Improved the Detection Ratio of Cyber Attack using Feature Reduction based on Support Vector Machine and Glowworm Optimization

Himanshu Shroti
M.E. Scholar, Dept. of CSE,
TRUBA Institute of Engineering and Information Technology,
Bhopal (M.P.), India

Kaptan Singh
Assistant Professor,
Dept. of CSE,
TRUBA Institute of Engineering and Information Technology,
Bhopal (M.P.), India

Amit Saxena
Associate Professor,
Dept. of CSE,
TRUBA Institute of Engineering and Information Technology,
Bhopal (M.P.), India

ABSTRACT
The swarm intelligence plays vital role in feature reduction process in cyber-attack detection. The family of swarm intelligence gives bucket of algorithm for the processing of feature reduction such as ant colony optimization, particle swarm optimization and many more. In family of swarm new algorithm is called glowworm optimization algorithm based on the concept of luciferin. The luciferin collects the similar agent of glow and proceeds the minimum distance for the processing of lights. Such concept used for the reduction of feature in cyber-attack classification. The reduce attribute classified by well know classifier is called support vector machine. The combination of support vector machine and glowworm swarm optimization performs very well in compression of pervious feature reduction technique. The proposed algorithm is implemented in MATLAB software, for the validation of algorithm used KDDCUP99 dataset.

Keywords
Cyber Attack, Feature Reduction, GSO, SVM

1. INTRODUCTION
Now a day's internet based services faced a problem of cyber threats and attack. The cyber-attack performs illegal activity over computer and network. The cyber-attack damages computer software and meaningful information over the internet communication. For the detection and prevention of cyber-attack various approach are used such as system level approach and algorithm level approach. The system level approach used hardware based firewall and intrusion detection system. The intrusion detection system is very famous tools for the detection and prevention of cyber-attack. The intrusion detection process is slow detection and compromised with the performance of cyber-attack data. In current scenario various authors used feature reduction technique for the improvement of cyber detection. The feature reduction process used various algorithms such as particle swarm optimization, ant colony optimization and many more swarm based algorithm. For the reduction of feature also used some pattern recognition based algorithm such as PCA, LDPA and neural network tools. The reduce feature set of intrusion increases the efficiency of intrusion detection system. In this paper proposed an efficient feature reduction technique based on glowworm optimization technique. The glowworm optimization technique works with combination of support vector machine. The support vector machine is well known feature based classification technique. In GSO, a swarm of agents are initially randomly distributed in the search space so that they are well dispersed. In fact, the agents start from the edge of the search space when they entered an unknown environment. Therefore, in this paper, all the agents are initially placed in the edge of the search space. GSO algorithm uses a fixed step-size movement mechanism. The applied glowworm optimization technique reduces the static nature of attribute in intruder file. The reduce feature set proceed for the process of classification. The process of classification used support vector machine. The support vector machine is binary regression tools used for the process for data categorization. For the process of classification in support vector machine used RBF kernel function as margin separate. For the evaluation of performance used KDDCUP99 dataset. The KDDCUP99 dataset obtained from MIT laboratory. The KDDCUP99 dataset contains 7 lacks instance of data. The 7 lacks instance contains all types of attack data and normal communication data. Rest of this paper is organized as follows In Section II. Discuss about glowworm optimization and support vector machine. Section III. Discuss the proposed algorithm. Section IV. Discuss Experimental analysis. Section V. Discuss about Comparative result analysis. Finally, Concluded in Section VI.

2. GLOWWORM OPTIMIZATION TECHNIQUE & SUPPORT VECTOR MACHINE
The glowworm optimization technique and support vector machine work combined and improved the performance of cyber-attack detection.

The KDDCUP99 dataset is basically mixed categories data. The mixed categories data cannot directly proceed for the feature reduction. For the feature reduction, the data transformation is required. The data transformation process used the min-max algorithm for the process of data conversion. The min-max algorithm converts the complete data in form of numeric data.

For the process of feature reduction the transform data passes through glowworm optimization algorithm. Initially all data of KDDCUP99 is distributed in from of glowworm and process for the local decision. The mapped data designed the objective function J(xi(t)) at its current location xi(t) into α luciferin value li and broadcasts the same within its neighborhood. The set of neighbor (Ni(t)) of glowworm i consists of those glowworm that have relatively higher luciferin value that are located within a dynamic decision domain and updating by formula 1 at each iteration.

Local decision range update is given by equation 1

$$r_d^i(t+1) = min\left\{rs, max\{0, r_d^i(t) + \beta(nt - |Ni(t)|)\}\right\} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

And $r_d^i(t+1)$ is glowworm is local decision range at the t+1 iteration, rs is the sensor range, nt is the neighborhood range. The number of glow in local decision range is given by equation (2)

$$N_{i(t)=\{j:\|xi(t)-xi(t)\|<r_d^i ; li(t)<li(t)\}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

And xi(t) is the glowworm I position at the t iteration(t) is the glowworm i luciferin at the t iteration the set of neighbor of glowworm i consist of those glowworm that have relatively higher luciferin value and that are located within dynamic decision domain whose range r_d^i is bounded above by a circular sensor range.

Each glowworm is given in equation (3)

$$p_{ij\,(t)=\frac{l_{i(t)-l_{i(t)}}}{\sum_{k\in Ni\,(t)} lk(t)-li\,(t)}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

Movement update is given in equation (4)

$$x_{i(t+1)=xi(t)+s\left(\frac{sj\,(t)-xi(t)}{\|xj\,(t)-xi(t)\|}\right)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (4)$$

Luciferin update is given in equation (5)

$$l_{i(t)=(1-\rho)li(t-1)+\gamma j\,(xi(t))} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (5)$$

And li(t) is a luciferin value of glowworm i at the t iteration, P belong (0,1) lead to the reflection of the cumulative kindness of the path followed by the glowworm in their current luciferin values the parameter Y only scale the function values, J(xi(t)) is the value of test function. Finally gets the reduce feature. The reduce feature set pass through support vector machine and finally cyber-attack detected.

We begin by discussing a soft margin SVM learning algorithm written by Cortes, which is sometimes called c-SVM. This SVM classifier has a slack variable and penalty function for solving non-separable problems. First, given a set of points $x_i \in R^d$, i=1,……,l and each point x_i belongs to either of two classes with the label $y_i \in \{+1,-1\}$.These two classes can be applied to anomaly attack detection with the positive class representing normal and negative class representing abnormal. Suppose there exists a hyper-plane $W^T x_i$ + b = 0 that separates the positive examples from the negative examples. That is, all the training examples satisfy:

$$W^T x_i + b \geq +1 \; for\; all\; x_i \in P$$

$$W^T x_i + b \geq -1 \; for\; all\; x_i \in P \qquad (1)$$

W^T s an adjustable weight vector, x_i is the input vector and is the bias term.

Equivalently:

$$y_i.(W^T.x_i - b) \geq 1 \forall i, = 1 \dots N \qquad (2)$$

In this case, we say the set is linearly separable.

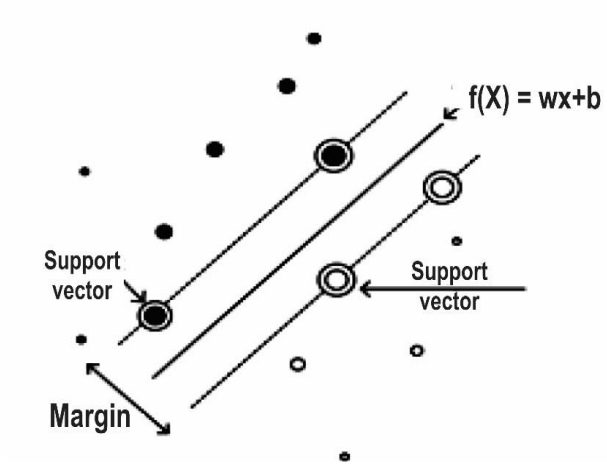

Figure 1: Separable hyper-plane between two datasets

3. PROPOSED ALGORITHM

In this section discuss the proposed algorithm for feature reduction of intruder file using glowworm swarm optimization and detection using support vector machine. The glowworm optimization algorithm based on the concept of shared neighbor for the collection of glow.

Step 1 Transformation of KDDCUP99 Dataset

Let us consider $\{D\,(i,j)\mathrm{I}i = 1,2,\dots\dots,n\,; j = 1,2,\dots,p\}$ is the KDDCUP99 data transform for the processing of feature reduction, the transform data mapped according to glowworm map function by equation (1) and equation (2)

For mapping, glowworm swarm processing.

$$G(i,j) = \frac{(x^*(i,j) - x_{min}\,(j))}{(x_{max}\,(j) - x_{min}\,(j))} \qquad (1)$$

Mapping feature data validated the attribute of agent for the processing of reduction:

$$R(i,j) = \frac{((x_{max}\,(j) - x^*(i,j))}{(x_{max}\,(j) - x_{min}\,(j))} \qquad (2)$$

Where $x_{min}\,(j)$the minimum agent for the reduction j, and $x_{max}\,(j)$is the dynamic change attribute for the processing of reduction.j.

Step 2 Calculate the reduces feature set for the processing of next data$F(a)$.

$\{R(i,j)\mathrm{I}j = 1,2\dots.,p\}$ Is mapping of data for the processing of grouping for the classification$k(i)$ through Grouping $G = [g(1), g(2), \dots\dots, g(n)]$ as:

$$K(i) = \sum_{j=1}^{n} g(j)k(i,j), \quad i = 1,2,\dots\dots n \qquad (3)$$

Then, $k(i)$ is the relative data for the processing of classification.

The derivation of SVM class for the categorization of data:

$$C(K) = L_z T_z \qquad (4)$$

Where $L_z$is the level derivation of relation data of $R(i)$; $T_z$is the training sample for define class in formula (5):

$$\begin{cases} L_z = \sqrt{\dfrac{\sum_{i=1}^{n}(z(i) - E(z))^2}{(n-1)}} \\ DT_z = \sum_{i=1}^{n}\sum_{j=1}^{n}\big(R - r(i,j)\big)u(R - r(i,j)) \end{cases} \qquad (5)$$

Step 3 Defining $R(z(k), z(h))$as the relative class of support vector machine

$$d(z(k), z(h)) = \sqrt{(z(k) - z(h))(z(k) - z(h))} = \sqrt{(z(k) - z(h))^2} \quad (6)$$

$k = 1, 2, \ldots\ldots\ldots, N; h = 1, 2, \ldots\ldots\ldots, N$

$N(n \geq N \geq 2)$ is evaluation level of class. And $D_q(q = 1, 2, \ldots\ldots, N)$ is used to training pattern of data of groupG_q($q = 1, 2, \ldots\ldots, N$),

Step 4 Determining glowworm agent

$$\begin{cases} s.t. \sum_{j=1}^{p} a^2(j) = 1 \\ 1 \geq a(j) \geq 0 \end{cases} \quad (7)$$

Step 5 Finally data are categorized in DOS, PROB, U2R and R2L

Step 6 Measure the performance precision, recall and accuracy.

4. EXPERIMENTAL ANALYSIS

For the validation of feature reduction technique based on glowworm swarm optimization and support vector machine used MATLAB software and personal computer for the simulation purpose. For the validation of proposed algorithm used KDDCUP99 dataset. The KDDCUP99 data set contains 7 lack instance data. These data contain a combination of normal activity of network and attack activity of network. The KDDCUP99 data set is collection of different types of attack data such as DOS, Prob, U2R, R2L and some other attack in limited scope. These categories of attack contain some other group of attack. The main categories of attack define here in form of Table.

Table 1. Different types of attacks in kdd99 dataset

Four main class of attack	Overall 22 categories of attack
Denial of Service (DoS)	back, land, neptune, pod, smurt, teardrop
Remote to User (R2L)	ftp_write, guess_passwd, imap, multihop, phf,spy, warezclient, warezmaster
User to Root (U2R)	buffer_overflow, perl, loadmodule, rootkit
Probing(Information Gathering)	ipsweep, nmap, portsweep, satan

For the evaluation of result used some standard parameter such as Precision, Recall and Accuracy. The Precision, Recall and Accuracy measure the performance of proposed algorithm instead of previous algorithm. The previous algorithm works on the basis of common attribute reduction process. The common attribute reduction process compromised with the selection of parameter for the analysis.

Precision = $\frac{TP}{TP+FP}$, Recall = $\frac{TP}{TP+FN}$, Accuracy = $\frac{TP+TN}{TP+TN+FN+FP}$

Table 2. Shows the performance evaluation of given input value such as 42, 10, 7 and 5 for the classification method such as KNNG and Proposed method.

o. of attribute	Attribute name	Method	Precision	Recall	Accuracy	Classification ratio
42	42	KNNG	89.79	86.27	83.81	85.81
		Proposed	96.30	89.80	88.06	88.94
42	10	KNNG	91.55	88.03	85.57	87.57
		Proposed	98.06	91.59	89.82	89.62
42	7	KNNG	93.26	87.29	88.23	89.67
		Proposed	96.78	91.56	91.36	91.58
42	5	KNNG	94.87	90.48	91.28	91.46
		Proposed	98.79	94.78	94.56	94.26

5. COMPARATIVE RESULT ANALYSIS

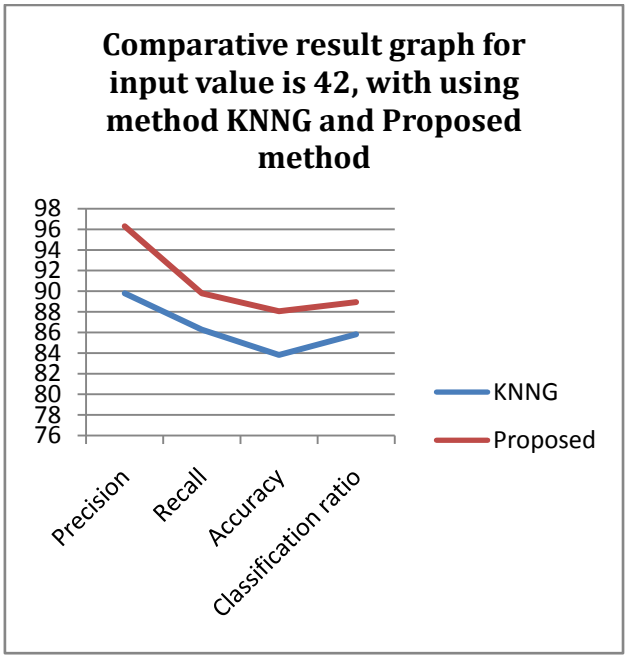

Figure 2: Shows that the comparative result graph for the KNNG and Proposed method and find the Classification Precision, Recall, Accuracy and Classification Ratio for the given number of input value, and the number of given input value is here 42.

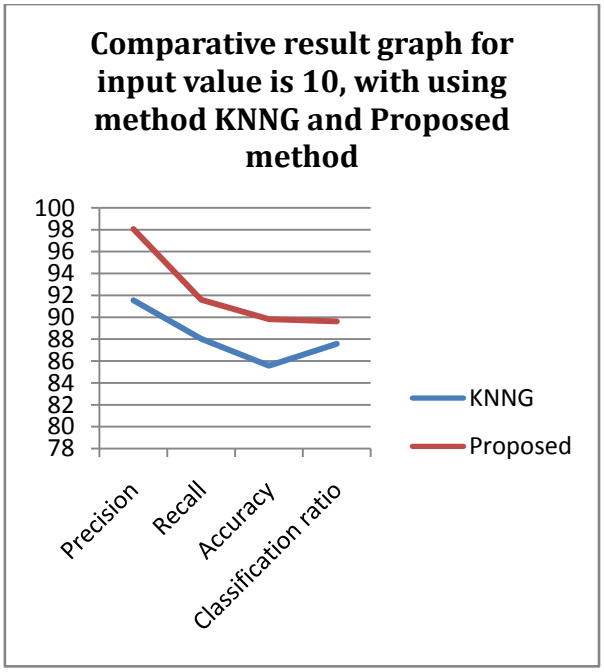

Figure 3: Shows that the comparative result graph for the KNNG and Proposed method and find the Classification Precision, Recall, Accuracy and Classification Ratio for the given number of input value, and the number of given input value is here 10.

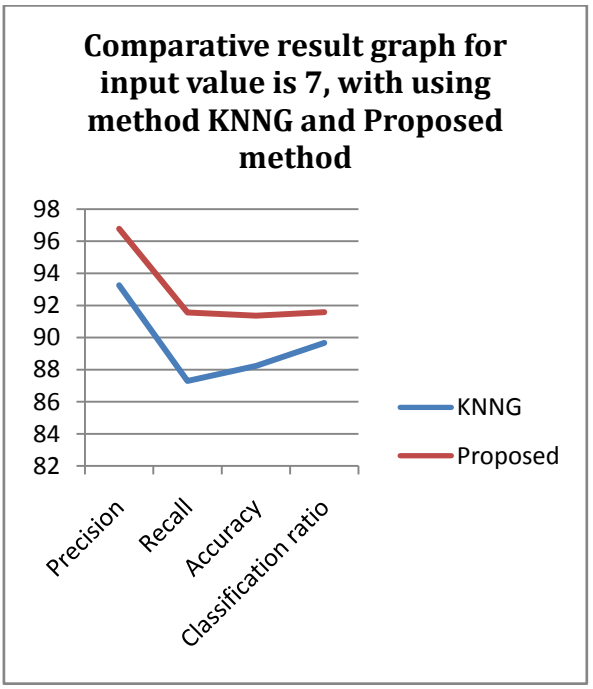

Figure 4: Shows that the comparative result graph for the KNNG and Proposed method and find the Classification Precision, Recall, Accuracy and Classification Ratio for the given number of input value, and the number of given input value is here 7.

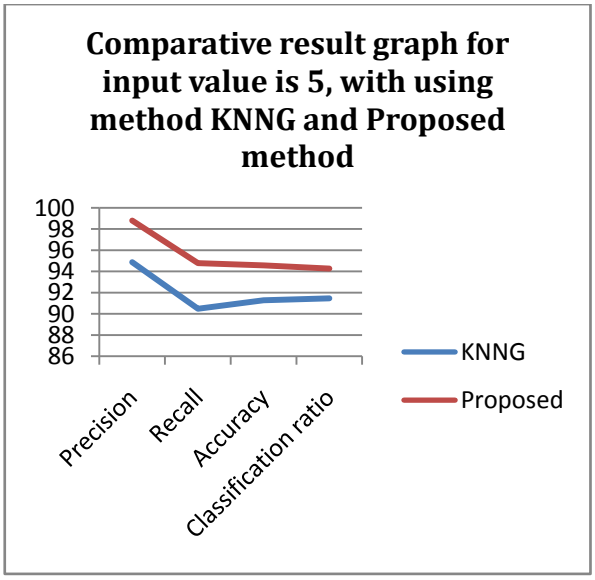

Figure 5: Shows that the comparative result graph for the KNNG and Proposed method and find the Classification Precision, Recall, Accuracy and Classification Ratio for the given number of input value, and the number of given input value is here 5.

6. CONCLUSION & FUTURE SCOPE

The efficiency and scalability of cyber-attack data depends on the reduction of feature. The process of feature reduction improves the capacity of cyber-attack classification and detection of attack. The scope of data and the diversity of attribute is minimum the detection ratio is approx. 100% in case of the diverse attribute of KDDCUP99 dataset the classification ratio is bit low. The combination of glowworm swarm optimization and support vector machine algorithm is better than KNNGA process of feature reduction and classification. The proposed algorithm is very efficient for dynamic attribute for the classification problem. The detection and classification process is better than previous method. In future, uses multi agent glowworm optimization algorithm.

7. REFERENCES

[1] Alvaro A. Cárdenas, Robin Berthier, Rakesh B. Bobba, Jun Ho Huh, Jorjeta G. Jetcheva, David Grochocki and William H. Sanders "A Framework for Evaluating Intrusion Detection Architectures in Advanced Metering Infrastructures", IEEE, 2014, Pp 906-915.

[2] Junho Hong, Chen-Ching Liu and ManimaranGovindarasu "Integrated Anomaly Detection for Cyber Security of the Substations", IEEE, 2014, Pp 1-11.

[3] Preeti Singh and Amrish Tiwari "An Efficient Approach for Intrusion Detection in Reduced Features of KDD99 using ID3 and classification with KNNGA", IEEE, 2015, Pp 445-452.

[4] Gaby AbouHaidar and CharbelBoustany "High Perception Intrusion Detection Systems Using Neural Networks", IEEE, 2015, Pp 497-501.

[5] D.P.Gaikwad and Ravindra C. Thool "Intrusion Detection System Using Bagging Ensemble Method of Machine Learning", IEEE, 2015, Pp 291-295.

[6] Robin Berthier and William H. Sanders "Specification-based Intrusion Detection for Advanced Metering Infrastructures", IEEE, 2011, Pp 184-193.

[7] Robin Berthier and William H. Sanders "Monitoring Advanced Metering Infrastructures with Amilyzer", IEEE, 2013, Pp 1-13.

[8] Adel NadjaranToosi and Mohsen Kahani "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers", Computer Communications, 2007, Pp 2201-2212.

[9] NasimBeigiMohammadi, Jelena Misic, Vojislav B. Misic and HamzehKhazaei "A framework for intrusion detection system in advanced metering infrastructure", SECURITY AND COMMUNICATION NETWORKS, 2012, Pp 195-205.

[10] Massimo Ficco, Salvatore Venticinque and Beniamino Di Martino "MOSAIC-Based Intrusion Detection Framework for Cloud Computing", COMPUTER SYSTEMS SCIENCE AND ENGINEERING, 2012, Pp 1-17.

[11] Mustafa Amir Faisal, Zeyar Aung, John R. Williams and Abel Sanchez "Securing Advanced Metering Infrastructure Using Intrusion Detection System with Data Stream Mining", Springer-Verlag Berlin Heidelberg, 2012, Pp 96-111.

[12] Vincenzo Gulisano, Magnus Almgren and Marina Papatriantafilou "Online and Scalable Data Validation in Advanced Metering Infrastructures", IEEE, 2014, Pp 1-6.

[13] Gianluigi Folino, Clara Pizzuti and GiandomenicoSpezzano "GP Ensemble for Distributed Intrusion Detection Systems", Springer Berlin Heidelberg, 2005, Pp 54-62.

[14] Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin and Wei-Yang Lin "Intrusion detection by machine learning: A review", EXPERT SYSTEMS WITH APPLICATIONS, 2009, Pp 11995-12000.

[15] Edward Guillen, Jhordany Rodriguez and Rafael Paez "Evaluating Performance of an Anomaly Detection Module with Artificial Neural Network Implementation", International Scholarly and Scientific Research & Innovation, 2013, Pp 1484-1490.

[16] Mustafa Amir Faisal, Zeyar Aung, John R. Williams and Abel Sanchez "Data Stream-based Intrusion Detection System for Advanced Metering Infrastructure in Smart Grid: A Feasibility Study", IEEE, 2014, Pp 1-14.

[17] Yu-Li Zhang, Xiao-Ping Ma, Ying Gu and Yan-Zi Miao "A Modified Glowworm Swarm Optimization for Multimodal Functions", IEEE, 2011, Pp 2070-2075.

[18] K.N. Krishnanand and D. Ghose "Glowworm swarm optimization for simultaneous capture of multiple local optima of multimodal functions", Springer, 2009, Pp 87-124.