Development of a GA based extraction technique for mitigation of intrusion attacks

[1]Manpreet Kaur, [2]Dr. Amandeep Verma
[1]raimanpritkaur@gmail.com, [2]vaman71@gmail.com
[1,2]Punjabi University Regional Centre for Information Technology and Management, Phase 7, Mohali

Abstract*— Intrusion detection system is a technique or software with which malicious activities of node in a network can be monitored. Intensive growth and high usage of internet are prone to malicious node activities so a question is raised how to protect network from malicious nodes. In this research a technique is designed in which intrusions are to be classified according to attacks in network. This paper a hybrid technique is used with which one can classify attacks based upon feature selection from pre-defined NSL-KDD data set. In this approach the features are selected according to their weights so that with the help of minimum features attacks can be classified in an efficient manner. The proposed algorithm is performing better than that of existing approach as shown in result section.

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

I. INTRODUCTION

Network security stars with reference to policies, ordinary with user name and password. It is an organizational practice that combines with security of its assets from any intruder and attacker which can hamper the normal operations in an organization. So access to network is provided by effective network security.

Network security is combination of numerous layers of defences in and at network. In this procedure access to network is provided by network administrator only so that any non legitimate node can-not access network and it reduce the chances of attack on network in manual way. To make it automatic intrusion detection system is used which are of 2 types, one is network based and other one is host based.

II. INTRUSION DETECTION SYSTEM

As defined in the previous section Intrusion Detection System is of two types.

Network based IDS

Host based IDS

In network based IDS is used to monitor the network traffic always whenever the network is alive and it protect network from all-over network attacks. It analyze all the packet flow into network and always searches for any kind of suspicious patterns. If any kind of suspicious activity is shown by network then based upon its polarity NIDS intimate network administrator or ban that IP source.

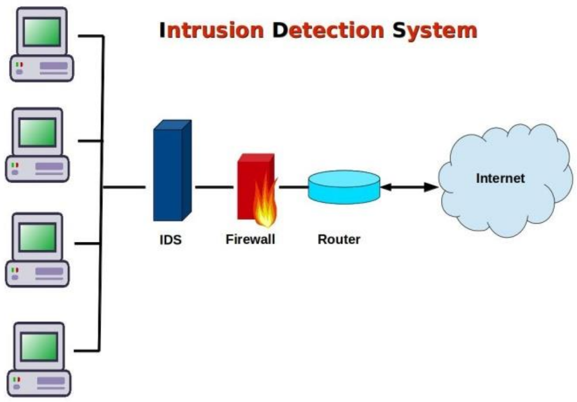

Fig 1: IDS

In case of Host based IDS HIDS analyze all the packet flow in a single system as opposed to NIDS which analyze all the packets on network.

The main difference is only about the analysis of packet. NIDS analyze packets on the network level whereas HIDS analyze packets on the single system level.

On the basis of polarity of attack both system will notify network administrator or ban that particular IP source.

III. GENETIC ALGORITHM

In Genetic rule a listing of candidate answer to a haul is evolved to form a higher answer. every candidate answer incorporates a set of property which can be iterated and updated historically.

A typical genetic rule requires:

- a genetic illustration of the answer domain
- a fitness perform to judge the answer domain.

Initialization:

The population estimate relies upon the character of the issue, however by and large contains numerous heaps of or a huge number of possible arrangements.

Selection

During each successive age, a portion of the present population is reared a fresh out of the plastic new age. Singular arrangements ar choose through a wellness based strategy, wherever fitter arrangements (as estimated by a wellness work) ar for the most part extra presumably to be choose. bound decision ways rate the wellness of each answer and specially pick the best arrangements. distinctive ways rate exclusively an irregular example of the population, in light of the fact that the previous technique could likewise be appallingly long.

Genetic operators:

The following stage is to think of a moment age population of arrangements from those choose through a blend of genetic operators: hybrid (likewise alluded to as recombination), and transformation. for each new response to be made, a join of "parent" arrangements is decided for rearing from the pool chose aforesaid. By assembling a "youngster" answer exploitation the over methods for hybrid and transformation, a fresh out of the box new answer is made which normally shares a few of the attributes of its "parents".

Advantages

- doesn't require any side-effect information (which won't not be possible for a few true issues).
- is speedier and extra temperate when contrasted with the typical ways.
- Has magnificent parallel capacities.
- Optimizes each persistent and unmistakable capacities and conjointly multi-target issues.
- Provides a posting of "good" answers and not just one arrangement.
- continually gets an answer for the issue, that gets higher over the time. Supportive once the inquiry house is to a great degree goliath and there are a larger than average assortment of parameters concerned.

IV. RELATED STUDY

Xiaofeng Zhang et al. [1] anticipated a genuine semi-administered mil system for interruption identification. In particular, the structure embraces Laplacian Support Vector Machine (LapSVM) as its business model and uses data pick up fundamentally based component determination philosophy to zest up the execution. Machine learning (ML) has been wide called a genuine approach for information essentially based interruption location investigation. Particularly, semi-administered mil approaches apply each named and unlabelled information to teach the location display, which can maintain a strategic distance from the high worth of marking information.

Jin vitality et al. [2] 2015 anticipated a half and half security andcompressive detecting based topic for transmission identifier task is given. it's light-weight securitymechanism and along these lines diminishes the standard and vitality utilization of framework. Execution analysisabout security and pressure is applied.The utilization of study strategies like cryptography and hashing for the first half can increment theenergy utilization of sensors, that disturbs the essential fundamental vitality requirement hindrance of wirelesssensor networks (WSNs). to diminish the weight of sensors, pressure region unit regularly utilized. Since the traditionalchaos-based plans don't have all the earmarks of being specifically relevant for WSNs, we tend to tend to blessing a crossover security determination. The hybridsecurity comprises of 8-bit extend disorganized square cryptography and a tumult based message confirmation codes.It intends to plug the insurance and execution of information gathering.

M. Cheng et al. [3] 2014 anticipated partner degreed through A test show a topic whereby hyperchaosand inadequate Fourier change (FrFT) strategies unit of estimation coordinated in AN orthogonalfrequency-division multiplexing (OFDM) uninvolved optical network framework. In degree experiment,both security issues relate degreed transmission execution unit of estimation researched underneath A general frame,and 7.64-Gb/s 16-quadrature-abundancy tweak OFDM information with a four-level encryptionscheme unit of estimation effectively transmitted over a 25-km ordinary single-mode fiber.

EdoardoBiagioni [4] 2014 started to use capacity to supply interpersonalcommunication each finished foundation networks (theInternet), and over impromptu and postponement tolerant networks composedof the cell phones themselves.This network is decentralized among the feeling that it'll functionwithout any framework, however can benefit of framework associations once out there. All social communicationis scrambled and recorded so parcels may even be conveyed bydevices joy to untrusted others. The decentralized model ofsecurity manufactures an adaptable trust network on high of the social networkof human action folks.This interpersonal organization territory unit ordinarily used to arrange bundles to or fromindividuals firmly associated by the informal community. totally extraordinary packetsare organized to support parcels most likely to expend less networkresources.

MuhamedElezi et al. [5] 2015 offered a gathering of reenacted secure information correspondence tunnelstogether with an examination of aftereffects of the speed factors estimated against the insurance through totally unique cryptography protocolsbetween remote LAN's. These cryptography conventions unit of estimation ran onto disseminated inquiries hone changed data functions.The universe of net all by itself is open and unreliable normally. organizations and associations abuse Brobdingnagian possibilities thatinternet offers to frame laptop frameworks to change correspondence and information sharing capacities among their organization platforms.In doing as such, they ceaselessly attempt towards giving a quick, sparing and in the meantime secure operational environment byprotecting their structure resources. Endeavors assemble their network framework with expectation to watch out dependable arrangements toprotect themselves from untrusted and wrongdoing exercises. all through this sense, Virtual private Networks (VPN) unit of estimation primarilyconcerned identifying with information protection. VPNs speak to degree augmentation of a private network made through further decisions like encapsulatingthe information bundles with a header on each end, on the lines of the correspondence also as all through setting communicationtunnels rehearse composite suite of conventions out there.

Harun Ozkisi et al. [6] 2015 delineated that with the ascent of the cell phones, fundamentally the understudies have started to utilize net further effectively. Thisstudy expects to appear at the amount of the college understudies' information of net and on-line applications.The web has moved toward becoming degree irreplaceable a territory of our elegant life as per the data advancements that has beengrowing rapidly. the imperatives identifying with time, place, instrumentation and worth territory unit eased as of late. Accordingly, the frequency,purpose and nature of net utilize zone unit enormously enhanced and conjointly the fluctuate of net clients has been increasedtremendously.

HartiniSaripan et al. [7] 2011 feature a procedure and a preparatory finding of eight different contextual analyses among banks giving net managing an account benefits in Asian country. The investigation, at the beginning uncovered that inside the truth circumstance', the advanced mark innovation is scarcely being received in securing net managing an account exchanges, that has thus formed the degree of the apparatus of the computerized signature law in Asian country. while the Digital Signature Act 1997 has perpetually been recognized along of the pioneers of atechnology-particular authoritative approach, the Act yet, has been extraordinarily presented to fluctuated studies, proposing itsinability to secure on-line exchanges, also as net saving money. Besides, the deficiency of any of its arrangements being tried in theMalaysian courts has thus, advised that the law has degree unimportant application in securing net keeping money exchanges.

V. PROPOSED WORK

Subset generation is comprised of three search approaches, heuristic search, complete search and random search where each generated subset is specified as a candidate subset for evaluation [4]. The process of subset generation is generally is followed by two directional approaches. Firstly, decide the search starting point (or points) i.e. the search direction. Search may proceed in forward direction starting with an empty set and successively adds features one at a time, called sequential forward selection(SFS) , or start with a full set and successively eliminate features one by one, called sequential backward elimination(SBE), or starting from both ends and add and remove features simultaneously (i.e., bi-directional). Another Search strategy is random search which may start with a randomly selected start point and proceed in one direction, so that the search may not be trapped into local optima [5]. Other than the aforementioned general search strategies one can also adapt some specific ones. With N features in the feature space, 2N candidate subsets are generated for evaluation. However, this search space often limits the exhaustive search even with normal N number of features. Therefore, various other strategies have also been investigated: complete, sequential, and random search.

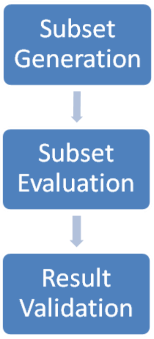

Fig 2: Feature Subset Selection Process

VI. RRESULTS AND DISCUSSION

Table 1: Feature Selected

The dataset used for evaluating and analysis of correlation based feature selection is the same NSL-Kdd Cup 99. The details of the dataset is already discussed in detail in the previous chapter in section 4.1.The dataset that we have taken is KDD 20% and KDD full data sets for training and testing respectively. The respective count of examples in the training and testing dataset are 25192 and 22544. The details of instances under normal, DoS, Probe, U2R and R2L classes for training and testing files are presented in previous chapters.

VII. RESULTS AND DISCUSSION

The dataset used for evaluating and analysis of correlation based feature selection is the same NSL-Kdd Cup 99. The dataset that we have taken is KDD 20% and KDD full data sets for training and testing respectively. The respective count of examples in the training and testing dataset are 25192 and 22544. The details of instances under normal, DoS, Probe, U2R and R2L classes for training and testing files are presented in previous chapters. The dataset contains 41 features with attack class as the 42nd feature. All the 41 features are categorized into basic features, content features, traffic features, and host based features.

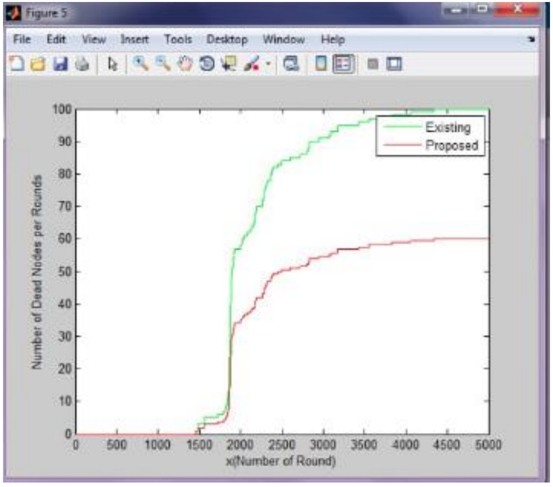

Fig 3: Dead Nodes vs Rounds

Fig 3 is the presentation of number of dead nodes against the number of rounds. This figure is comparative study of proposed technique and existing technique. From the figure it is clear that the dead nodes in proposed technique is less than that of existing technique.

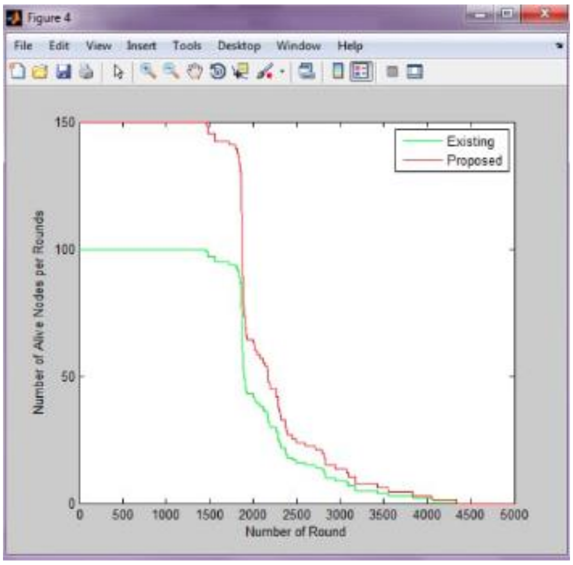

Fig 4: Percentage of Alive Nodes

Fig 4 is the presentation of percentage of alive nodes against the number of rounds. This figure is comparative study of proposed technique and existing technique. From the figure it is clear that the alive nodes in proposed technique are less than that of existing technique.

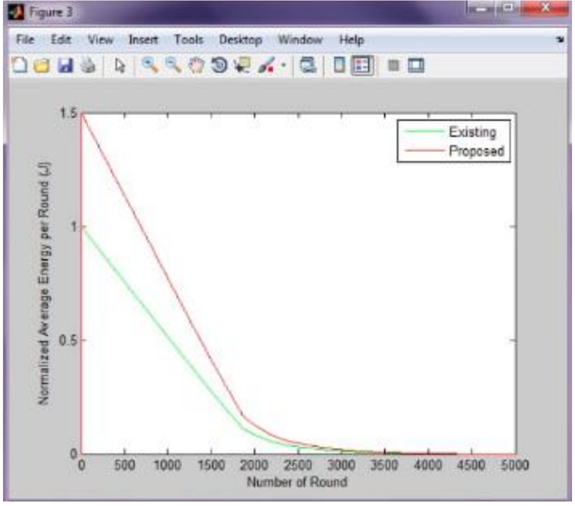

Fig 5 Energy Residual

Fig 5 is the presentation of energy residual against the number of rounds. This figure is comparative study of proposed technique and existing technique. From the figure it is clear that the energy residual in proposed technique is more than that of existing technique.

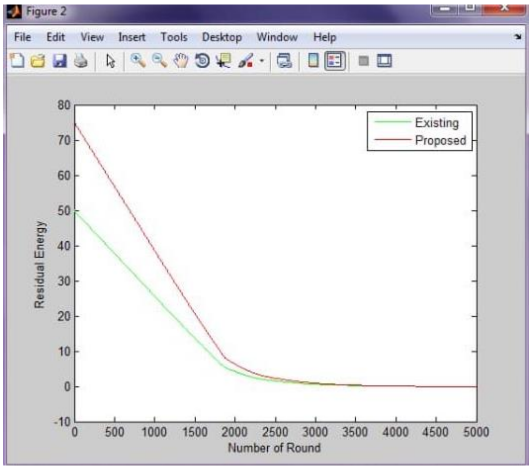

Fig 6: Residual Energy

Fig 6 is the presentation of energy residual against the number of rounds. This figure is comparative study of proposed technique and existing technique. From the figure it is clear that the energy residual in proposed technique is more than that of existing technique.

Table 2 Comparative study of various parameters for both algorithms

Algorithm Parameters	Existing	PDORP
Delay(sec)	80	60
Throughput(packets)	100	110
Residual power	8 rounds	10 rounds
Jitter(sec)	7	3.5
Average Load(bits/sec)	7000	6000

VIII. CONCLUSION

This study gives a structure to having a general game plan as for the intrusion revelation systems and furthermore gives this examination work that is happening in the midst of this field. There are changed IDSs worked for the security of pc systems from perils caused by the aggressors. Of these systems are prepared for distinguishing ambushes inside the network and issue cautions once found pernicious activities. however still there's a need to endeavor to more incorporate this field as strikes are growing well ordered; what's more, software engineers recognize new courses that of manhandling the network resources by misuse different shirking frameworks. There is a need for a generous interference area structure which can observe each and every potential ambush as in front of timetable as potential. Multi-expert advancement is that the future development in the midst of this field since it is an impressive measure of ascendible, strong and may likewise lessen network movement. the long run work will be to make administrator based IDS for police examination ambushes inside the network.

IX. REFERENCES

[1] Xiaofeng Zhang, Jianwei Tian, Peidong Zhu, Jiexin Zhang, "An Effective Semi- supervised Model for Intrusion Detection Using Feature Selection Based LapSVM", IEEE, ISBN: 978-1-5090-5957-7, 2017, page no: 475-480

[2] Jin Qi, Xiaoxuan Hu, Yun Ma, Yanfei Sun, "A Hybrid Security and Compressive Sensing-Based Sensor Data Gathering Scheme", IEEE Access, volume: 3, 2015, page no: 718-724

[3] M. Cheng, L. Deng, X. Wang, H. Li, M. Tang, C. Ke, P. Shum, D. Liu, "Enhanced Secure Strategy for OFDM-PON System by Using Hyperchaotic System and Fractional Fourier Transformation", IEEE Photonics Journal Secure Strategy for OFDM-PON System, volume: 6, 2014, page no: 854-860

[4] Edoardo Biagioni, "Ubiquitous Interpersonal Communication over Ad-Hoc Networks and the Internet", 47th Hawaii International Conference on System Science, 2014, page no: 5144-5153

[5] Muhamed Elezia, Bujar Raufia, "Conception of Virtual Private Networks using IPsec suite of protocols, comparative analysis of distributed database queries using different IPsec modes of encryption", World Conference on Technology, Innovation and Entrepreneurship, Procedia - Social and Behavioral Sciences, volume: 195, 2015, page no: 1938-1948

[6] Harun Ozkisia, Murat Topaloglu, "The University Students' Knowledge of Internet Applications and Usage Habits", 4th World Conference On Educational Technology Researches, WCETR, volume: 182, 2015, page no: 584-589

[7] Hartini Saripan, Zaiton Hamin, "The application of the digital signature law in securing internet banking: some preliminary evidence from Malaysia", Procedia Computer Science, volume: 3, 2011, page no: 248-253

[8] Goel R., Sardana A., Joshi R. C., "Parallel Misuse and Anomaly Detection Model," International Journal of Network Security, volume 14, July 2012, page no: 211-222.

[9] Davis J J, Clark A J., "Data pre-processing for anomaly based network intrusion detection. A review", Computers and Security, volume: 30, 2011, page no: 353-375.

[10] Varun, C., Arindam, B., Vipin, K., "Anomaly Detection, A Survey", ACM Computing Surveys, volume: 41, 2009, page no:1-58.