Mitigation of robustness of intrusion detection system based upon dynamic attacks

[1]Sandeep Kaur, [2] Dr. Amandeep Verma

[1]sandeepdhaliwal31@gmail.com, [2]vaman71@gmail.com

[1,2]Punjabi University Regional Centre for Information Technology and Management

***Abstract*— Network security involves the authorization of access to data in a network, which is controlled by the network administrator. It covers a variety of computer networks, both public and private, that are used in everyday jobs; conducting transactions and communications among businesses, government agencies and individuals. In network security Intrusion Detection System plays an important role to prevent attacks. Existing approaches developed a classifier ensemble method for intrusion detection that is diversified by using two different approaches i.e. different feature sets and training sets or both. The methodology also makes use of resampling technique that emphasizes the attack of rare category. In these approaches the resampling count for accuracy in quite high so that the memory consumption is high and inefficient resource utilization which results into lower throughput. To avoid these drawbacks proposed approach adjust of the ensemble size dynamically according to the size of dataset may be done. That is, decision of number of base classifiers to be used for constructing ensemble should be done dynamically. This may be analyzed by comparing it with existing approach.**

***Keywords*— Intrusion Detection System, Machine Learning, Firefly Optimization.**

I. INTRODUCTION

Internet has become a powerful platform for business, telecommunication, banking and e- commerce. Nowadays it is the only medium that brought the entire world on a common platform to share information, to do business, entertain people and spread education globally. However with the leading graph of internet usage, the graphs of internet threats and attacks have reached its threshold. Internet security thus is a global area of concern, as the attacks through internet are increasing at a massive pace. These attacks affects the normal working of the machine, threatens the confidentiality of the data both in personal and corporate sector.

Apart from the corporate world millions of household are connected to internet nowadays. Every machine, point to point network and distributed networks that is connected through the internet or not, are at risk. If the computer is left unattended, that are connected to the internet, could be easily accessed, misused or injected with unwanted stuff by any suspect [1].

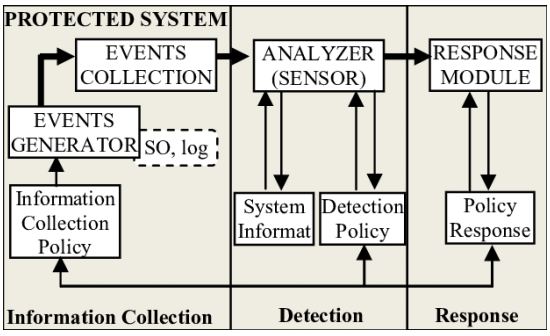

Fig 1: IDS

Intrusion detection system is a secondary wall of defense which monitors the usual behavior of the user for any anonymous or irregular action, either in the network or in the machine. Intrusion detection systems raise alerts for anomaly detection as well as misuse detection. It can be implemented as a centralized as well as distributed set up. It basically monitors the internet log for network activities and application, system and data server logs for machine related actions.

II. MACHINE LEARNING IN INTRUSION DETECTION

Machine learning may be a specific part of artificial intelligence that acquires data from coaching knowledge supported illustrious facts. Machine learning principally focuses on prediction. Machine learning techniques square measure classified into 3 broad classes like supervised learning, unsupervised learning, and reinforcement learning.

- Semi-supervised learning:
- Active learning:
- Unsupervised learning:

III. FIREFLY OPTIMIZATION

Firefly algorithm is metahuristic in nature and inspired by flash behavior of fireflies. Its primary feature is to act as a signal system which can attract other fireflies. It is formulated by considering following conditions:

- All fireflies are unisexual in nature therefore one firefly can be attracted to others
- if a firefly is brightest then it can move randomly
- Attractiveness is in directly proportionate to brightness.

Advantages of Firefly

- High Convergence Rate
- High exploration ability
- Generate optimal Global solution

IV. RELATED STUDY

Jorge L.M. Amaral et al. [3] 2012 developed K nearest neighbor (KNN), call trees, artificial neural networks (ANN) and support vector machines (SVM) were compared thus on the planning for the foremost effective classifier. Four feature alternative ways that} at intervals that were place on utilized so on ensure a reduced set of the foremost relevant parameters. The offered dataset consists of seven doable input picks (FO parameters) of one hundred fifty measurements created in fifty volunteers (COPD, n = 25; healthy, n = 25). The performance of the classifiers and reduced information sets were evaluated by the determination of sensitivity (Se), specificity (Sp) and house below the imagination curve (AUC). Among the studied classifiers, KNN, SVM and ANN classifiers were the foremost adequate, reaching values that amendment a awfully correct clinical designation (Se > eighty seven, Sp> ninety four, and United protection Force of South yank nation > zero.95). the employment of the analysis of correlation as a ranking index of the FOT parameters, allowed USA to vary the analysis of the FOT parameters, whereas still maintaining a high degree of accuracy.

CüneytDirican et al. [4] 2015 created public that With the event of web and mobile technologies, physics, nano technology, advances in medicine, health and digital applications then on speed up mechatronics studies of late. Last World Economic Forum holds a awfully necessary place on the agenda of computing and AI and together the economists like Roubini, Stiglitz place on entered at intervals the discussion of computing and artificial in intelligence impacts on political economy and business. though man of science criticized on the risks throughout this regard, on a usual we've associate inclination to face live witnessing tremendous news and articles in business pages, regarding on these topics and clearly company life and professionals won't resist to those changes. slashing moderately the business terms and work forces, the suggests that of doing business by exploitation new technologies will have serious impacts on the daily vocation and account from these on countries and on world political economy. many things and headlines like discharged relation, Philips Curve, performance, management, CRM Analytics, shopper relationship management, sales, strategic arising with, production, shopping for Power Parity, GDP, inflation, money, Central Banks, industry, coaching, training, accounting, taxes etc. regarding to business and political economy will face serious dangers, hits, change, exposures any as opportunities and gains with the enhancements in AI and computing.

Ming Jiang [5] 2015 analyzed the ability that unit obligatory for a mechanism to serve in very robust and intelligence hard-to-please applications. Above all, a mechanism is there that is known as Distributed Collaboration and Continuous Learning (DCCL) mechanism, as a results of the key capability of a mechanism i.e. a mechanical or associate automatic man, maintained to appreciate applications that unit printed on prime of. By making support of hottest Brobdingnagian information Analytics tools with distributed machine learning technologies that unit integrated as services, a singular DCCL middleware platform is developed to facilitate the realisation of the DCCL mechanism.

A. Medina-Santiago et al. [6] 2014 given the event and implementation of neural management systems in mobile robots in obstacle rejection in real time exploitation silent sensors with refined ways that within which of decision-making in development (Matlab and Processing).

Danilo S. Jodas et al. [7] 2013 given the event of a system to manage the navigation of associate autonomous mobile golem through tracks in plantations. Track footage unit accustomed management golem direction by pre-processing them to extract image decisions. Such decisions unit then submitted to a support vector machine and a unreal neural network thus on ascertain the foremost applicable route. A comparison of the 2 approaches was performed to see the one presenting the foremost effective outcome. the general goal of the project to that this work is connected is to develop a true time golem system to be embedded into a hardware platform.

Christos N. Moridis et. al. [8] planned students mood recognition for on-line self-assessment take a glance at. Exponential logic and formulas were used during this regards. The inputs were student's previous answers and slide bar standing. The exponential logic variables were an entire vary of queries for the net selfassessment take a glance at, student's goal, and slide bar price. applicable feedbacks unit recorded supported current standing of moods of the students. Student's manual selection of their mood exploitation slide bar with none automation is that the limitation of the system.

V. PROPOSED WORK

Intrusion detection is need that is for technical world wherever knowledge is generating and ever-changing at an awfully fast rate. In last decade feature choice is that the science that has given a brand new perspective to analysis within the space of Intrusion Detection System. Objective of this paper is to perform AN analysis and comparison of assorted feature choice techniques with a brand new technique of hybrid Particle Swarm optimisation (PSO). applied mathematics Analysis: during this paper standard filter and wrapper feature choice techniques are explored along side a hybrid PSO technique on the quality KDDCup99 dataset. A comparative analysis is performed over four filter techniques and 2 wrapper based mostly techniques. Four completely different classifiers are compared to pick the one providing smart accuracy on the dataset. Findings: The hybrid PSO feature choice technique offers important improvement in prediction capability as compared to ancient feature choice approaches. Analysis shows that SVM classifier provides higher classification results. SVM is employed as classifier due to its high accuracy. The analysis over four filter and 2 wrapper techniques shows that Hybrid PSO provides higher results with ninety eight.6% accuracy and twenty four feature set.

Step1: Generate Initial population of firefly xi where i=1, 2,.., n, n=number of fireflies (documents).
Step2: Initial Light Intensity, I=total weight of document.
Step3: Define light absorption coefficient γ, initial γ=1.
Step4: Define the randomization parameter α, α=0.2
Step5: Define initial attractiveness $\beta_0 = 1.0$
Step6: While t < Number of iteration
Step7: For i=1 to N
Step8: For j=1 to N
Step9: If (total weight Ii < total weight Ij){
Step10: If Similarity (i, j) >= Threshold {
Step11: Calculate distance between i, j using Eq.11.
Step12: Calculate attractiveness using Eq. 1.
Step13: Move document i to j using Eqs. 2,13,14,15.
Step14: Update light intensity using Eq. 12.
Step15: End For j
Step16: End For i
Step17: Loop
Step18: Rank to find best document.

Fig 2: Firefly algorithm

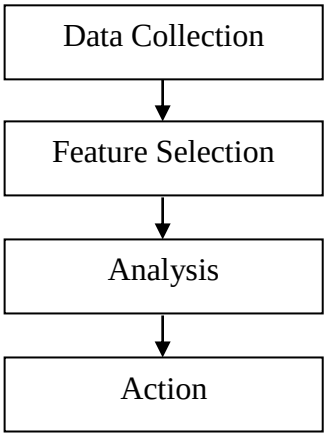

Fig 2: Steps for IDS

Data collection: In this module data is collected by passing data input to Intrusion Detection System. Now data is collected and recorded on file and then analyzed. Network based Intrusion Detection System collects data packets and made mandatory updations.

Feature Selection: In this module features are selected and choose a particular feature from huge data that is available from NIDS is then evaluated. For eg Internet Protocol(IP) extract the data from source and destination node and then take considerable actions.

VI. RESULTS

Accuracy: It is a description of systematic errors, a measure of statistical bias; as these cause a difference between a result and a "true" value, ISO calls this trueness.

Accuracy = (TP+TN)/(TP+TN+FP+FN)

Precision: It is a description of random errors, a measure of statistical variability.

Precision(P) = TP / (TP + FP)

Recall: Recall is defined as amount of related instances that extracted over total related number of instances in the image.

Recall (R) = TP / (TP + FN)

True positive (TP) = the number of cases correctly identified as true

False positive (FP) = the number of cases incorrectly identified as true

True negative (TN) = the number of cases correctly identified as false

False negative (FN) = the number of cases incorrectly identified as false.

Table 1: Comparative Study for Existing and Proposed approaches

of ords	Accuracy (Proposed)	Accuracy (Existing)	Recall (Existing)	Recall (Proposed)
DS1	90	83	0.89	0.96
DS2	90	85	0.86	0.94
DS3	92	88	0.83	0.92

Table 2: Comparative Study for Existing and Proposed approaches

Dataset	Precision (Existing)	Precision (Proposed)	F-Measure (Existing)	F-Measure (Proposed)
DS1	0.9	0.94	0.87	0.92
DS2	0.82	0.92	0.86	0.9
DS3	0.82	0.92	0.83	0.9

Table 1 and 2 is the comparative study for the statistical analysis and Proposed approaches in IDS. From the table it is clear that the Proposed values of results are better than that of statistical analysis. From table 5.1 it is clear that the values of accuracy, recall, precision and F-Measure are better in case of Proposed approaches that are nearly 84, 0.92, 0.93 and 0.87 approx.

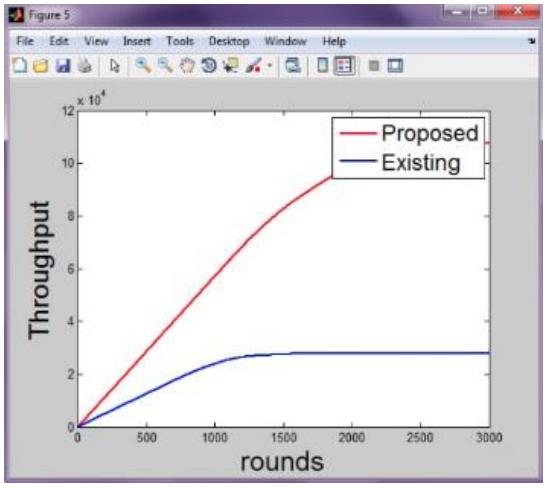

Fig 3: Throughput

Fig 3 is representation of throughput in case of existing and proposed approach. In proposed approach feature set is selected in an efficient manner so that the probability of Intrusion in network is reduced. From the figure it is clear that the throughput in proposed approach is better i.e. approx $11*10^4$ bits where as in case of existing approach it is $3*10^4$ bits.

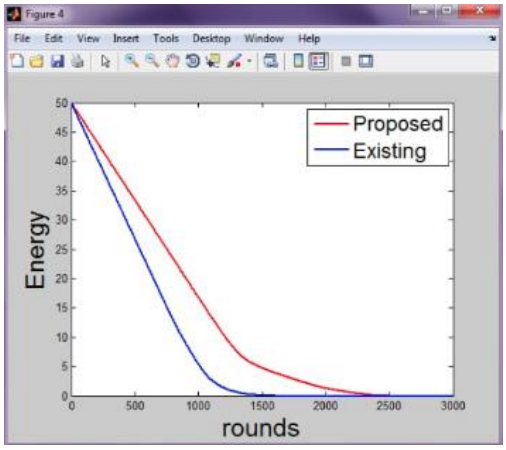

Fig 4: Energy

Fig 4 is representation of residual energy in case of existing and proposed approach. From the figure it is clear that the residual energy in proposed approach is better i.e. approxupto 2500 rounds where as in case of existing approach it is approx 1500 rounds.

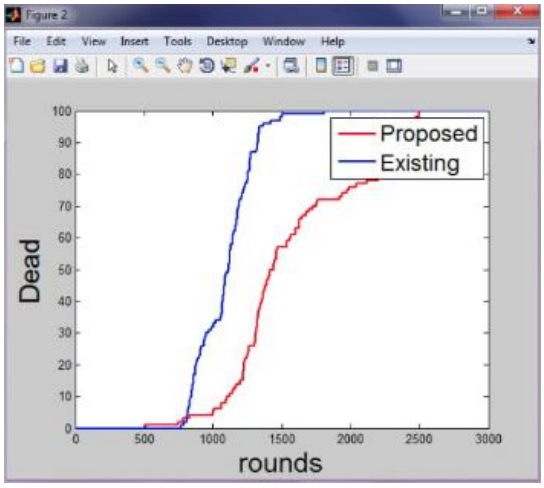

Fig 5: Dead Nodes

Fig 5 is representation of residual energy in case of existing and proposed approach. From the figure it is clear that the dead nodes in proposed approach is better i.e. approxupto 2500 rounds where as in case of existing approach it is approx 1500 rounds.

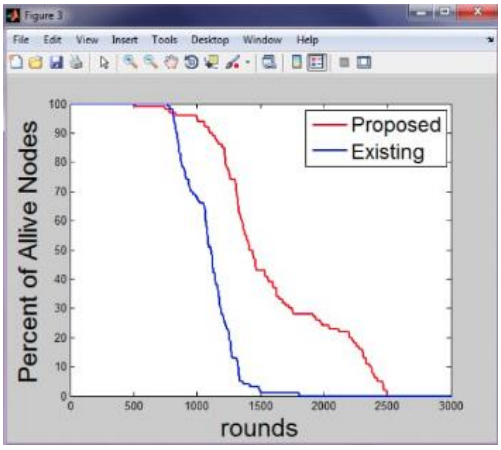

Fig 6: Alive Nodes

Fig 6 is representation of residual energy in case of existing and proposed approach. From the figure it is clear that the alive nodes in proposed approach is better i.e. approxupto 2500 rounds where as in case of existing approach it is approx 1500 rounds.

VII. CONCLUSION

This research provides a framework forhaving a general plan regarding the intrusion detection systemsand additionally provides this analysis work that is takingplace during this field. There are varied IDSs built for the safety of pc systems from threatscaused by the attackers. of these systems are capable ofdetecting attacks within the network and issue alarms oncefound malicious activities. however still there's a desire to try to more add this field as attacks are increasing day byday; what is more, hackers realize new ways that of exploiting thenetwork resources by victimisation numerous evasion techniques.There is a desire for a robust intrusion detection systemwhich can observe all potential attacks as early as potential. Multi-agent technology is that the future technology

during this field because it is a lot of ascendible, strong and may additionally cut backnetwork traffic. the long run work are going to be to develop agentb ased IDS for police investigation attacks within the network.