An Enhanced Hybrid Intrusion Detection Mechanism Based on Chicken Swarm Optimization and Naïve-Bayes Method

A.Shanthi Sona, N. Sasirekha

Abstract : ***The major important factor of network intrusion detection is to avoid malicious process in network. Since, existing modules are out-dated because of improper authentication and the network may get affected because of new attacks and malwares. In this research, Hybrid module is formed by using Chicken Swarm Optimization and Naive Bayes classifier (HCSO-NB) for classification of intrusion data. The hybrid method is introduced to detect the features efficiently in complex dataset because strategy which is designed to be capable of detecting huge data in network. Some traditional methods results in serious limitations in case of complex datasets. The algorithms are shared their properties together to discover better optimization results and the classification precisions values. This paper examines the feature selection performance by utilizing NSL-KDD-99 dataset and comparing it with the Swarm Intelligence (SI), Naïve-Bayes classifier and proposed HCSO-NB algorithms. The proposed classification process designed in NETBEANS 8.2 tool. Experiments show that proposed HCSO-NB successfully improved the accuracy.***

Keywords--- Chicken swarm optimization, Classification, Network intrusion detection, Naïve-Bayes classifier, Swarm Intelligence

I. INTRODUCTION

Because of the tremendous improvement in the field of information development, one of the wonderful testing issues is security. Thus, interruption discovery structure or Intrusion Detection System (IDS) which is a basic piece of the framework that should to be verified. The conventional IDS techniques like Genetic Algorithm (GA) [1] and improved Support Vector Machine (SVM) [2] can't deal with recently emerging attacks or assaults. The fundamental objective of IDSs is to recognize and distinguish the ordinary and unusual system tasks in a precise also, snappy way which is considered as one of the principle issues in interruption identification framework in terms of the extensive measure of qualities or highlights. In such cases, information mining based system interruption detection is generally used to recognize how and where the interruptions are occurring.

Related to realizing ongoing interruption identification mechanism, researchers have examined a few techniques for performing highlight determination process. Lessening the amount of features by choosing the crucial features is essential to upgrade the accuracy and speed of characterization calculations. Thus, choosing the exact features from a complex dataset is a best way to build up the classifier.

The examination on machine learning or data mining considers the interruption recognition as a classifier issue. Some traditional algorithms that are considered for executing calculations are combined in to set of algorithms that may apply into small subset [3]. For example, Naıve Bayes, GA, neural Network, SVM and decision tree are considered for classifier modules. In request to improve the exactness of an individual classifier, surely understood methodology is to consolidate with the classifiers. In recent days, application of swarm based methodology for interruption discovery has been picked up with an evident quality among the exploration network [4]. Swarm insight can be a measure displaying the collective behaviour of social insect crawly provinces or other social orders to actualize calculations [5].The potential of swarm intelligence makes it an ideal competitor for IDS, which needs to separate typical and anomalous practices from complex dataset.

Generally, IDS items utilize three unique techniques to distinguish interruptions. Initially, they can search for recognized attacks, which are streams or examples of information recently distinguished as an assault. Second, they can search for framework misuse, for example, unapproved endeavours to get to documents or refused traffic inside the firewall. Third, they can search for exercises that are not quite the same as the client's or frameworks typical module. These anomaly based items are intended to recognize inconspicuous changes or new attack. Therefore, the detection of attacks in busy network is difficult and it is less effective. Hence, this research provides the exact detection of data modules with the complex dataset.

The IDS strategies are dynamic to keep frameworks and systems from pernicious practices. Anyhow, conventional system interruption, for example, firewalls, user authentication and information encryptions have deserted to totally shield systems. It is very difficult to maintain the network free from refined assaults and malwares. Hence, Chung and Wahid (2012) proposed an intelligent dynamic swarm based rough set for highlight determination and streamlined swarm advancement for interruption information grouping [6].

Manuscript published on 30 September 2019
* Correspondence Author

Mrs. A.Shanthi Sona*, PG and Research Department of Computer Science,Tiruppur Kumaran College for women,Tirupur, India.Email: shanthisona@yahoo.in

Dr. N. Sasirekha, PG and Research Department of Computer Science,Vidyasagar College of Arts and Science, Udumalpet, india.Email: nsasirekhaudt@gmail.com

It holds the combination of SSO classifier and additional Weighted Local Search for detecting the large dataset. The motivation behind this mechanism is to find the better arrangement of the present method delivered by SSO with the help of NSL-KDD-99 dataset. They accomplished the characterization precision as 93.3% and it very well may be one of the aggressive classifier for the interruption recognition framework.

The intrusion detection depends upon various factors such as attack signature database, reporting and automated response capabilities and remote network updating. To avoid such intrusion, the network should be aware of attack. Since, the sensor network adapts the concept of dynamic model for upgrading [7]. The main motivation is investigating, examining, analysing and representing the records from activities of network. IDS are a hardware or software or combination of both hardware and software with aggressive defensive process that protects information, networks and systems. It analyse the traffic in network, detects the suspicious communication, reporting to security manager and controlling communications and ports. Generally, the detection units are considered for monitoring the attacks and avoid misbehaves of the system. Random forest classifier is also utilized in the intrusion concept, Tesfahun and Bhaskari (2013) [8]. It provides better accuracy and good performance in designing efficient and effective Intrusion Detection System for network intrusion identification .

SVM is one of the leading concept that applied and growingly continuously for anomaly detection. It is mostly used because it has an ability to learn effectively even for complex data. SVM are used in Wireless Sensor Networks (WSN) to evaluate spatial and temporal correspondences and compromised behaviour of nodes. SVM can be applied to large data sets to find the possible methods, and it is found that Sequential Minimal Optimization (SMO) utilized for Quadratic Programming problems [8] and an adaptive network IDS has been implemented based on Principal Component Analysis (PCA) and SVM [9][10].

Kohavi and John (1997) [11] represented the element subset choice issue in managed realizing, which includes recognizing the applicable or helpful highlights in a dataset and giving only that subset to the learning calculation. The genuine interruption identification dataset contains repetitive highlights or unimportant highlights. The excess highlights make it harder to identify conceivable interruption designs, Lee and Stolfo (2000) [12].With the expanding uses of characterization calculations and highlight determination techniques for interruption location dataset, an exhaustive research guidelines are considered [13].

The role of swarm intelligence in intrusion is more because of its remarkable attributes. It describes complex issues by various researchers such as Ant colony optimization based network intrusion feature selection by Gao et al., (2005) [14], PSO-SVM [15], Radial basis function (RBF) neural network are applied for intrusion concepts Chen and Qian (2009) [16] and so on. Each research works together with others toward finding the ideal arrangement. This happens by means of immediate or circuitous interchanges (collaborations) while the operators continually processed in the inquiry space. In this appreciation, specialists can be utilized for a few hard assignments like discovering grouping rules for mis-location, find nodes for peculiarity discovery and so forth. To be sure, these self-arranging also, dispersed characteristics are profoundly considerable by advertising the way to separate a troublesome IDS issue into various straightforward techniques are considered.

Most existing frameworks that depend on swarm inspired calculations for IDS receive a type of guidelines extraction strategy which must have a maximum probability. The low complication of such swarm calculation builds up it as a significant aspirant for the production of quick, powerful and versatile IDS. The hybrid model of some traditional learning strategy is required to prompt robust IDS. From this survey, it is noticed that the algorithms are to be considered as hybrid and determines the concept of classifier and complex dataset.

Rest of this examination is portrayed as pursues: Section 2 delivers hybrid swarm intelligence procedure based HCSO-NB design methodologies. Section 3 presents experimental analysis in terms of various parametric values. At last, section 4 gives the closing up comments and recommendations are provided.

II. PROPOSED METHODOLOGY

An IDS is a security tool that every network needs by undergoing many enhancements and it is very important to define expectations from its implementation. IDS technology provides some automation in case of detecting malicious activity but still it requires some manual activity. It is important to have a well-defined plan in an organization if any intrusion is perceived and described by IDS. The success of any IDS is depends on the deployment. Hence lot of suggestion is required in both design phase and application phase.

In most of the case, the solution of hybrid models results in exact scenario. But the decision in organization can differ starting with one then onto the next. System based IDS is often used in many organizations as it have capability to display numerous schemes and does not need any software to be installed in the system as the host based IDS need to install software in the system. Hybrid solutions are implemented in some of organizations, which must test some adequate resources IDS software is memory intensive. If IDS is badly configured then the sensors may send many numbers of false positives. Hence it is essential to design baseline policy before the implementation of IDS to avoid false positives.

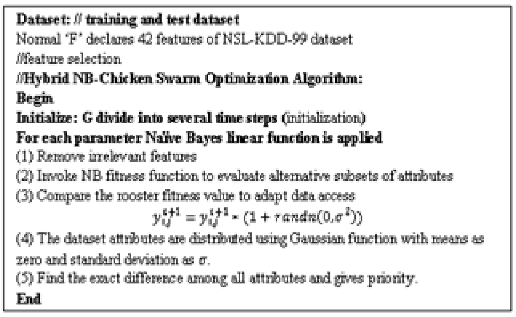
3.1 Naive Bayes classifier

NB classifier is often used in WSN to find out the intrusions because of its features like simplicity, robustness and smoothness. In addition to these characteristics many modifications is applied by data mining, pattern recognition, statistical analysis and machine learning to make it more flexible. Usually Naive Bayes Classifier can be applied largely in WSNs to identify the faulty hubs in the system. This analyse the end-to-end transmission time of every arriving packet in the sink node to determine the status of network. The main advantages of using Naive Bayes Classifier are it does not require any extra protocol and it simply suggests for faulty node. Hence, the classifier feature set is extended with the intelligence swarm based network.

3.2 Hybrid Naïve Bayes-CSO based classifier Algorithm

Meng et al., (2014) described about the CSO and its utilization in complex problems solved by chicken behaviour. It is framed with the following four different rules of chicken behaviour. The chicken swarm contains various chicken groups. Each group involves a dominant chicken, several hens, and chicks. Instructions to separate the chicken swarm into a few gatherings and decide the character of the chickens, hens and chicks. Some chickens with a few wellness esteems would be gone about as chickens, each of which would be the head chicken in a gathering.
Initially,

the chicken swarm adapted with major groups (attributes), based on the dominant rooster (high priority attributes) are particularly based on the fitness value. The terms hens (normal) and hens (anomaly) are distributed with the hierarchical order. In case of repeating the attributes the mother child relationship between each group will be maintained. Each attributes are selected with respect to the random value (1 to N).
The chickens with most exceedingly terrible a few wellness esteems would be assigned as chicks and remaining will be hens. Hens normally haphazardly pick which gathering to living in. The mum-youngster connection among the hens and the chicks is likewise arbitrarily settled. The hierarchal request, strength connection and mother-child affiliation in a gathering will stay unaltered. These statuses just refresh each few (G) time steps. Chickens take after their chicken collection to scan for nourishment, while they may shield the ones from eating their very own sustenance. Expect chickens would subjectively take the incredible sustenance formally found by others. The chicks filter for sustenance about their mother (hen). The predominant people require benefit in rivalry for sustenance.



```
Dataset: // training and test dataset
Normal 'F' declares 42 features of NSL-KDD-99 dataset
//feature selection
//Hybrid NB-Chicken Swarm Optimization Algorithm:
Begin
Initialize: G divide into several time steps (initialization)
For each parameter Naïve Bayes linear function is applied
(1) Remove irrelevant features
(2) Invoke NB fitness function to evaluate alternative subsets of attributes
(3) Compare the rooster fitness value to adapt data access
        $y_{i,j}^{t+1} = y_{i,j}^{t+1} * (1 + randn(0, \sigma^2))$
(4) The dataset attributes are distributed using Gaussian function with means as
zero and standard deviation as $\sigma$.
(5) Find the exact difference among all attributes and gives priority.
End
```

Figure 1: Concept of hybrid optimization algorithm

Figure 1: explains the concept of hybrid optimization algorithm & CSO algorithm based Cocks, hens and chicks were allocated based on the respective models as shown below:

3.2.1 Rooster search strategy

The motion behaviours of cock represented as

$$y_{i,j} = y_{i,j}(1 + Randn(0, \sigma^2)) \quad (1)$$

$$\sigma^2 = \begin{cases} 1, f_i \le f_k \\ \dfrac{f_k - f_i}{e|f_i| + \varepsilon}, else \end{cases} \quad k \in [1, N], k \ne i \quad (2)$$

From equation 2, Gauss distribution function is represented as Randn (0, σ²). Its value is displayed as 0 and the standard deviation is σ². Ɛ is utilized to keep away from zero division error which is the littlest steady in a figuring. ḳ is a record of a chicken by haphazardly browse a gathering of cockerel and *f* is the relating wellness estimation of chicken.

3.2.2 Hen search strategy

Hens pursued by the chicken and accumulate its very own sustenance. Likewise, in spite of the fact that the hens will be smothered by various chickens, they can in like manner haphazardly take sustenance which is found by various chickens. Along these lines, the development conduct of hens is mathematically represented as:

$$y_{i,j}^{(t+1)} = y_{i,j}^{t} + s_{1*} Ran\,d\left(y^t r_{1,j} - y_{i,j}^{t}\right) + s_{2*} Ran\,d\left(y_{r_{2,j}}^{t} - y_{i,j}^{t}\right) \quad (3)$$

Where,

$$s_1 = e^{\frac{f_i - f_{r_1}}{|f_i| + \varepsilon}}$$

$$s_2 = e^{f_{r_2} - f_i}$$

Among them, Randn is a uniform arbitrary numeral in [0,1]. r1ε [1 to N] it's chicken's file, and r2ε [1 to N] hen's gathering friends list is the determined record estimation of the chicken, which is arbitrarily chicken chosen from their gathering, and speak to connection of haphazardly chosen chicken r1≠r2.

3.2.3 Chicken hunt scheme

Chickens are around about their moms searching for nourishment. Along these lines, the development conduct of chickens is detailed as (3):

$$y_{i,j}^{(t+1)} = y_{i,j} + FL * \left(y_{m,j}^{t} - y_{t,j}^{t}\right) \quad (4)$$

y ᵗ m, j (mƐ[1,… n]) is speaks to the situation of the Quantity of "I" chick's mom and *FL (FL ε [0,2])* is a constraint, that characterized the chicken will gather the nourishment and chickens pursue their mom. Chicks sustenance looking system considering singular contrasts of every chick will be pursue mother to accumulate its own nourishment arbitrarily chose somewhere in the range of 0 and 2.

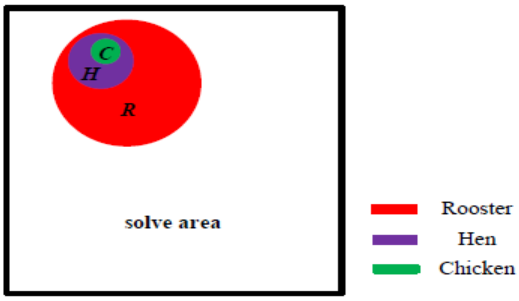


Figure 2: Representation of solving area

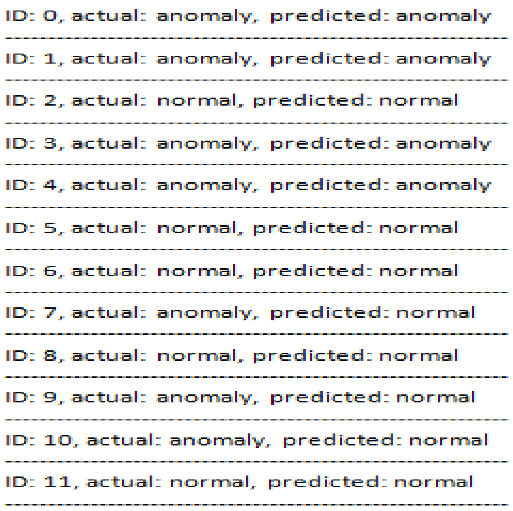
As shown in Figure 4: Performance comparison of exact predictions after a specific timeframe, as indicated by the present prosperity of every chicken, the entire chicken masses was revised to accomplish invigorate and after that again keep looking at for the ideal strategy rely on the lead of the chicken.

III. RESULT ANALYSIS

The examinations are directed utilizing NSL-KDD-99 dataset that has 60438 preparing examples, 22544 occasions for testing with

42attributes and 38 assault sorts for five class groupings to manufacture a proficient system interruption discovery framework [18-19]. The proposed classifier unit is constructed using NETBEANS8.2 IDE tool to compute and evaluating the feature selection for corresponding dataset. The performance of the Hybrid CSO-NB classifiers depends essentially on computing a possibility which is represented from the classification of a dataset. For a framed Hybrid CSO-NB classifier, the eventuality table consists of the following factors:

- True Positive (TP): number of attributes a classifier appropriately allots to the chicken swarm.
- False Positive (FP): number of attributes a classifier erroneously assigns to chicks swarm.
- False Negative (FN): number of attributes that belong to the class but which the classifier incorrectly assigns to other chicks or hens.
- True Negative (TN): number of attributes a classifier does not assign to inappropriate groups.



```
ID: 0, actual: anomaly, predicted: anomaly
------------------------------------------------------------
ID: 1, actual: anomaly, predicted: anomaly
------------------------------------------------------------
ID: 2, actual: normal, predicted: normal
------------------------------------------------------------
ID: 3, actual: anomaly, predicted: anomaly
------------------------------------------------------------
ID: 4, actual: anomaly, predicted: anomaly
------------------------------------------------------------
ID: 5, actual: normal, predicted: normal
------------------------------------------------------------
ID: 6, actual: normal, predicted: normal
------------------------------------------------------------
ID: 7, actual: anomaly, predicted: normal
------------------------------------------------------------
ID: 8, actual: normal, predicted: normal
------------------------------------------------------------
ID: 9, actual: anomaly, predicted: normal
------------------------------------------------------------
ID: 10, actual: anomaly, predicted: normal
------------------------------------------------------------
ID: 11, actual: normal, predicted: normal
------------------------------------------------------------
```

Figure 3: Sample Predicted values

As per the NSL-KDD-99 dataset, 42 subset characteristics are framed with the classification module with false positive Rate and True positive rate.

Table 1: Result Summary of Naïve Bayes,CSO and Hybrid CSO-NB

Parameters	Naive Bayes	CSO	Hybrid CSO-NB
Correct Predictions	17160	17913	18543
Incorrect Predictions	5384	4631	4001
Precision (%)	92	96	96
Recall Features (%)	63	65	74
Accuracy (%)	76	79	87.27

Comparison of two models with proposed model is shown in Table 1 and Figure 3 gives sample predicted values. Figure 4 shows comparison of correct predictions and Figure 5 incorrect predictions. The predictive model with Naïve Bayes accuracy is 76.0% and result recorded by proposed hybrid CSO-NB is 87.27%. proposed model improved the accuracy as 11.27% and 18.27% when compared to Naive Bayes system and Swarm Intelligence respectively.

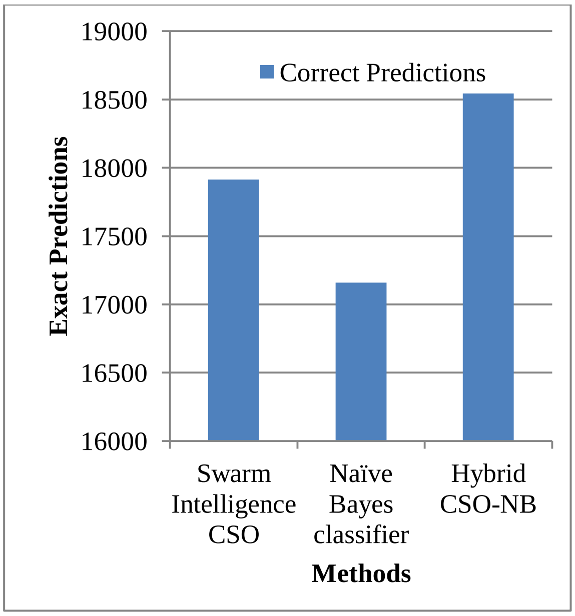


Figure 4: Performance comparison of exact predictions

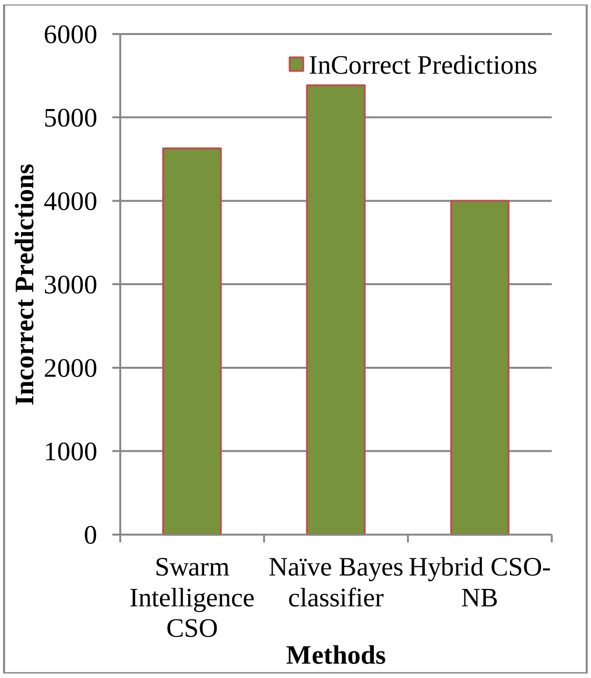


Figure 5: Incorrect predictions on three different methodologies

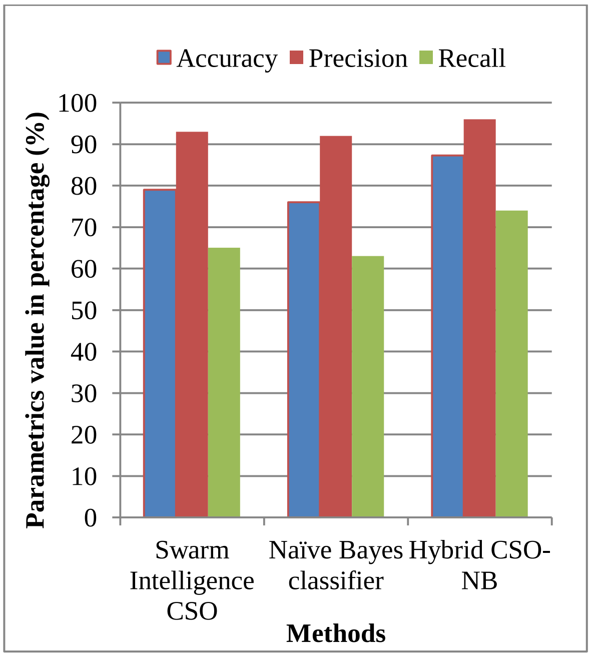


Figure 6: Comparison of accuracy, precision and recall metrics for swarm intelligence, Navies Bayes and Hybrid CSO-NB

Generally, the precision determines the positive predictive value. Likewise, the proposed method holds good segment of 96% appropriate occurrences between the recovered orders. The recall determines the sensitivity of whole detection module. In such case, the proposed intrusion detection holds the recall factor as 74% when compared with all other classifier.

IV. CONCLUSION

In this proposed research, designed hybrid model by combining the Naives Bayes and Chicken swarm optimization algorithm improves performance. The experimental results show that with the selection of base algorithm integrated with swarm optimize concept results in different outcome. The obtained results for proposed hybrid CSO-NB shows that algorithm is faster in convergence and more efficient in detecting the unidentified attacks, by filtering normal data, can easily detect intrusion by using various network fields and comparing other swarm intelligence and Naïve Bayes technique. The accuracy is greatly improved from existing accuracy 76% to 87.27% because of hybrid optimization.

REFERENCES

1. Li, W. (2004). Using genetic algorithm for network intrusion detection. Proceedings of the United States Department of Energy Cyber Security Group, 1, 1-8.
2. Kim, D. S., Nguyen, H. N., & Park, J. S. (2005, March). Genetic algorithm to improve SVM based network intrusion detection system. In 19th International Conference on Advanced Information Networking and Applications (AINA'05) Volume 1 (AINA papers) (Vol. 2, pp. 155-158). IEEE.
3. Sabhnani, M., &Serpen, G. (2003, June). Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context. In MLMTA (pp. 209-215).
4. Wu, S. X., &Banzhaf, W. (2010). The use of computational intelligence in intrusion detection systems: A review. Applied soft computing, 10(1), 1-35.
5. Bonabeau, E., Marco, D. D. R. D. F., Dorigo, M., Théraulaz, G., &Theraulaz, G. (1999). Swarm intelligence: from natural to artificial systems (No. 1). Oxford university press.
6. Chung, Y. Y., & Wahid, N. (2012). A hybrid network intrusion detection system using simplified swarm optimization (SSO). Applied Soft Computing, 12(9), 3014-3022.
7. Huo, G., & Wang, X. (2008, June). DIDS: A dynamic model of intrusion detection system in wireless sensor networks. In 2008 International Conference on Information and Automation(pp. 374-378). IEEE.
8. Tesfahun, A., &Bhaskari, D. L. (2013, November). Intrusion detection using random forests classifier with SMOTE and feature reduction. In 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies (pp. 127-132). IEEE.
9. Xu, X., & Wang, X. (2005, July). An adaptive network intrusion detection method based on PCA and support vector machines. In International Conference on Advanced Data Mining and Applications (pp. 696-703). Springer, Berlin, Heidelberg.
10. Cao, L. J., Keerthi, S. S., Ong, C. J., Uvaraj, P., Fu, X. J., & Lee, H. P. (2006). Developing parallel sequential minimal optimization for fast training support vector machine. Neurocomputing, 70(1-3), 93-104.
11. Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. Artificial intelligence, 97(1-2), 273-324.
12. Lee, W., &Stolfo, S. J. (2000). A framework for constructing features and models for intrusion detection systems. ACM transactions on Information and system security (TiSSEC), 3(4), 227-261.
13. Nguyen, H., Franke, K., &Petrovic, S. (2010, February). Improving effectiveness of intrusion detection by correlation feature selection. In 2010 International Conference on Availability, Reliability and Security (pp. 17-24). IEEE.
14. Gao, H. H., Yang, H. H., & Wang, X. Y. (2005, August). Ant colony optimization based network intrusion feature selection and detection. In 2005 International Conference on Machine Learning and Cybernetics (Vol. 6, pp. 3871-3875). IEEE.
15. Wang, J., Hong, X., Ren, R. R., & Li, T. H. (2009). A real-time intrusion detection system based on PSO-SVM. In Proceedings. The 2009 International Workshop on Information Security and Application (IWISA 2009) (p. 319). Academy Publisher.
16. Chen, Z., &Qian, P. (2009, November). Application of PSO-RBF neural network in network intrusion detection. In 2009 Third International Symposium on Intelligent Information Technology Application (Vol. 1, pp. 362-364). IEEE.
17. Meng, X., Liu, Y., Gao, X., & Zhang, H. (2014, October). A new bio-inspired algorithm: chicken swarm optimization. In International conference in swarm intelligence (pp. 86-94). Springer, Cham.
18. R. P. Lippmann, Richard P. Lippmann, David J. Fried, Isaac Graf, Joshua W. Haines, Kristopher R. Kendall, David McClung, Dan Weber, Seth E. Webster, Dan Wyschogrod, Robert K. Cunningham, and Marc and A. Zissman., "Evaluating intrusion detection systems: the 1998 DARPA off-line intrusion detection evaluation," Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00, Hilton Head, SC, USA, 2000, pp. 12-26 vol.2.
19. Peddabachigari, Sandhya, Ajith Abraham, and Johnson Thomas. "Intrusion detection systems using decision trees and support vector machines." International Journal of Applied Science and Computations, USA 11.3 (2004): 118-134.

Retrieval Number: C4737098319/19©BEIESP
DOI:10.35940/ijrte.C4737.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication