high performance metrics and low false alarm and classification time. The novel ensemble learner algorithm combines the idea of boosting and bagging techniques which is based on multi-layer ensemble classifiers. The novel algorithm consists of multi-level of classification to achieve high performance metrics and low classification time.**

***Index Terms*— Intrusion Prevention System (IPS), Concept Drift, Ensemble learner, Bagging ensemble learner, Multi-layer perceptron (MLP).**

I. INTRODUCTION

The changes are taking place in our world that plays an important role in our lives. The dynamic environment like the internet is receiving data over time, which is carrying more threats to the network. The feature of traffic pattern changes over time due to vary the original data distribution. The model is built on the original data to learn a new data. The original data may be conflicting with a new data. This trouble is called the concept drift. We need an automated mechanism to detect the change of traffic pattern to improve the accuracy of the IPS. A difficult problem in the concept drift is that it can't distinguish between true concept drift and noise.

Drift has many different forms which are gradually, sudden and cyclic (seasonal) drift. Gradually drift occurs when the pattern changes from one class to another smoothly.

When gradually drift repeats over time, it leads to cyclic or seasonal drift. Sudden drift occurs when the entire pattern of concept changes into a new one which takes place in the old pattern.

There are two solutions to prevent a drop in prediction accuracy that are an active and a passive solution. An active solution; the changing in the original data leads to change in the prediction result. When the concept drift is detected, the model is replaced with another one to protect the prediction accuracy. In the passive solutions; the model is continuously updated by retraining the model on the recently observed events.

There are two types of concept drift; Real concept drift and Virtual concept drift. Real concept drift refers to the changes of the subsequent probabilities of the classes which lead to the change of prediction.

The idea of the virtual concept drift is that the original data is changed, the current model is changed. These changes lead to an error in a model which is not acceptable by adding a new data distribution.

The concept drift can be categorized into three concepts that are vertical representation, horizontal representation and unstable attribute value. Vertical representation includes unstable instance, which may be increased or decreased. Horizontal representation represents unstable attributes. Unstable attribute value changes the feature of value over time.

There are three approaches to handling concept drift where can be classified in the system as follows instance selection, instance weighting and ensemble learners. Instance selection selects instances which are related to the current concept. It includes two methodologies that are fixed and adaptive windows. Two methodologies are used to learn new or forgetting data. The fixed windows adjust the size of windows according to the current concept drift to delete noise, irrelevant and redundant cases. Fixed windows measure the characteristics of new or forgetting data by measuring the rate and the type of the drift. There are two types of fixed windows that are small and large windows. Small windows lead to the low stability because there are a few samples in the windows that are used to train the model. A large window may lead to less responsiveness to the change. The second methodology is adaptive windows that it adjusts the size of the widows according to the change over time.

Instance weighting uses the ability of some learning algorithms such as Support Vector Machines (SVMs) to process weighted instances. Instances can be weighted according to the current concept which consists of the decreasing time of the important samples so it is hard to decide that which instances that should be assigned higher weights for them [22].

Ensemble learning algorithms are used for the dynamic environments. An ensemble learner learns a set of concept descriptions over different time intervals. It maintains a set of concept descriptions and combines the predictions by

Samah osama M. Kamel, Dep. of Informatics Electronic Research Institute, Egypt.
Hany M. Harb, Dep. of Computer Science, Faculty of Engineering in Azahr University, Egypt.
Nadia Hegazi, Dep. of Informatics Electronic Research Institute Egypt.
Adly S. Tag El Dein, Dep. of Communication Faculty of Engineering in Benha University, Egypt.
Hala M. Abd El Kader, Dep. of Communication Faculty of Engineering in Benha University, Egypt.

using voting, weighted voting, or the most selected which related to the description.

Ensemble learning algorithms can be classified into three types that are dynamic combination, continuous update of the learners and structural update. Dynamic combination; when the change of environment occurred, the base learners trained the original data. The base learners are combined with changing the combination rule. In the continuous update of the learners, the learners are either retrained in batch mode or updated online using the new data. Structural update, the ensemble designer can add more new learners to the ensemble learner algorithm.

The ensemble learner is not changeable and is unlike the hybrid constructions where algorithms can't change. So the ensemble designer can easily replace one or more learning algorithms with a more accurate one. The characteristics of the ensemble learner are combining multiple models into one model to improve the accuracy, obtaining more accurate model by increasing the numbers of algorithms and robust classifier, the training speed is high, the classification time is low, decreasing the number of false alarms (false positive and negative) to obtain a high accuracy and low cost.

Ensemble learner types are supervised ensemble, unsupervised ensemble and semi-supervised ensemble.

There are two combine methods in the ensemble learner which are combined with learning and combining by consensus.

The combining by learning includes two types of ensemble model. One of them learns labeled data, such as boosting supervised ensemble and the other learns labeled and unlabeled data, such as multi-view learning in semi-supervised ensemble. The advantage of combining by learning is gaining the useful feedbacks from labeled data and it can improve the accuracy. The disadvantages of combining by learning are that it needs to keep the labeled data to train the ensemble and it may cause overfitting of the labeled data. It can't work when no labels are available.

The combining by consensus doesn't need to learn the labeled, such as bagging supervised ensemble, consensus maximization in semi-supervised and clustering in supervised ensemble. The advantage of combining by consensus is the avoiding the overfitting of the labeled data so it can improve the generalization performance. The disadvantages of combining by consensus are that it doesn't need feedbacks from the labeled data so it can't improve the accuracy.

The learning mode technique is a mechanism which is used for the generalization from data and updating the models. The learning mode can be categorized into two types of learning that are batch learning and incremental learning. The batch learning learns a large collection of instances at once and build a single model. If an error occurs in the model, it will be discarded this model and build new model using the original data. The batch learning depends on the static data. The incremental learning deals with data arriving over time.

It is suitable for dynamic environment. It learns new data by using Adaptive incremental learning which deals with continuous network traffic arriving over time. The system self must be adjusted by updating the model to accommodate new data over time.

II. RELATED WORK

[1] The authors introduced the proposed system used a self -learning component. The proposed system consists of retraining the anomaly detector periodically by using traffic which has been flagged as normal. The verification is compared with the last training sample. The result displayed 99% detection rate with no false positives.

[2] The authors introduced the proposed system presented self-learning systems which have been proposed to detect anomalous SIP messages and filter them.

The proposed system has been analyzed a new dataset which has a minimal difference between normal and anomalous messages. These messages are called statistical analysis. The dataset demonstrates self-similar characteristics with the highest constraints. The proposed system has been used k-means clustering algorithm to detect the difference between the normal messages and anomalous messages.

[3] The authors introduced the proposed system which identifies the anomalous behavior of the user in real time environment by using a supervised ensemble technique to improve the anomaly detection rate. The concept drift is changing continuously so the detection of user behavior is critical object.

The proposed system used real data streams which are generated artificially by using the KDDCUP99 Dataset. The ensemble learning consists of three base learners that are perceptron, ML-Ozabagadwin and Binary class SVM.

The proposed system used ADWIN algorithm which is a change detector and an estimator. ADWIN keeps a variable length window of recently seen items. ADWIN automatically detects and adapts to the current rate of change. The worst classifier of the ensemble is removed and a new classifier is added to the ensemble. The important performance parameter for a classifier is kappa values. When interpreting kappa, it is also important to keep in mind that the estimated kappa itself could be due to chance. When the value of kappa is high, the stronger the agreement is existed.

[4] The authors introduced the proposed system, which used intrusion detection system (IDS) that is used the recirculation neural networks (RNN) as an anomaly detector as well as a misuse detector, an ensemble of anomaly and misuse detectors, a fusion of several detectors for correct detection and recognition of attack types.

The proposed system reconstructed the input information in the same kind on an output which applies to compression and restoration of the information RNN. Ensemble learner made of two RNN based detectors that are anomaly detector and misuse detector to analyze not only binary vectors of their decisions, but also to construct the decision basing on their output data. RNN-based detectors can compare reconstruction errors of anomaly and misuse detectors. The classifier can grow from one normal detector to many parallel neural detectors. The proposed system used both KDD'99 data and real network traffic data show by using BroIDS.

[5] The authors introduced the ensemble approach of different soft computing and hard computing techniques for intrusion detection. The ensemble learner consists of Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and Multivariate Adaptive Regression Splines (MARS).

The proposed system helps to indirectly combine the synergistic and complementary features of different learning algorithms without any complex hybridization. The proposed could be helpful in several real world applications.

[6] The authors introduced an ensemble of one-class classifiers where each adopts different learning paradigms. The techniques deployed in this ensemble model are; Linear Genetic Programming (LGP), Adaptive Neural Fuzzy Inference System (ANFIS) and Random Forest (RF).

The experimental results illustrated an improvement in detection accuracy for all classes of network traffic; Normal, Probe, DoS, U2R and R2Lby using KDD CUP'99 datasets. The individual result was able to address an imbalanced dataset problem that many of machine learning techniques fail to sufficiently address it.

[7] The authors introduced the proposed system, which is an ensemble approaches which fed with appropriate features sets. The proposed system can be helped in reducing both the number of false positives and false negatives. This system used the KDD Cup 99 dataset.

The supervised ensemble learner consists of four ensembles of decision trees. Each of the four ensembles is in charge of detecting one class of attacks and composed of four decision trees trained on different sets of features. The first three decision trees were fed with sets of five features selected in and the last decision tree was fed with the union of these three sets of five features from which the redundant features were removed.

The proposed system is shown in two experiments. The first experiment used the set of features which is selected by linear genetic programming and gave the worst results, except for the class DoS which the set of features selected by SVM performed poorly. The second experiment gave less interesting results because of the unstable distribution of the examples between the training and test sets of the KDD99 data. There were misclassified in the types of attacks which are performed by the ensemble algorithm. The result of this work showed that the accuracy obtained was not good enough for a real world application.

[8] The authors introduced the proposed system, which is a hybrid architecture for combining different feature selection algorithms for real world intrusion detection.

The proposed system investigated the performance of two feature selection algorithms involving Bayesian networks (BN) and Classification and Regression Trees (CART). The new techniques for intrusion detection are performed by using DARPA benchmark intrusion data. The result of the hybrid model Normal, Probe and DOS could be detected with 100% accuracy and U2R and R2L with 84% and 99.47% accuracies.

[9] The authors introduced the proposed system, which presented a new approach for IDS adaptability by integrating a Simple Connectionist Evolving System (SECOS) and a Winner-Takes-All (WTA) hierarchy of eXtended Classifier System (XCS) by using artificial neural network (ANN).

The proposed system has been implemented by the hybrid intrusion detection system. The proposed system consisted of two stages. The first stage is XCS local adaption, which involves an automatic adaptive learning that updates the signature base hosted in the XCS. The second stage is SECOS local adaptation which deals with the evolutionary hidden layer with the addition of a new neuron or modification of the connection weights of the most activated neuron. After the learning algorithm execution, the SECOS applies its aggregation algorithm to prune the evolution layer. It determines a subset of neurons, which are much closer one than the other and replaces them by a new single neuron.

[10] The authors introduced the proposed system which is an adaptive network intrusion detection system. The proposed system is implemented in two stage architecture (TCP connection). The first stage a probabilistic classifier is used to detect anomalies in the traffic. The second stage a Hidden Markov Model (HMM) based traffic model is used to narrow down the potential attack IP addresses.

The proposed system used DARPA dataset for implementing hybrid intrusion detection system by combining the Naive Bayesian (NB) model and HMM. NB is used for online classification and HMM model is used for offline analysis of traffic.

[11] the authors introduced the proposed system, which is a Learnable Model for Anomaly Detection (LMAD), as an ensemble real-time intrusion detection model using incremental supervised machine learning techniques.

The proposed system is based on making use of two different machine learning techniques, decision trees and attribution rule classifiers. The classifiers comprise an ensemble that provides bagging for decision making. The proposed model used the NSL-KDD'99 datasets which automatically learns new rules from continuous network stream.

III. The Proposed Model

The proposed system uses Adaptive incremental learning to detect and classify types of attacks by using adaptive incremental classifiers. The proposed system combines the idea of boosting and bagging techniques. The proposed system is based on multi-layer ensemble classifiers. It consists of two classification levels as shown in figure 1. First one contains a heterogeneous ensemble which consists of five base classifiers (different types). The different base classifiers make different error and accuracy. These classifiers are chosen according to the performance parameters which are used to evaluate each algorithm and produce a good performance. Second one is bagging ensemble based on GA.

The proposed system presents a novel ensemble learner, which is based on the combining of two ensembles. The process between them is implemented in a sequential process to create a model which gives a high prediction rate and low classification time and error metric.

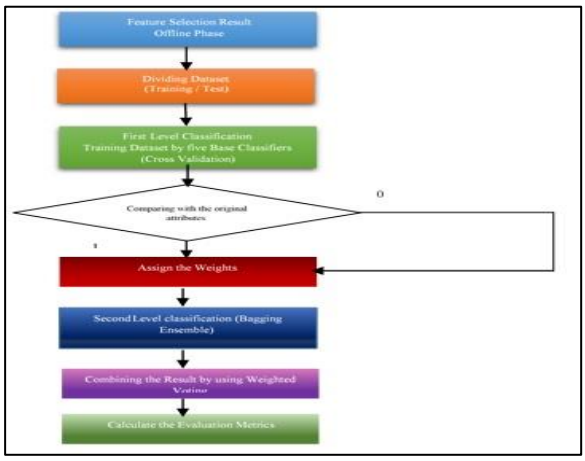

Fig. 1 The proposed system

The novel ensemble learner consists of two layers of confidence. The first layer of confidence is the first level of

classification process. The first step in the first level of classification is that preparing the dataset and classifier algorithms.

The selected number of base learner must be an odd number because of the final result is combined by weighted voting. The base learner trains the dataset by using a cross validation process. The training/test dataset is divided into a number of subsets. The number of subsets is an even number because one of subset is used for estimating the misclassification rate for each classifier algorithm and is considered as the test dataset in the cross validation process. The proposed system divides the training/test dataset into six subsets and uses five base classifiers. The next step is that the base classifiers use cross validation (6 folds) for training dataset and estimating the misclassification error rate of each base classifier. The result of each classifier algorithm is the first level of classification. The weight of the first prediction for each classifier algorithm is either zero or one. After the training process has been implemented, it will be produced in a new dataset. By comparing the new dataset by the original dataset, the new dataset is divided into two classes which called 0 or 1 class. The weights of records can be either 0 or 1 which is the probability values of records. Class 1 represents correct classified records and class 0 represents incorrect classified records. The next step is assigning the weights of records and send it to the second level of classification.

The second layer confidence is the second level of classification process which depends on the bagging ensemble learner. For every base classifier has its bagging ensemble learner. So we have five bagging ensemble learners. The final results for each bagging ensemble are combined by using weighted voting. The proposed system calculates the performance metrics for each level of classification to display the important role of second level of classification.

The proposed system uses a homogenous bagging ensemble learner, which is based on Genetic Algorithm. The bagging ensemble learner uses an independent base classifier to reduce the classification time and error. The base classifier in the bagging ensemble is Genetic Algorithm where the number of GA is ten algorithms. The process of bagging ensemble learner based on bootstrap aggregating that the dataset is sampled with replacement to create a new dataset.

The dataset consists of a number of records which is sampled a number of times. The sample is taken randomly from the original dataset and train it by using cross validation. The next sample contains records which have been picked in the previous sample and some records which have not been picked (new records in the second sample). The new records in the second sample used as a test dataset during the cross validation. The sample which is taken by randomly is replaced with another sample and so on.

Not that:

The number of records = The number of times to sample the records of the dataset.

The size of samples = the size of the dataset (training/test).

The probability of the records which have been picked in each time = 1⁄n

The probability of the records which have not been picked = 1- 1⁄n

The bootstrap process is repeated several times with different replacement samples to obtain good estimation error rate for a very small dataset.

The probability of records in the dataset which are being selected at least once = $1-(1-1/n)^n$

Where: n is the number of times in the sampling process.

For each base classifier in the bagging ensemble learner, there are around 36.8 % of the original training records which have not been picked in the training process and they will be considered as a test data set. So the cross validation process is used in this case. This process will be repeated until each record will be trained in the dataset. The final prediction of each GA in the bagging is combined by majority voting. All five results of bagging ensembles can be combined by using the weighted voting.

The model generation
x = records of dataset
N = number of records in the dataset "S"
i = number of iterations
Ṡ = bootstrap sample from S
for each iteration: i = 1 to T
Ṡ = generate bootstrap sample from S
$h = learn\left(\dot{S}\right)$ (1)
end
return $h_{1,\ldots\ldots\ldots\ldots,T}$
The output $H(x) = argmax\ \sum_i^T h_s\ (x)$ (2)

The advantages of novel ensemble learner algorithm: -

- It combines the idea of boosting and bagging ensemble learner, which it has been implemented by multi-layer confidence.
- It is used in the multicore processor.
- The novel ensemble learner produced the robust ensemble learner, which is used to classify the types of attack.
- It has been achieved the goal of anomaly based detection and prediction in the network based IPS to detect the novel attack with identifying the type of the novel attack to deal with such attack and prevent it.
- It can be used as an adaptive incremental learning which accommodates the new records without destroying the old records in dynamic environments (continuous network traffic).

III. The Adaptive Incremental Learning Algorithms

Five adaptive incremental learning algorithms are used to implement our experiment. These learning algorithms are selected according to the metrics of performance that are used to evaluate each algorithm and produce good performance that are mentioned in the previous section. The algorithms are Random Forest Tree (RFT), PART, IBK, Bayes Network and Multilayer perceptron.

A. Random Forest Tree (RFT)

Random Forest Tree (RFT) is an ensemble classifier which consists of individual decision trees. RFT combines bagging idea and a random selection of features which is independent of the structure a collection of decision trees with controlled variation. RFT is one of the highest accurate classifier among classification algorithm.

B. PART

PART algorithm is one of the decision rules and it has the highest performance among the decision rules. It uses

separate and conquers. In each iteration, it obtains the best leaf when it builds a fractional of C4.5 decision tree.

C. IBK

IBK is a lazy classifier algorithm which makes the use of the k-nearest-neighbor classifier. k-NN is a type of instance-based learning, or lazy learning. The principle of k-NN is that the input space is similar to the output space and stores the training set. When it trains the test set, k-NN identifies k instances from training set which is close to the majority class.

D. Averaged N Dependence Estimators

A1DE supports incremental learning where the classifier can be updated as new records. It achieves a high accurate classification than naïve Bayes by averaging over all of a small space of alternative naïve Bayes. The resulting algorithm is computationally efficient while delivering highly accurate classification of many learning tasks.

E. Multilayer perceptron (MLP)

It is an adaptive incremental learning and uses backpropagation to classify instances. The backpropagation calculates the difference between the outputs and obtain an error of the neural network to reduce the error rate by adjusting the weights. The proposed system uses a multilayer perceptron with 20 hidden layers. If we increase the hidden layer, the complexity may be increased. The input and output layer are considered the attributes and classes predictions respectively. The training time is called epoch is the number of iterations and is used for updating the weights and biases. The bias gives the system more flexible. Epoch is a single iteration of applying all the inputs. The learning rate ά is 0.3 which is the amount of the updated weights so the system must learn the data very fast. If ά is high, the system learns the dataset very slowly. The figure 2 illustrates the MLP construction which is captured by WEKA.

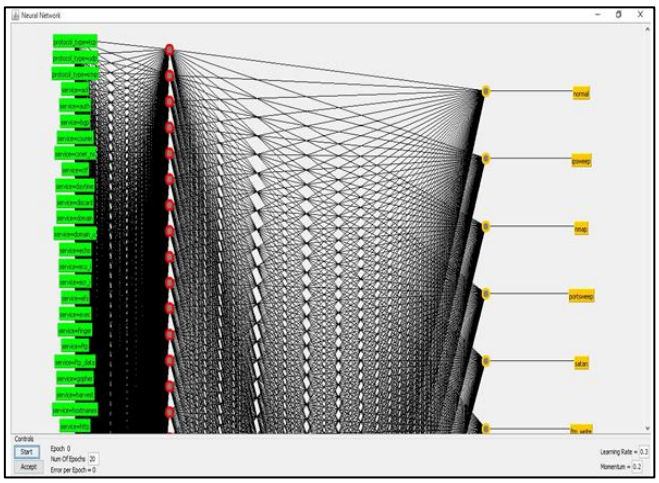

Fig. 2 MLP network by WEKA

IV. THE PROPOSED MODEL PROCESS

The following figure illustrates the process of the novel ensemble learner, which consists of two levels of classification. The first classification level contains five steps that dividing dataset, training dataset by using five classifier algorithms, creating new dataset, comparing process, assigning weights of results. The second classification level contains three steps that are bagging ensemble learner based on GA, combining the final result by using the weighted voting and calculation the performance metrics.

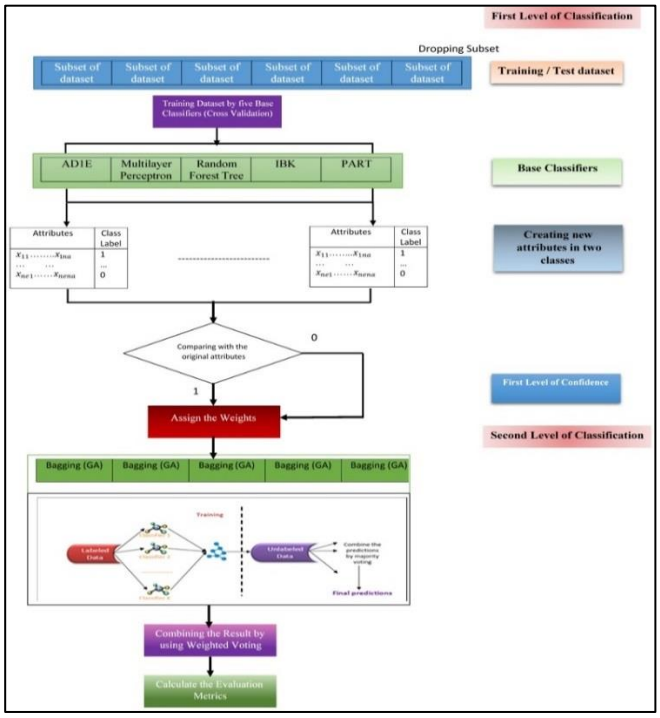

Fig. 3 The novel ensemble leaner algorithm

V. EXPERIMENT SETUP

The experiment has been implemented by using KDD and NSL-KDD dataset. Our experiments have been used to run with the used platform which is Intel Core i7 4500CPU, 2.40GHz and RAM is 8GB and MS windows 8 professional 64 bits. The development environment is Waikato Environment for Knowledge Analysis (Weka) version 3.7.12 which is an open source machine learning package. Weka applications are the Explorer, Experimenter, Knowledge Flow and Simple CLI. For Explorer section, it contains tools for data pre-processing, classification, clustering, association rules, select attributes and visualization.

The proposed system has been performed by using KDD and NSL-KDD dataset. The experiment goal is that classify and identify unknown record in the test dataset to deal with an attack and prevent it. After we removed the redundant and duplicated records, we get the different types of threats. The training dataset gave a broad diversity of intrusions and normal activities which is simulated in the network environment and indicates the legitimate network traffic. The training and test dataset belong to one of the following five categories: Normal, DoS (denial of service), R2L (root to local), U2R (user to root) and Probing (surveillance). Every attack category gives different types of attacks. The numbers of instances in the training data set are 48916 instances. The dataset includes 41 features which can be categorized into three types; numeric (or continuous), nominal (or symbolic), and binary (or discrete). The training dataset includes 41 features which can be categorized into three types; numeric (or continuous), nominal (or symbolic), and binary (or discrete). The feature selection is applied to the proposed system before the experiment will be implemented. The wrapper model is applied on the 41 features of training and test dataset. The best first search

technique is implemented to obtain the best first features in the training dataset.

The best first selected features are protocol_type, service,flag,src_bytes,land,logged_in,count,serror_rate,same_srv_rate,diff_srv_rate,dst_host_count,dst_host_srv_count,dst_host_diff_srv_rate,dst_host_same_src_port_rate,dst_host_srv_diff_host_rate and dst_host_rerror_rate.

VI. The Experiment Results and Discussion

The experiment goal is that classify and identify unknown record in the test dataset to deal with an attack and prevent it. The experiment uses the training dataset to build the model which is used to classify process of the test dataset. The experiment is divided into four experiments to decide which the classifier algorithm will give the best performance and builds a model to apply it to test dataset.

For exp. 1, the selected five base classifiers for first level of classification are RFT, PART, IBK, A1DE and MLP and the second level of classification is bagging ensemble leaner based on GA.

For exp. 2 the selected d five base classifiers for first level of classification are RFT, PART, IBK, A1DE and GA and the second level of classification is bagging ensemble leaner based on MLP.

By applying the proposed system to exp. 1 and exp. 2, the following results are illustrated in table which contains the training and testing dataset.

Table. 1 Exp. 1 and Exp. 2 for training and test dataset

	Exp. 1		Exp. 2	
	Training dataset	Test dataset	Training dataset	Test dataset
AR	99.9836	99.925	99.97	99.9069
KS	99.97	99.86	99.96	99.83
MAE	0	0.01	0	0.01
RMSE	0.39	0.82	0.43	0.92
TPR	100	99.9	100	99.9
TNR	100	99.85	100	99.77
FPR	0	0	0	0
FNR	0	0	0	0
Precision	100	99.97	100	99.9
Time to build model	11993.9 S	----	30881.1 S	----
Classification Time	5007.99 S	660 S	6021.1 S	1140 S

The conclusion illustrates the following results for training dataset:

- AR of exp. 1 is greater than exp. 2 by 0.01 %.
- RMSE of exp. 1 is less than exp. 2 by 0.04 %.
- KS of exp. 1 is greater than exp. 2 by 0.01 %.
- Time is used to test model of exp.1 is less than exp.2 by 1013.12 seconds.

The conclusion illustrates the following results for test dataset: -

- AR of exp. 1 is greater than exp. 2 by 0.02 %.
- RMSE of exp. 1 is less than exp. 2 by 0.1 %.
- KS of exp. 1 is greater than exp. 2 by 0.03 %.
- TNR of exp. 1 is greater than exp. 2 by 0.08 %.
- Classification time of exp. 1 is less than exp. 2 by 480 Seconds.

For exp. 3, the selected five base classifiers for first level of classification are RFT, PART, IBK, A1DE and MLP and the second level of classification is GA.

For exp. 4, the selected five base classifiers for first level of classification are RFT, PART, IBK, A1DE and GA and the second level of classification is MLP.

By applying the proposed system to exp. 3 and exp. 4, the following results are illustrated in table which contains the training and testing dataset.

Table. 2 Exp. 3 and Exp. 4 for training and test dataset

	Exp. 3		Exp. 4	
	Training dataset	Test dataset	Training dataset	Test dataset
AR	99.97	99.88	99.97	99.906
KS	99.96	99.8	99.96	99.83
MAE	0	0.01	0	0.01
RMSE	0.43	1.01	0.43	0.92
TPR	100	99.9	100	99.9
TNR	99.962	99.85	99.95	99.77
FPR	0	0	0	0
FNR	0	0	0	0
Precision	100	99.9	100	99.9
Time to build model	2526.21S	----	2594S	----
Classification Time	3457.6 S	660 S	2998 S	720 S

The conclusion illustrates the following results for training dataset:

- Most of results in exp. 3 equal to exp. 4 results.
- Time is used to test model of exp. 4 is than exp. 3 by 459.62 seconds.

The conclusion illustrates the following results for test dataset: -

- AR of exp.4 is greater than exp. 3 by 0.026 %.
- RMSE of exp. 4 is greater than exp. 3 by 0.09 %
- KS of exp. 4 is less than exp. 3 by 0.03 %.
- TNR of exp. 3 is greater than exp. 4 by 0.08 %.
- Classification time of exp. 3 is less than exp. 4 by 60 Seconds.

The following figures illustrate the comparison among four experiments, which is applied to test dataset (unknown dataset). The comparison includes AR, RMSE and the classification time.

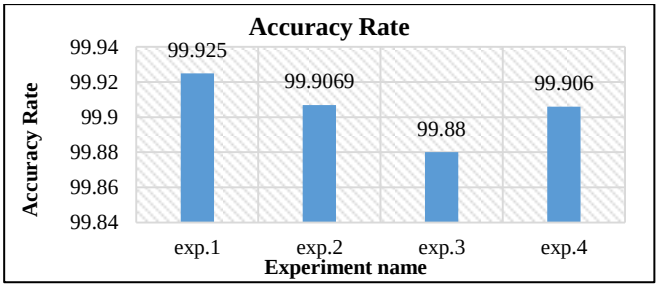

Fig. 4 Accuracy rate for four experiments

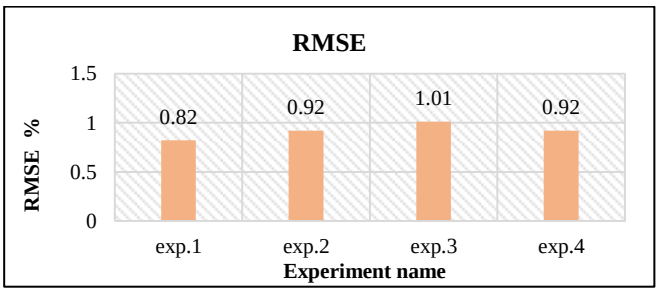

Fig. 5 RMSE for four experiments

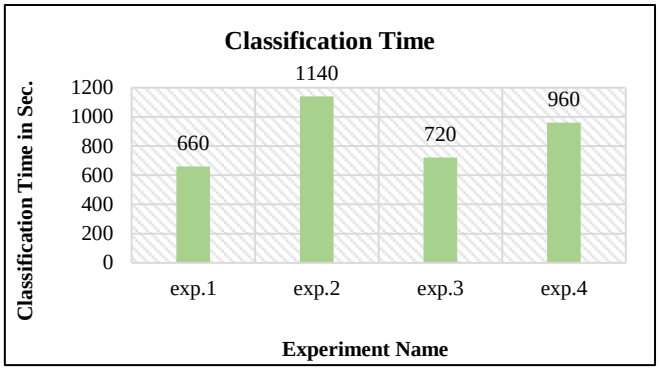

Fig. 6 Classification time for four experiments

The conclusion of all experiments displays the important role of using bagging ensemble learner as the following:

- When we used GA as a second level of classification, the performance metrics are less than the using MLP as a second classification level.
- The bagging ensemble learner based on GA developed the results of the using GA as a second level of classification.
- The performance of using bagging ensemble learner based on MLP is greater than the performance of MLP because of the complexity of the hidden layer increase and the error metrics due to increasing the number of iterations of bagging ensemble.
- The experiment 1 is the best experiment which is implemented with high performance and applied it to test dataset.

The following figures illustrate the number of records each class of attack in training and test dataset using experiment 1.

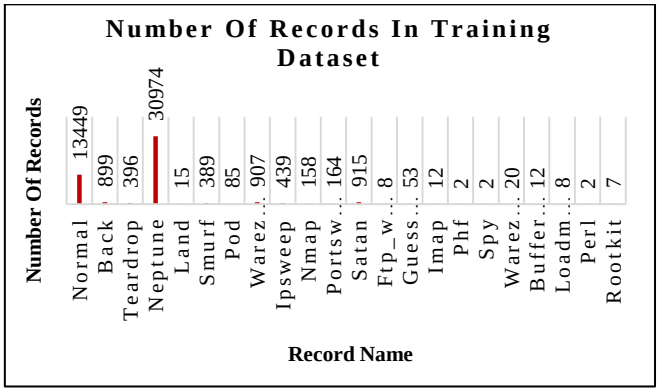

Fig. 7 The number of records in the training dataset

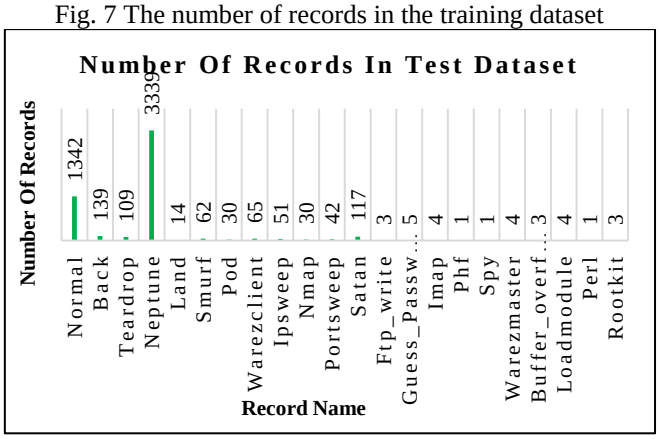

Fig. 8 The number of records in the test dataset

The Confusion Matrix of the test dataset (unknown dataset) can be displayed as shown in the following figure which achieved low confusion matrix.

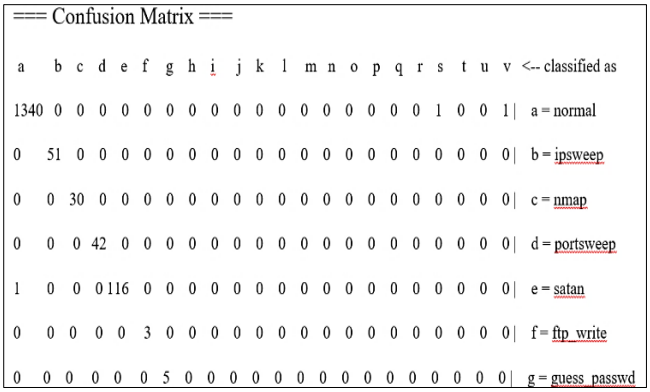

Fig. 9 Confusion matrix of test dataset

VII. CONCLUSIONS

This research proposed a novel ensemble learner algorithm. The novel system has been implemented by combining the idea of boosting and bagging ensemble learner, which has been achieved by using multi-layer confidence. It consists of two classification levels to increase the performance metrics with minimum classification time. The first level of classification is composed of five learner algorithms. The result of the first level classification is the input on the second level, which is bagging ensemble learner based on GA. The process of two levels of classification has been performed by sequence process.

This novel ensemble learning algorithm achieved the goal of anomaly based detection and prediction in the network based IPS to detect the novel attack with identifying the type of the novel attack to deal and prevent attacks. It is considered as the robust ensemble learner, which is used to analysis and classify different types of threats. The contribution of the novel ensemble learner algorithm is that it is used in the multicore processor because of the increasing the classification speed and decreasing the classification time.

It can be used as an adaptive incremental learning which accommodates the new records without destroying the old records in dynamic environments.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions that improve the presentation of this paper.

REFERENCES

[1] Konrad Rieck, Stefan Wahl, Pavel Laskov, Peter Domschitz and Klaus-Robert Müller "A self-learning system for detection of anomalous sip messages", Principles, Systems and Applications of IP Telecommunications – Chapter in Principles, Systems and Applications of IP Telecommunications. Services and Security for Next Generation Networks Volume 5310 of the series Lecture Notes in Computer Science, 2008, pp 90-106.

[2] Neda Hantehzadeh, AnilMehta, Vijay K. Gurbani, Lalit Gupta, TinKamHo and Gayan Wilathgamuwa, "Statistical Analysis of Self-Similar Session Initiation Protocol (SIP) Messages for Anomaly Detection", New Technologies, Mobility and Security (NTMS), 4th IFIP International Conference on, Paris, 7-10 Feb. 2011, pp. 1 – 5.

[3] Ravinder Reddy, Y Ramadevi and K.V.N Sunitha, "Real time anomaly detection using Ensembles", Information Science and Applications (ICISA), International Conference on, Seoul, 6-9 May 2014, pp. 1 – 4.
[4] Pavel Kachurka, Vladimir Golovko, "Fusion of recirculation neural networks for real-time network intrusion detection and recognition", International journal of computing, vol. 11, Issue 4, 2012, pp. 383-390.
[5] Srinivas Mukkamalaa, Andrew H. Sunga and Ajith Abraham, "Intrusion detection using an ensemble of intelligent paradigms", journal of network and computer applications, vol. 28, Issue 2, April 2005, pp. 167–182.
[6] Anazida Zainal, Mohd Aizaini Maarof and Siti Mariyam Shamsuddin, "Ensemble Classifiers for Network Intrusion Detection System", Journal of Information Assurance and Security 4, 2009, pp. 217-225.
[7] Alexandre Balon-Perin and Bjorn Gamb¨ ack, "ensembles of decision trees for network intrusion detection systems", International Journal on Advances in Security, vol. 6 no 1 & 2, 2013, pp. 62-77.
[8] Srilatha Chebrolua, Ajith Abrahama,b and Johnson P. Thomas, "Feature deduction and ensemble design of intrusion detection systems", Computers & Security, vol. 24, Issue 4, June 2005, pp. 295–307.
[9] Hassina Bensefia and Nacira Ghoualmi, "A New approach for adaptive intrusion detection", Computational Intelligence and Security (CIS), 2011 Seventh International Conference on, Hainan, 3-4 Dec. 2011, pp. 983 – 987.
[10] R Rangadurai Karthick , Vipul P. Hattiwale and Balaraman Ravindran, "Adaptive network intrusion detection system using a hybrid approach", Communication Systems and Networks (COMSNETS), Fourth International Conference on, Bangalore, 3-7 Jan. 2012, pp. 1–7.
[11] Abdurrahman A. Nasr, Mohamed M. Ezz and Mohamed Z. Abdulmageed, "Use of decision trees and attributional rules in incremental learning of an intrusion detection model", International Journal of Computer Networks and Communications Security, Vol. 2, No. 7, July 2014, pp. 216–224.
[12] Mrutyunjaya Panda, Ajith Abrahamb and Manas Ranjan Patra, "A Hybrid Intelligent Approach for Network Intrusion Detection", International Conference on Communication Technology and System Design (ICCTSD), 2011, pp. 1-9.
[13] Surendra K. Singhi and Huan Liu, "Error-Sensitive Grading for Model Combination", Machine Learning: ECML, vol. 3720 of the series Lecture Notes in Computer Science, 2005, pp. 724-732.
[14] Lior Rokach "Datamining and knowledge discovery handbook", "ensemble methods for classifiers", 2nd ed., Springer, 2010, ch.45.
[15] L.Dhanabal and S.P. Shantharajah, "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms", International Journal of Advanced Research in Computer and Communication Engineering vol. 4, Issue 6, June 2015, pp. 446-452.
[16] Alexander K. Seewald, "Towards a Theoretical Framework for Ensemble Classification", IJCAI'03 Proceedings the 18th Int. Joint Conference on Artificial Intelligence (IJCAI-03), 2003, pp. 1443-1444.
[17] Yi-Bin Lu, Shu-Chang Din, Chao-Fu Zheng and Bai-Jian Gao, "Using Multi-Feature and Classifier Ensembles to Improve Malware Detection", Journal of C.C.I.T., vol.39, no.2, nov 2010, pp. 57-72.
[18] "Real Life Applications of Soft Computing", Anupam Shukla, Ritu Tiwari and Rahul Kala, CRC, Press, 2010.
[19] R Rangadurai Karthick , Vipul P. Hattiwale and Balaraman Ravindran, "Adaptive network intrusion detection system using a hybrid approach", Communication Systems and Networks (COMSNETS), 2012 Fourth International Conference on, Bangalore, 3-7 Jan. 2012, pp. 1 – 7.
[20] Alexey Tsymbal, "The problem of concept drift: definitions and related work", Technical report, TCD-CS-204-15, Department of Computer Science Trinity College Dublin, Ireland, April 29, 2004.
[21] J. Zico Kolter and Marcus A. Maloof, "DynamicWeighted Majority: An Ensemble Method for Drifting Concepts", Journal of Machine Learning Research 8, 2007, pp. 2755-2790.
[22] JO ao gama, Indr˙ e ˇ zliobait.e, Albert bifet, Mykola Pechenizkiy and Abdelhamid bouchachia, "A Survey on Concept Drift Adaptation", ACM Computing Surveys, vol. 1, no. 1, Article 1, January 2013.
[23] Abdelhamid Bouchachia, "Incremental learning with multi-level adaptation", Neuro-computing, Vol. 74, Issue 11, May 2011, pp. 1785–1799.
[24] "Ensemble Methods: Foundations and Algorithms", Zhi-Hua Zhou, Chapman & Hall/Crc Machine Learnig & Pattern Recognition, June 6, 2012, CRC Press.
[25] The NSL-KDD Dataset." [Online]. Available: http://nsl.cs.unb.ca/NSL-KDD.
[26] "KDD Cup 1999 Dataset." [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

Samah Osama M. Kamel is Research Assistant in Dep. of Informatics at Electronic Research Institute, Egypt. Received B.S. degree in electronics and communications from Zagazig Faculty of Engineering, Zagazig University, in 2001. Her research interests Secure IP Telephony Attack Sensor, A survey on threats, vulnerabilities and security solutions for cellular networks, A proposed ip multimedia subsystem security framework for long term evaluation (LTE) networks.

Hany M. Harb is a professor in the faculty of Engineering at Al-Azhar University. He is currently the Head of the Computers and Systems Engineering Department. He received his Doctor of philosophy (Ph.D.), Computer Science, Illinois Institute of Technology (IIT), Chicago, Illinois, USA, 1986. He received his Master of Science in Operations Research (MSOR), IIT, 1987. He received his Master of Science in Computers and Systems Engineering, Faculty of Engineering, Azhar University, Cairo, Egypt, 1981. He received his Bachelor of Science in Computers and Control Engineering, Faculty of Engineering, Ain Shams University, Cairo, Egypt, 1978. His area of interest includes object Oriented Software Engineering Modeling Using UML, Software engineering, Distributed Systems and Computer Networks.

Nadia Hegazi is a Professor in Dep. of Informatics at Electronic Research Institute, Egypt. Her main research interests are in artificial Intelligence, Arabic computational linguistics, automatic translation, computers in education, multimedia systems, more than 25 years in language engineering. She was senior expert for strategic Planning since 2007, ministry of telecommunication and Information Technology, senior Expert for International Relations, Ministry of telecommunications and Information Technology since 2002, professor of Computer Engineering since 1986 Electronics Research Institute, Head of the Informatics Department 1980-1994, and vice President of the Electronics Research Institute 1994-2001. She Supervised More than 45 M.Sc. And PH.D. theses and published more than 90 papers in the fields of her interest.

Adly S. Tag El Dein is Professor (Associate) in Benha University, Dep. of Electrical Engineering (Faculty of Engineering at Shoubra), Egypt. Received the B.S. Degree in Electronics and communication, Benha University in 1984 and the M.Sc. in computer based speed control of single phase induction motor using three level PWM with harmonic elimination, Benha University, in 1989. The Ph.D. in optimal robot path control, Benha University, in 1993. He is currently an Associate prof. in shoubra faculty of engineering and Manager of Benha university network and information center. And his research interests include Robotics, Networks, and Communication

Hala M.Abd El Kader is a Professor at Benha University, Dep. of Electrical Engineering (Faculty of Engineering at Shoubra), Egypt. Received the B.S. Degree in Electronics and communication, Cairo university. M.Sc. And PhD Degree in Electronics and communication, Benha University. She Supervised More than 29 M.Sc. And PH.D. theses and published more than 80 papers in the fields of her interest. Her main research interests are mobile robot, analog matched filter, circuit, wireless, data acquisition system.