A Hybrid Intrusion Detection System for SDWSN using Random Forest (RF) Machine Learning Approach

Indira K[1]
School of Computing
Sathyabama Institute of Science and Technology
Chennai, India

Sakthi U[2]
Department of Computer Science and Engineering
St.Joseph's Institute of Technology
Chennai, India

***Abstract*—It is indeed an established fact which network security systems had certain technical problems that mostly tends to lead to security risks. Nowadays, Attackers could still continue to abuse the security vulnerabilities as well as shatter the systems and networks, and is quite pricey and even sometimes extremely difficult to resolve all layout and computing faults. The above appears to suggest that methodologies relying on preventive measures seem to be no longer secure and perhaps tracking of intrusion is necessary as a last line of defense. A Hybrid in Software Defined Wireless Sensor Network (SDWSN) the Intrusion Detection System is designed for this paper which really incorporates the benefits of Salp Swarm Optimization (SSO) algorithm as well as the classification of Machine Learning method it is based upon Random Forest (RF). We propose SSO optimization procedures to guarantee that the ideal features for the intrusion detector are chosen and in addition for improving the Random Forest (RF) classifier detection efficiency. To assess / calculate the reliability of the proposed approach here we make use of the generic NSL KDD dataset. Therefore, our proposed hybrid IDS-SSO-RF classifier further analyzes these detected abnormal activities. The known and unknown attacks are also identified. Hybrid framework also shown by the experimental results can reliably detect anomaly behavior and obtains better results in terms in terms of delay, delivery ratio, drop overhead, energy consumption and throughput.**

Keywords*—SDWSN; IDS; Salp Swarm Optimization; Random Forest Classifier

I. INTRODUCTION

Increasing computer data size has made the protection of information more critical [1]. Protection of information means protecting from unauthorized access to information and information systems [2]. When data is accessed in a network environment and transmitted via an unreliable medium, information security becomes more important [3]. Network security approaches can be typically divided in two main categories. They are (1) prevention-based techniques (2) detection based techniques [4]. The Detection-based technique strategies seek to detect intrusions that impact data centers after prevention-based techniques have failed [5]. Hence a detection system for intrusion is a detection-based strategy that detects malicious or anomalous activity by either networks or other devices [6]. By admiring defensive technologies like those of firewalls, fast encryption [30] and user access IDSs has become a key component of corporate IT security management and they are defined as systems based on misuse or anomaly [7]. In [27], only IDS are deployed in cluster heads not in every node. This method saves energy for remaining nodes and minimizes computational cost.

Technology is evolving rapidly every day and so many advancements and software developments are always being designed to protect Computer Systems from every network intrusion assault that involves various machine learning, deep learning [33] and heuristic approaches [8]. Some of the sophisticated methods in this sense, such as those focused on machine-learning techniques, e.g. Support Vector Machines (SVMs), Artificial Neural Networks, Fuzzy Logic, Bayesian Networks, Decision Trees, Random Forests, Clustering and Methods Ensemble [9][10][11][31][32]. Likewise heuristic methods like those of Genetic Algorithm (GA), Particular Swarm Optimization Algorithm (PSO), Ant Colony Optimization Algorithm (ACO), and Cuttlefish Optimization (GWO) [12][13]. Each of these approach-based IDSs, though, must produce low false-positive levels as well as comply with a large database for learning and estimation of imbalanced datasets, categorical and continuous features and a large number of features [14]. In [29], classification is done by combining SVM and KNN. In addition, some researchers are working on Data Mining (DM) technique. In a secure network environment increasingly used to detect these attacks, anomalies or intrusions [15][16].

In this work, we intend to introduce a Knowledge and Behavior-based Hybrid Intrusion Detection System (IDS) in a Software Defined Wireless Sensor Network in which Salp Swarm Optimization (SSO) and Random Forest (RF) classifier based on Decision Tree approaches plays a major role in ensuring the ideal features for intrusion detector selection and improved detection efficiency. The manuscript remaining segment is structured in the following section. In section II using machine learning algorithm and other similar works on IDS a comprehensive literature survey has convey the various intrusion detection systems. Section III addresses about the Preliminaries i.e. in proposed approach has various modules. Section IV discusses about our proposed framework. Section V by assess the performance on several metrics discuss about the simulation results of the proposed framework. Section VI provides the ending remarks.

II. Related Work

Some of the researchers past research works have been briefly described in this section among the various research works on intrusion detection method. Table I indicates the writers ' literature works with its merits and demerits.

Saurabh Dey (2019) et.al[17] presented to involve heterogeneous client networks an intrusion detection scheme for mobile clouds based on machine learning. The suggested strategy doesn't really probably require regular updates to the rules, and its level of complexity can be tailored to meet client network requirements. Technically, there are two steps in the presented scheme: multi-layer traffic screening and Virtual Machine (VM) selection based on decisions. Their experimental results show however presented scheme was really pretty constructive at intrusion detection.

Huijun Peng (2018) et.al[18] provides flow detection method based by SDN, develops anomaly SDN flow detection structures and performs flow classification detection for K-nearest neighboring algorithms using transductive confidence machines double P-value. The experiment demonstrate results that perhaps the presented algorithm reaches higher accuracy, a lower false positive rate and better adaptation SDN setting than other related algorithms.

D. Jianjian (2018) et.al[19] has primarily introduced an intrusion detection algorithm depend upon enhanced AdaBoost-RBFSVM, developed a WSN denial of service (DoS) intrusion detection system (IDS) on based the presented method. The learning result was achieved to render the RBF-SVM algorithm as the soft classifier of AdaBoost. The IABRBFSVM algorithm was presented using the impact of parameter π on RBF-SVM on the smoothness of AdaBoost weights and the template training error effect. But the other hand, the eigen space for all the attack were proposed after evaluating the DoS attack, as well as the corresponding framework for intrusion detection were developed. The presented IDS can be modeled.

Due to the inherent transparency of the communication channel, wireless networking is vulnerable to specific amount of cyber-attacks and intrusion attempts. Among other threats, electronic jamming attack stands out. As the sophistication of the attacks continues to increase, it is necessary to develop new and more reliable detection mechanisms. To address the issue of IEEE 802.11 networks electronic jamming attacks, D. Santoro (2017) et.al[20] present a novel Hybrid-NIDS (HNIDS) on based the proof theory from Dempster-Shafer (DS). The suggested approach is intended to combine the advantages of NIDSs based on signature and anomaly.

Kai Lin (2016) et.al [21] introduced a new globoid model to assess for the quality of all-directional detection during efficiently network saving the energy, dividing the sensing field into the outermost shell and interior region. Initially, they present an outermost shell coverage algorithm to ensure intruding events ' recognition performance. Then a model of Markov prediction was designed to predict the probability of movement in the adjacent area based on intruders ' historical trajectories. Using SDR technology various working frequencies will allocated to the protected nodes, according to the expected performance. In addition, a path correction plan was suggested during the operation to retrieve the missing intruders. The quality evaluations demonstrate the efficacy. Our scheme is in terms of network life, path estimation exactly and correction strategy success rate.

TABLE. I. Literature Survey

Year	Author	Contribution	Merits	Demerits
2019	Saurabh Dey, et.al	A Machine Learning Based Intrusion Detection Scheme for Data Fusion in Mobile Clouds involving Heterogeneous Client Networks	rule updates does not need. in terms of intrusion detection highly effective	Feature extraction based on interpacket delayis used here in which, the sender application times out and resends the packet sometimes when the queuing delay is through.
2018	H. Peng,. et.al	A Detection Method for Anomaly Flow in Software Defined Network	reaches a lower false positive rate. higher precision. better adaptation to the SDN environment	KNN does not work well with large dataset. KNN is sensitive to noise in the dataset
2018	D. Jianjian, T. Yang and Y. Feiyue	For Wireless Sensor Networks a Novel Intrusion Detection System based on IABRBFSVM	improved performance of network by detecting and removing malicious nodes in the network	Less Classification accuracy.
2017	D. Santoro, et.al	For Virtual Jamming Attacks on Wireless Networks a Hybrid Intrusion Detection System	to generate the hybrid IDS 100% DR, 3.8% of FPR and 0% FNR.	Dempster Shafer theory of eveidence has a problem of Potential computational complexity. It lacks a well-established decision theory
2016	Kai Lin, et.al	In SDR-based 3D WSNs node Scheduling for All-directional Intrusion Detection	improved lifetime and trajectory prediction accuracy	Complex Algorithm
2014	S. Shamshir, et al.,	In wireless sensor networks for detecting intrusion Cooperative fuzzy artificial immune system utilized	improves detection accuracy, successful defense rate	Random and uneven distribution of cluster heads by LEACH

S. A bio-inspired process, the cooperative-based fuzzy artificial immune system (Co-FAIS) has been implemented in the paper by Shamshir (2014) et al [22]. It is a modular defense strategy. It is based on the human immune system's risk concept. In terms of background antigen value (CAV) or attackers the agents accompany and collaborate with each other to measure the abnormality of sensor activity, to change the protection response threshold for fuzzy activation. By evaluating the packet components and sending the log file to the next layer sniffer module adapts to the sink node to inspect information in such a multi-node situation. To identify hazardous signal sources the fuzzy detector module combines with a hazard detector system. The contaminated origins were passed to the Fuzzy Q-learning vaccination modules (FQVM) to improve device capabilities in general. To order to produce optimal security techniques, the Cooperative Decision Making Modules (Co-DMM) combines risk detector module with the fuzzy Q-learning vaccine module. Using a network simulator the Low Energy Adaptive Clustering Hierarchy (LEACH) was tested to determine as the efficiency of the presented model.

III. PREMILINARIES

A. Random Forest (RF): A Machine Learning Approach [23]

Random Forest [26][28] is a learning model for an ensemble that takes tree choice as a fundamental classifier. As the name suggests, with an amount of trees, this algorithm produces the forest. If more trees in the forest, it appears the more robust forest. Similarly, in forest greater amount of trees provides the outcomes of elevated precision in the random forest classification. When entering a sample to be categorized, the final outcome of classification is determined by a single decision tree's output vote. Random forest overcomes decision trees ' over-fitting issue, has excellent noise and anomaly values tolerance, and has excellent scalability and parallelism to the issue of high-dimensional data classification. In contrast, random forest is a data-driven, non-parametric method of classification. It trains rules of classification through sample learning and does not involve previous classification understanding.

The model of random forests is based on forests of K choice. Each tree votes on which class a specified independent variable X belongs to, and the class it deems most suitable is given only one vote. The K decision trees description is as follows:

$$\{h(X,\theta_k), k = 1,2,.....,k\} \tag{1}$$

K represents amount of decision trees in random forests, among them. θ_k reflects random vectors that are autonomous and identically distributed. The technique of random repeated sampling is implemented to randomly extract K samples as a self-service sample set from the initial training set, and then generate regression trees for classification K. Assuming the initial training set has n characteristics, m characteristics are chosen randomly at each tree node (mn). By computing the quantity of data in each feature, a feature with the most classification capacity is chosen for node splitting among the m characteristics. Every tree develops without cutting to its peak. The trees produced are made up of random forest, and random forest classifies the fresh information. The outcomes of the classification are determined by the amount of tree classifiers votes. The resemblance and correlation of decision trees are significant characteristics of random forest to represent efficiency of generalization, while generalization error represents the system's capacity to generalize. Generalization capacity is the system's capacity to make right decisions outside the training sample set on fresh information with the same distribution. Smaller mistake in generalization can cause the scheme.

The similarity and correlation of decision trees are important features of random forest to reflect generalization performance, while generalization error reflects generalization ability of the system. Generalization ability is the ability of the system. To make correct judgments on new data with the same distribution outside the training sample set. Smaller generalization error can lead to better results of the scheme and increased generalization capability.

The error of generalization is specified as follows:

$$PE^* = P_{x,y}(mr(X,Y) < 0) \tag{2}$$

Where PE* represents a generalization error, datatype X, Y shows the probability definition area and margin function is mr(X, Y).The margin function is defined as follows:

$$mr(X,Y) = avg_k I(h(X,\theta_k) = Y) - \max_{j \neq y} avg_k I(h(X,\theta_k) = J) \tag{3}$$

In which, X is the sample of the input, Y is the right classification and J is the wrong classification. I (g) is an indicative function, avgk(g) is an average function, and h(g) is a classification model series. The margin function reflects the extent to which the numbers of votes corresponding to sample X for the correct classification exceeds the maximum number of votes for other incorrect clauses. The greater the margin function value, the greater the classifier's credibility will be. The generalization error convergence expression is described as follows:

$$\lim_{k\to\infty} PE^* = P_{x,y}(P_\theta(I(h(X,\theta_k) = Y)) - \max_{j \neq y} P_\theta(I(h(X,\theta_k) = J))) \tag{4}$$

The formula above shows that the generalization error will tend to an upper limit, and the model will not over-fit with the rise in the amount of decision trees. Depending on the classification intensity of the single tree and the correlation between the trees, the upper limit of the generalization error is accessible. The random forest model aims to establish a random forest with low correlation and high classification intensity. Classification intensity S is the mathematical expectation of mr(X, Y) in the whole sample space:

$$S = E_{x,y} mr(X,Y) \tag{5}$$

Both θ and θ′ vectors are autonomous and identically distributed and the correlation coefficients of mr(θ,X,Y) and mr(θ′,X,Y) are described as follows:

$$\bar{\rho} = \frac{\mathrm{cov}_{x,y}(mr(\theta, X, Y), mr(\theta', X, Y))}{sd(\theta)sd(\theta')} \quad (6)$$

Hence, Sd(a) can be articulated as follows, among them:

$$sd(\theta) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(mr(x_i, \theta) - \frac{1}{N}\sum_{i=1}^{N} mr(x_i, \theta))^2} \quad (7)$$

In Equation (6), it is possible to measure the correlation between the trees of h(X,θ) and h(X,θ′) on the X and Y dataset by means of $\bar{\rho}$. The greater the $\bar{\rho}$, the greater the coefficient of correlation, the upper limit of generalization error can be obtained from Chebyshev's inequality:

$$P_{x,y}(mr(X, Y) < 0) \leq \frac{\bar{\rho}(1 - \bar{s}^2)}{s^2} \quad (8)$$

To see the random forest boundary generalization error is negatively linked with a single decision tree's classification intensity S and strongly correlated with the decision trees correlation P. Therefore, the higher the intensity of classification S, the lower the correlation P, The lower the generalization error limit, the greater the accuracy of the classification.

B. Salp Swarm Algorithm (SSA): A Metaheuristic Technique [24]

Salps have a transparent body in the form of a barrel. Salps belong to the Salpidae family. Their skin cells are very comparable to those of jelly fish they comparable to jelly fish and also migrate, where water is pumped as propulsion through the body to move forward. In profound oceans Salps frequently form a swarm called the Salp chain. The primary reason of this behavior is not yet evident, but few scientists think is accomplished by using fast coordinated adjustments and foraging to achieve better locomotion. To mathematically model the Salp chains population is first split into two groups (1) leader and (2) supporters. The leader is perhaps the Salp in front of chain, while the remaining salps were regarded as followers. The leader guides swarm and supporters pursue each other (and leader straight from indirectly), as the name of these salps suggests.

In an n-dimensional search space, where n is the number of variables of a given problem, salps position is defined like other swarm-based techniques. Hence a two-dimensional matrix called x, is stored the location of all salps also thought the search space there is a food source called F as the goal of the swarm. The following equation is suggested for updating the leader's position.

$$x_j^1 = \begin{cases} F_j + c_1((ub_j - lb_j)c_2 + lb_j) & c_3 \geq 0 \\ F_j - c_1((ub_j - lb_j)c_2 + lb_j) & c_3 < 0 \end{cases} \quad (9)$$

Where x_j^1 shows the j-dimensional position of the first Salp (leader), F_j is the j-dimensional position of the food source, ub_j indicates the j-dimensional upper bound, lb_j indicates the j-dimensional lower bound, c1, c2, and c3 are random numbers. The above equation demonstrates that only with regard to the food source, the leader updates his stance. The coefficient c1 is the SSA's most significant parameter because it balances exploration and exploitation as follows:

$$c_1 = 2e^{-(\frac{4l}{L})^2} \quad (10)$$

Here, the present iteration is l and the highest amount of iterations is L. The parameters c2 and c3 are evenly produced random numbers in the [0,1] interval. In reality, they determine whether the next j-th dimension position should be towards positive infinity or negative infinity along with step size. The following equations are used to update the followers ' position.

$$x_j^i = \frac{1}{2}at^2 + v_o t \quad (11)$$

If i≥2, x_j^i indicates the position of the i-th follower salp in the j-th dimension, t is time, v 0 is the original velocity, and a calculation is as follows:

$$a = \frac{v_{final}}{v_o} where v = \frac{x - x_o}{t} \quad (12)$$

Since iteration is the time in optimization, the difference between iterations is equivalent to 1, and considering v 0=0, the following equation can be expressed as;

$$x_j^i = \frac{1}{2}\left(x_j^i + x_j^{i-1}\right) \quad (13)$$

Where i≥2 and x_j^i shows the position of i-th follower salp in j-th dimension. With equation (9) and (13), the salp chains can be simulated.

IV. Proposed Methodology

Internet security has become even more essential for personal computer subscribers, companies and indeed the army. With those of the invention of the internet, privacy has now become a significant issue, as well as the continuity of privacy enables for a better comprehension of the development of safety features. The entire sector of internet security is massive even in a developmental phase. The research range includes a good summary referring back to the origins of the internet and the current network security growth. In order to comprehend the analysis about the significance of safety to be carried out today and varieties of assaults in the networks, this article focuses on the unique hybrid Intrusion Detection System design and its evaluation in the Software Defined Wireless Sensor Network (SDWSN).

Major obstacles for recognizing intrusion from the Software Defined Wireless Sensor Network (SDWSN) were as stated [25].

The type of attack is diverse and the sources and features of attacks in wireless sensor networks differ more widely from

traditional computer networks, including most of the attacks in link layer and network layer assaults that are unique to wireless sensor networks.

Standard computer network assets, like those of networks, directories, system records as well as functions, should never be used in software defined wireless sensor networks, and we have to take into account the functionalities of info which could be used to monitor intrusion in a wireless network sensor.

There are still many different attacks on SDWSN that seem to be different from typical networks. The main issue is enhancing the effectiveness for intrusion detection system to pinpoint unidentified threats as well as to choose the applicable methodologies. Some methods seem to be appropriate for only the identification of noted threats, whilst others are ideal for the identification of unidentified threats.

A. Intrusion Detection System

Intrusion detection system monitors scheme activities in a specified setting dynamically and chooses whether these activities look like an assault or not. A primitive intrusion detection system is a detector that processes data that is to be protected from the attackers. The detector's mechanism is to remove unnecessary data from the inspection trial and portray a synthesized point of view of the users ' attitude related to safety. An ultimate decision is then taken to assess the likelihood that these actions can be regarded as intrusion symptoms. Reliable IDS is usually created by using information mining methods because they can detect intrusions excellently and execute generalizations appropriately. Indeed, it can be obviously complicated to implement and install such systems. The intrinsic complications of the schemes could classify into separate problem sets. It is based upon skill, precision and usability parameters. In Fig. 1 the basic structure of IDS represents below. Intrusion Detection System works on specific systems in which the network access to the system, i.e. sending and receiving packets, is monitored and controlled and at the same time the device file auditing is done, and the system administrator is notified about the same if there is any discrepancy. This IDS device installed in the system frequently track the computer's operating system and the benefit of this device is that it can track the entire system reliably and does not allow any other equipment to be mounted.

For attackers, valuable data is always appealing and therefore susceptible to focused network assaults. Intrusion relates with phase when intruder joins the system or system server that transfers malicious packets to client scheme for any private or significant data it can be steeled, modified or corrupted i.e. an attack relates for illegitimate network packets transmission such as user misuse, system misconfiguration, or program failures, the intrusion may occur over the server or system because of current system vulnerabilities. By placing together various vulnerabilities, one can also create a smart intrusion. In a worldwide network large numbers of internet services and millions of large servers run in the scheme. Around the same moment, such networks become more appealing to more attackers and therefore need smart intrusion detection models to protect their network system. Nevertheless, IDS built using data mining methods, primarily these methods on based anomaly detection show a greater proportion of false positive occurrences compared to earlier detection methods. Thus, processing information audit and detecting internet intrusions is hard for these methods. In addition, the learning method of the system needs big quantities of training data and excellent complexity compared to the present methodologies available. Building effective intrusion detection is therefore essential in the protection of the network system and helps to detect assaults over the network. Hence, a hybrid model for intrusion detection based on classification and a range of features are suggested here.

B. Proposed Hybrid IDS

Well into the network, the typical behavior of individuals is nothing more than an unusual practice, and one that allows the free flow of information imbalanced by usual activities and abnormalities. To enhance the IDS detection efficiency, this article presents a hybrid Intrusion Detection System (IDS). It is based upon optimized machine learning algorithm. An Intrusion Detection (ID) was described as an operational fault that is malicious and externally caused. IDS play a key role in identifying attacks, in order to detect attack, in the paper it is suggested to use a hybrid IDS technique using Salp swarm optimized random forest classifier. Our main goal is to improve the detection rate and decrease the false alarm discovery rate while identifying attacks. Salp Swarm Optimization (SSO) eliminates redundant features, and Random Forest (RF) detects attack and initiates the alert system. We suggest SSO techniques to ensure that ideal features are selected for the intrusion detector, in addition to enhance the detection efficiency of the Random Forest (RF) classifier, we used SSO for optimization.

Our Hybrid IDSs is the combination of knowledge-based approaches as well as behavior-based approaches. They generally include two detection functions; i.e. one is to accountable for identifying well-known attacks using signatures, while the other module is accountable for identifying and discovering ordinary and harmful patterns or monitoring change from ordinary profile. Hybrid IDSs are more precise with fewer false positives in terms of attack detection. However, in Software Defined Wireless sensor network the precision of the knowledge-based system completeness needs regular updating of the information of attacks. Potential for very low false alarm rates is the advantages of knowledge-based methods and that intrusion detection suggests situational analysis. Detection methods for behavior or anomaly by observing a deviation from the system's normal behavior suppose an intrusion can be detected. The normal behavior model is obtained by multiple means from the reference data gathered. This model is then compared to the present activity by the intrusion detection system and if a deviation is detected, an alarm is raised. Thus our proposed strategies of hybrid intrusion detection can identify efforts to exploit fresh and unforeseen vulnerabilities. They also assist to detect kinds of assaults that do not effectively involve exploiting any vulnerability to security previously. The structure of our proposed Intrusion Detection System (IDS) was shown in Fig. 2.

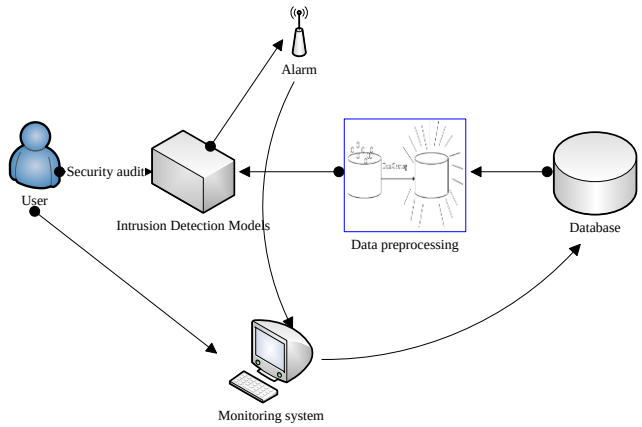

Fig. 1. Structure of IDS.

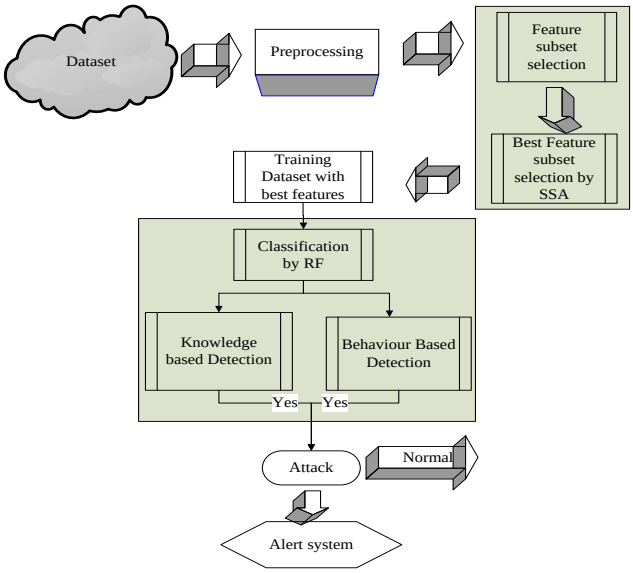

Fig. 2. Proposed Hybrid IDS System.

Data preprocessing step is usually initially utilized information mining. It is efficient for reducing dimensionality and removes irrelevant characteristics which diminish the precision. Here, a Salp Swarm Optimization algorithm is chosen to find an optimal dataset with best features as the input to the classifier. Our proposed Random forest (RF) is a group category. It is used to enhance the precision. Random forest has many decision trees. When compared other traditional classification algorithms Random forest has low classification error. For splitting each node number of trees, minimum node size and number of features are used. In random forest when constructing individual trees, to select the type of attack by split on randomization is applied. But in this work instead of randomly selecting the features we make use of Salp Swarm Optimization algorithm to find an optimal dataset with bestfeatures.

We first choose a set ofdata from the dataset and optimal featuresof the selected data is selectedby using the SSO algorithm. The entropy of each feature is calculated by using equation (14).

$$E = \sum_{\forall c} p(c) \ln \frac{1}{p(c)} \tag{14}$$

Assume that if there is an data i, feature j is used to define the split quality, which is stated as

$$Q(i, j) = \exp\{-(E_i + E_r)\} \tag{15}$$

Based on the amount of information contained in each feature, a feature with the most classification ability is selected among the k features to split the type of data withmost important and unimportant features until the decision tree grows to the maximum.

Which and how many features are important we do not know. Hence, to find the important features we take an SSO algorithm strategy. Initially, from the ranked list, we mark top features as ‘important’ and rest of the features as ‘unimportant’.

Consider A is the set of important features and B is the set of unimportant features. At each iteration these sets of features are updated.

Based upon selection criteria the features which satisfy the condition will be separated as important and unimportant features and new dataset is formed. Selection means selecting the minimum number of features that are essential for the classifier to define the normal and intrusive activity effectively and efficiently.

The generated new set of data with best features θ_k, $\{h(X, \theta_k), k = 1,2,.....,k\}$ are the input to the random forest classifier, random forest is used to classified new optimized dataset which is shown in Fig. 3. To detect the attack the final categorization solution are decided by the number of votes of the tree classifiers.

Random forest technique operates on dividing the rule and conquering system used in the task of classification. It amalgamates a group of vulnerable learners as it is an ensemble method to create well-built leaner that can exactly categorize the information. It unites the bagging system and random feature choice. In random forests, N number or tresses are generated. Each tree reflects malicious classes that are regular and different.

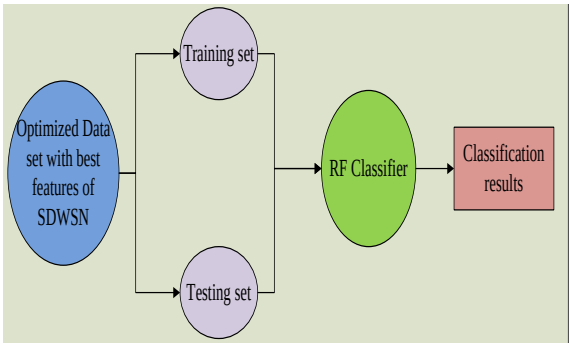

Fig. 3. Random Forest Classifier.

A big amount of datasets are readily managed by an algorithm of random forests. However, the choice of features by our suggested SSO algorithm improves several problems of the Intrusion Detection System. The pseudo code for our proposed approach in Algorithm 1 represents below.

Algorithm 1: Pseudo code for proposed Hybrid IDS

```
Input: NSL KDD dataset of SDWSN
Output: Classified result as attack or not
Initialize the Salp Population by the Dataset of SDWSN by considering ub and lb
While ( end condition is not satisfied)
Calculate the fitness the data with specific feature conditions
F= Data with best features
Update the optimal data list one by one
For each salp of xi
 If (i==1)
 Update the data with most important features
  Else
 Update the data with unimportant features
 end
End
Amend the salps based on the upper and lower bound of the data
Return f
The "N" characteristics of optimal data set are randomly selected where k << n
Create the root node N;
if (T belongs to same category C)
  {leaf node = N;
Mark N as class C;
 Return N;
 }
For i=1 to n
{Calculate Information gain (Ai);}
: ta= testing attribute;
N.ta = attribute having highest information gain;
if (N.ta == continuous )
create "n" the number of trees that a forest builds.
{ find threshold;}
For (Each T in splitting of T)
if (T is empty)
{child of N is a leaf node;}
else
{child of N= dtree T)}
calculate classification error rate of node N;
return N;
```

For the proposed system can operate in both during the offline and online phase; the training dataset is passed through the classifier while offline. This RF classification module builds the patterns that are useful for detecting intrusion. Feature selection algorithm and parameter construction with random forest algorithm is employed in this module which handles the imbalanced intrusions and after the patterns are mined, they are sent as an input to the hybrid Intrusion detector module. Similarly the intrusions are identified during the internet stage also in which the Network traffic captures the packets. For each link captured from network traffic, the pre-processors produce the feature characteristics. The detector module classifies the ordinary traffic or intrusion relation. It utilizes the models constructed in the stage of offline. Finally, the system raises an alert when attack is identified.

V. Performance Evalution

The simulation framework developed with Network Simulator (NS2) to test the data set as well as the methodology with code and devices to monitor data processing and performance. The performance evaluation using KDDCup99 dataset is the standard which includes a broad range of threats that represent serious-world intrusions in a database server. For training and testing the configuration of the workbench is carried out via the dataset KDD cup 99 among that 20% of the dataset is used here. The metrics examined to determine the feasibility of the solution proposed were listed below.

Delay: The system delay determines that how much time the data is needed to migrate to the destination across the network from the source node. Additionally, time needed to compute and find the malicious data travels on the proposed IDS model will also determine network delay. Fig. 4 demonstrates our suggested solution to traditional methods in accordance with the delay study it display the existing approach in red lines and the blue line indicates the model proposed. The overall delay is defined as time the packet / data take to reach senders through the SDWSN's network recipients. The figure shows that for our proposed strategies, the overall performance delay is minimal and the average delay for the proposed solution wireless network is 75% less than that of other KNN based approach.

Delivery Ratio (DR): The number of packets / data received at the receivers end is calculated with respect to the amount of packets transmitted at the transmission end is shown in Fig. 5. For an effective network performance, the effective SD wireless sensor network must have a significant DR quality. The DR is greater than the KNN based design for the proposed IDS-RF-SSO. For the proposed method maximum value for DR is 0.93, 0.90 and 0.720, 0.721, 0.602 at the same time current values which the existing approach are 0.736 and 0.667 and 0.493 respectively.

If the nodes are delivered to the destined user accurately then the possibility of delivery ratio is maximum and it is shown deliberately in Fig. 5. In which the blue line depicts the proposed approach with maximum delivery ratio of all is about 0.95 and the red line depicts the existing approaches with minimum delivery ratio of all is about 0.49 than that of

our proposed approach i.e. our proposed RF-SSO approach delivery ratio is 51 % better than that of KNN based IDS system.

Drop: The drop is evaluated in Fig. 6 using the number of packets / data received at the recipient end to determine the amount of data drops. For successful system performance, effective SD wireless sensor network needs a significant decrease value. The probability of a drop is small if the Identification system works effectively.

In Fig. 6 the drop analysis of our proposed approach with the existing KNN classifier is shown. In whichfor the proposed IDS-RF-SSO the drop is minimum but for KNN it is maximal. Minimum drop value of the proposed approach is 2313, 1120, 3111, 2735, 3277 whereas the existing approaches values are 2920, 22285, 14100, 14500, 20440 respectively.

Energy consumption: That node only devotes the number of resources that are not required for transmission of packets in the output queue to intrusion detection. In fact, once a packet is identified as malicious it will be excluded, as long as its evaluation is not done or labelled as good; it will be redirected according to a proposed algorithm to the destination. The values of energy consumption relating to non-malicious packets are also shown in Fig. 7 when the intrusion detection is carried out at the destination; in addition, the expense of the evaluation itself remains the same and packets are not discarded long before reaching the destination.

As the number of malicious packets increases, the amount of energy saved by early detection and discarding them also increases. In Fig. 7 because of the IDS-RF-SSO algorithm, for attacks the detection rate is better than other algorithms; malicious nodes could be detected and removed faster, slowing down the average node residual power. But energy is exhausted as some nodes, is gradually completed data packet transmission task and gradually reduced the average node's energy consumption.

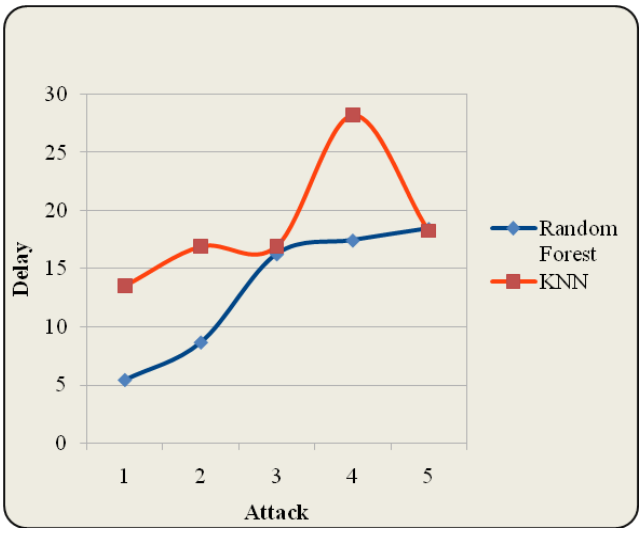

Fig. 4. Comparative Delay Analysis of RFS-SSO and KNN.

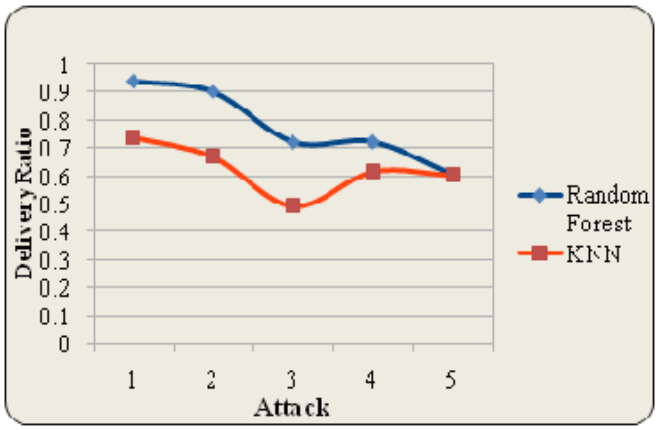

Fig. 5. Comparative Delivery Ratio Analysis of RFS-SSO and KNN.

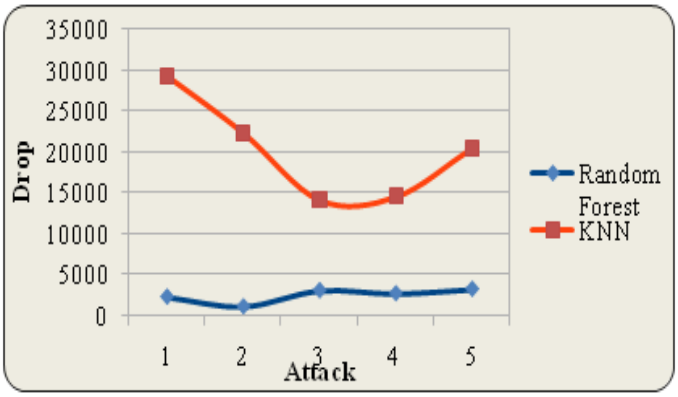

Fig. 6. Comparative Drop Analysis of RFS-SSO and KNN

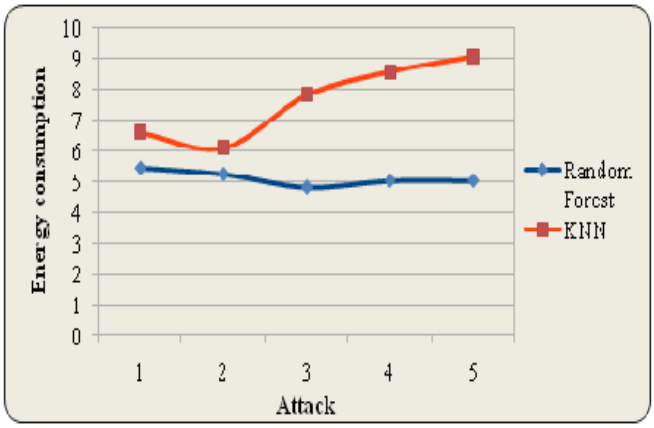

Fig. 7. Comparative Energy Consumption Analysis of RFS-SSO and KNN.

Overhead: The sum of extra data delivered during the communication activities on the network is called the overhead. Fig. 8 demonstrates the overhead during attacks on the existing IDS network and proposed IDS systems. To visualize the performance the proposed method provides using the blue line and the performance of the traditional technique provides using the red line. It depicts that, attack does not affect the overhead output in the suggested approach.

In Fig. 8 the overhead analysis of our proposed approach with the existing KNN classifier is shown. In which for the proposed IDS-RF-SSO the overhead is maximum but for KNN it is minimal. Maximum overhead value of the proposed approach is 8458, 6936, 9696, 6754, 7265 whereas the existing approaches values are 59639, 53514, 58401, 45968, and 25433 respectively.

Throughput: Successful message delivery over a communication medium of usual rate is known as Throughput. In terms of data packets per time slot or data packets per second is calculated by throughput. Compare the performance of the proposed and traditional technique. It represents red line for traditional IDS with KNN Classifier and for proposed green line is used. During IDS attacks according to the observation of results the throughput is increased for the proposed approach significantly and decreased for the existing approach. Therefore from attack the performance of the proposed IDS is not affected.

The analysis of our proposed approach with the existing KNN classifier is shown in Fig. 9. In which the throughput is optimum for the proposed IDS-RF-SSO but minimal for KNN. The proposed solution has a maximum throughput value of 31739, 24984, 24056, 19487 and 17524, while the existing approach values are 21952, 17855, 14423, 14152 and 8038 respectively.

Finally, the paper examined the quality of the full KDD dataset statistically and suggested a new methodology to deal with the challenges. Hybrid IDS with RF-SSO approach can effectively reduce the problem of complexity and multiclass dataset relative to other current algorithms. Pre-processing can easily extract and record as normal or attack the most relevant feature sub-set form network traffic. It is clear that the suggested model eliminates ordinary documentation and reduces the list of attributes, thereby increasing the IDS strain of dealing with a wide set of features. Swarm awareness strategies combined with RF Classifier can therefore effectively increase reliability of identification and deliver optimal solutions. Through finding an optimal solution in the pre-processing system, detection of invasion is rendered more reliable and inaccurate detection of attacks is minimized. The analytical result shows that perhaps the integration of the Hybrid RF-SSO algorithm is quicker and more efficient in solution.

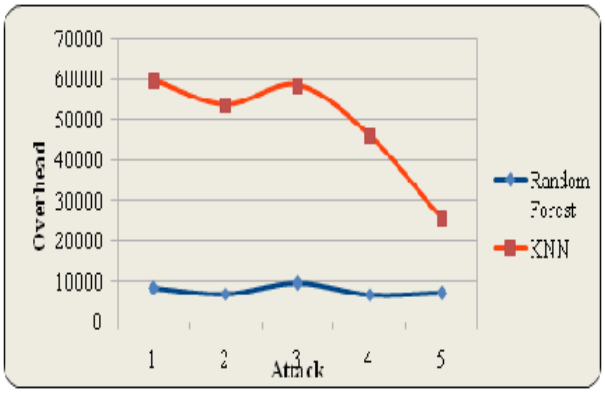

Fig. 8. Comparative Overhead Analysis of RFS-SSO and KNN.

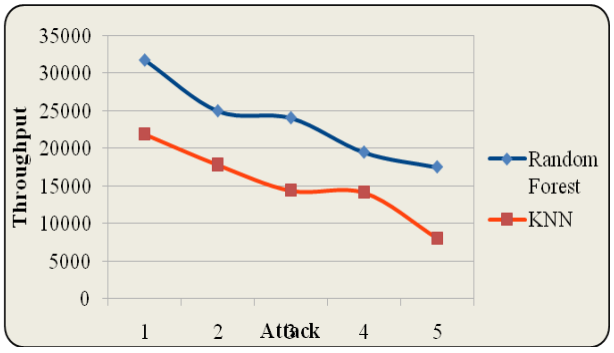

Fig. 9. Comparative Throughput Analysis of RFS-SSO and KNN.

VI. Conclusion

The contributions of this research are indeed the proposal for a hybrid aesthetic system for effective intrusion detection for service provider by utilizing the classification and optimization algorithms to enhance intrusion detection system performance. The reliability of the hybrid invasion detection system was measured in terms of delay, delivery ratio, drop overhead, energy consumption and throughput. To testing an effective hybrid intrusion detection system for SDWSNs in this research work the KDD CUP 1999 Dataset being utilized to test the proposed hybrid IDS. The system was designed on the basis of a combination of Knowledge and Behavior based IDS with RF as classifier and SSO approaches. The experimental study conducted on NSL-KDD dataset found our methodology significantly increased overall system efficiency when relative to the system performance with KNN classifier based IDS system.

References

[1] G. V. Nadiammai, S. Krishnaveni and M. Hemalatha, A Comprehensive Analysis and Study in IDS Using Data Mining Techniques, IJCA, vol. 35, pp. 51–56, November–December (2011).

[2] Arif Jamal Malik, Waseem Shahzad and Farrukh Aslam Khan, Network Intrusion Detection Using Hybrid Binary PSO and Random Forests Algorithm, Security and Communication Networks, (2012).

[3] P. Natesan and P. Balasubramanie, Multi Stage Filter Using Enhanced Adaboost for Network IDS, International Journal of Network Security and its Applications, vol. 4, no. 3, (2012).

[4] Mrutyunjaya Panda, Ajith Abraham and Manas Ranjan Patra, A Hybrid Intelligent Approach for Network Intrusion Detection, UCCTSD, pp. 1–9, (2012).

[5] Md. Al Mehedi Hasan, Mohammed Nasser, Biprodip and Shamim Ahmad, Support Vector Machine and Random Forest Modeling for IDS, JILSA, pp. 45–52, (2014).

[6] Ujwala Ravale, Nilesh Marathe and Puja Padiya, Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K Means and RBF Kernel Function, ICACTA, pp. 428–435, (2015).

[7] Aleksandar Lazarevic, Vipin Kumar and Jaideep Srivastava, Intrusion Detection: An Survey, p. 31.

[8] Araujo, Oliviera, Shinoda and Bhargava, Identifying Important Characteristics in the KDD99 Intrusion Detection Dataset by Feature Selection Using a Hybrid Approach, International Conference on Telecommunications, (2010).

[9] Nanak Chand, Preeti Mishra, C. Rama Krishna, Emmanuel Shubhakar Pilli, and Mahesh Chandra Govil, "A comparativeanalysis of SVM and its stacking with other classification algorithm for intrusion detection", In International Conferenceon Advances in Computing, Communication, & Automation, 1–6,2016.

[10] Gianluigi Folino and Pietro Sabatino. 2016. Ensemble based collaborative and distributed intrusion detection systems: A survey. 66 (May 2016), 1–16.

[11] Nabila Farnaaz and M. A. Jabbar, "Random forest modeling for network intrusion detection system", Procedia Computer Science 89 (2016), 213–217,2016.

[12] N. Cleetus and K. A. Dhanya, "Multi-objective functions in particle swarm optimization for intrusion detection", In 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI'14). 387–392,2014.

[13] Adel Sabry Eesa, Zeynep Orman, and Adnan Mohsin Abdulazeez Brifcani,"A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems",2015.

[14] Susan M. Bridges and Rayford B. Vaughn, "Fuzzy data mining and genetic algorithms applied to intrusion detection", In National Information Systems Security Conference (NISSC'00), 16–19, 2000.

[15] Anna L. Buczak and Erhan Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection", IEEE Communications Surveys & Tutorials 18, 2 (2016), 1153–1176, 2016.

[16] I. Butun, S. D. Morgera, and R. Sankar, "A survey of intrusion detection systems in wireless sensor networks", IEEE Communications Surveys Tutorials 16, 1 (2014), 266–282, 2014.

[17] S. Dey, Q. Ye and S. Sampalli, "A machine learning based intrusion detection scheme for data fusion in mobile clouds involving heterogeneous client networks", Information Fusion, vol. 49, pp. 205-215, 2019.

[18] H. Peng, Z. Sun, X. Zhao, S. Tan and Z. Sun, "A Detection Method for Anomaly Flow in Software Defined Network", IEEE Access, vol. 6, pp. 27809-27817, 2018.

[19] D. Jianjian, T. Yang and Y. Feiyue, "A Novel Intrusion Detection System based on IABRBFSVM for Wireless Sensor Networks", Procedia Computer Science, vol. 131, pp. 1113-1121, 2018.

[20] D. Santoro, G. Escudero-Andreu, K. Kyriakopoulos, F. Aparicio-Navarro, D. Parish and M. Vadursi, "A hybrid intrusion detection system for virtual jamming attacks on wireless networks", Measurement, vol. 109, pp. 79-87, 2017.

[21] Lin, Kai, et al. "Node scheduling for all-directional intrusion detection in SDR-based 3D WSNs." IEEE Sensors Journal 16.20 (2016): 7332-7341.

[22] S. Shamshirband et al., "Co-FAIS: Cooperative fuzzy artificial immune system for detecting intrusion in wireless sensor networks", Journal of Network and Computer Applications, vol. 42, pp. 102-117, 2014.

[23] Paul, Angshuman, et al. "Improved random forest for classification." IEEE Transactions on Image Processing 27.8 (2018): 4012-4024.

[24] Mirjalili, Seyedali, et al. "Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems." Advances in Engineering Software 114 (2017): 163-

[25] Atiku Abubakar and Bernardi Pranggono, "Machine Learning Based Intrusion Detection System for Software Defined Networks",2017 Seventh International Conference on Emerging Security Technologies (EST), pp.138-143,2017.

[26] Bosh, A., Zisserman, A., Munoz, and X.: "Image classification using Random Forests and ferns". In: IEEE ICCV2007.

[27] Indira K, Christal Joy E, "Energy Efficient IDS for Cluster-Based VANETS", Asian Journal of Information Technology, vol 14(1) ,2015, 37-41

[28] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[29] Indira K, Christal Joy E, "Prevention of Spammers and Promoters in Video Social Networks using SVM-KNN", International Journal of Engineering and Technology, Vol 6, No.5, Oct – Nov 2014, Pg 2024-2030.

[30] Imran Memon, Ibrar Hussain, Rizwan Akthar, Gencai Chen, "Enhanced privacy and authentication: An efficient and secure anonymous communication for location based service using asymmetric cryptography scheme", wireless personal communication, 2015.

[31] Abirami Devaraj, Karunya Rathan, Sarvepalli Jaahnavi, K Indira, "Identification of Plant Disease using Image Processing Technique", International Conference on Communication and Signal Processing (ICCSP) 2019.

[32] K .Indira, U.Sakthi, "Security issues, countermeasures and dynamic scheduling for SDWSN", 2nd International Conference on signal processing and communication (ICSPC), 2019.

[33] K. Indira, P. Ajitha, V.Reshma, A.Tamizhselvi, "An efficient secured routing protocol for Software Defined Internet of Vehicles", International Conference on Computational Intelligence in Data Science (ICCIDS), 2019.