HYBRIDIZED INTRUSION DETECTION SYSTEM USING GENETIC AND TABU SEARCH ALGORITHM

Oluwakemi Christiana Abikoye [1], Taye Oladele Aro [1], Racheal Oyeranti Obisesan [2], Akinbowale Nathaniel Babatunde [3]

[1] Department of Computer Science, University of Ilorin, Ilorin, Nigeria
[2] Computer Science Department, Kwara State College of Education, Ilorin, Nigeria
[3] Department of Computer Science, Kwara State University, Malete, Nigeria

Corresponding author: Oluwakemi Christiana Abikoye, kemi_adeoye@yahoo.com

ABSTRACT: *As transactions, data communication and information systems are drastically increasing in the society, so many people are connected through internet for e-commerce and other electronic activities. The introduction of internet technology in business brings about great relief in reaching the end users. Also this technology invites numerous security threats of misuses and intrusions. Intrusion detection systems are significant element for network security infrastructure which plays key role in the detection of several attacks along the network. They are several techniques being employed in intrusion detection, but these methods are not completely flawless. In quest for an efficient Intrusion Detection System (IDS), this study employs hybridization technique which involves the Genetic Algorithm and Tabu-search to produce a robust Intrusion Detection System. Evaluation of the system on KKD 99 intrusion database, shows that the performance of proposed hybridized IDS is better than that of Genetic algorithm or tabu search method alone which can significantly detect almost all anomaly data in the computer network.*

KEY WORDS: *Intrusion detection, data communication, Genetic Algorithm, Tabu Search, Information System, Electronic Transaction.*

1. INTRODUCTION

The detection of intrusion system is a monitoring technique that checks networks or systems for several hostile activities, acceptable use of policies, standard security practices and future threats of abuse of computer security rules ([SM07; Deb00]).The system for intrusion detection is used to identify all forms of system network attacks contrary to vulnerable services, data driven attacks on softwares, host based attacks such as privilege escalation, unauthorized logins, access to sensitive files, and malware (Viruses, Trojan Horses, and Worms) that cannot be detected by commonly available firewall systems ([G+09]). Intrusion is referred to any activity that significantly deviates from the normal behaviour to malicious action or operation that provides potential possibility to comprise the integrity, confidentiality and availability of computer resource ([HMB12]). IDS introduces a second level of defence before the use of common methods of securities such as access control system and authentication ([MIS12]).

In the society today, there is increase in number of computer networks compared with electronic businesses, especially on the internet while streaming, video conferencing and chatting ([Dha13]). These different trades introduce many intrusions and anomalies into network system. Intrusion detection systems are powerful tools which are employed by any organization in fight against hackers in order to secure various computing resources. System for Intrusion detection has emerged a recent field of research in the computer security owing to its challenge of certifying that information system will be totally free of security errors ([Deb00]). The insistent interest in IDS technology has constantly improved performance and accuracy of intrusion detection ([KPJ12]).

Numerous techniques of intrusion detections proposed by different organizations to play very important roles in securing infrastructure in network and communication by the use of the internet like antivirus applications, firewall softwares and intrusion detection systems ([Sri08]). Existing firewalls are unable to protect against every form of intrusion, whereas a number of intrusions take advantages of vulnerabilities in computer system.

Information is acquired by intrusion detection system about an information system to perform diagnosis on the security status in order to discover breaches of security and vulnerabilities that might result to potential breaches, which its level has developed to keep up with computer crime advancement. However, performance remains great challenge in the development of IDS. The performance of intrusion detection can be enhanced according to selection of optimal feature subset ([Bij16]), this can be improved by increasing the

recognition rates and diminishing false positives. In this study, multi-level algorithms which are Genetic Algorithm combine with K-Nearest Neighbour and Tabu Search were applied in the development of intrusion detection system in order to obtain a robust intrusion detection system.

2. CATEGORY OF INTRUSION DETECTION SYSTEM

Intrusion Detection System can be divided into two major types subject to the type and information source used to identify security flaws ([PYM14]) . These include:

i. Host Based Intrusion Detection (HIDS):- This is a application software set up on a system which observes activity only on that local system against internal or external attacks ([Bij16]). It connects directly to the operating system with no any information of low-level network traffic. HIDS assesses information located on one or many host systems including contents of files in application and operating systems.

ii. Network Based Intrusion Detection (NIDS): It evaluates information captured from network communications, study the network in which stream of packets travel across as shown in figure 1. The NIDS examines network traffic at the whole layers of the open systems interconnection (OSI) model to conduct decision about goal of the traffic and analyze for suspicious activity ([K+13]). NIDS can be a dedicated hardware machine or a software application running on a computer network.

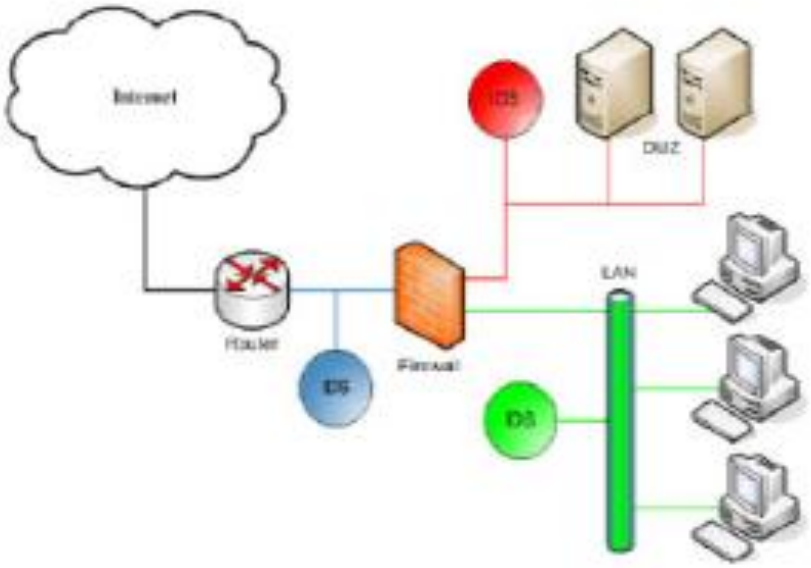


Figure 1. Intrusion Detection System in Computer System ([SRI08])

3. TECHNIQUES IN INTRUSION DETECTION

The two basic types of intrusion detection techniques are Anomaly Detection and Misuse Detection ([KV02]).

(i)Anomaly Detection: - This is the comparison of user's activities with the recognized behaviour of attackers trying to gain access to a system ([RS14]). This approach is designed to reveal the forms of behaviour extremely not normal and anything that widely deviates from it is considered as a potential intrusion ([KI10]).

(ii)Misuse Detection (Pattern Based Detection): - It is referred to as signature based detection, which identifies threats using predefined pattern ([DM15]). The technique employs signature detection to discriminate between anomaly and attacks, it checks for a new activity in the knowledge base ([DL12]).

4. GENETIC ALGORITHM (GA)

This is a computational method motivated by evolutionary biology such as inheritance, selection and recombination ([Deb00]). The technique is based on Darwinian's principle of evolution and survival of fittest to optimize a population of candidate solutions towards a predefined fitness ([Paw13]). GA uses an evolution and natural selection that uses a chromosome-like data arrangement and evolve the chromosomes using selection, recombination and mutation operators. The GA method usually begins with randomly generated population of chromosomes, which denote all possible solution of a problem that are considered by candidate solutions. This technique is used to obtain optimal solutions to specific problem in computer security. According to the attributes of the

problem, different positions (genes) of each chromosome are encoded as bits, characters, or numbers. An evaluation function is used to calculate the "goodness" of each chromosome. Two basic operators; crossover and mutation are used to simulate the natural reproduction and mutation of species during evaluation. The selection of chromosomes for survival and combination is biased towards the fittest chromosomes. The architecture of Genetic Algorithm and the process in Genetic Algorithm are illustrated in figure 2 and figure 3.

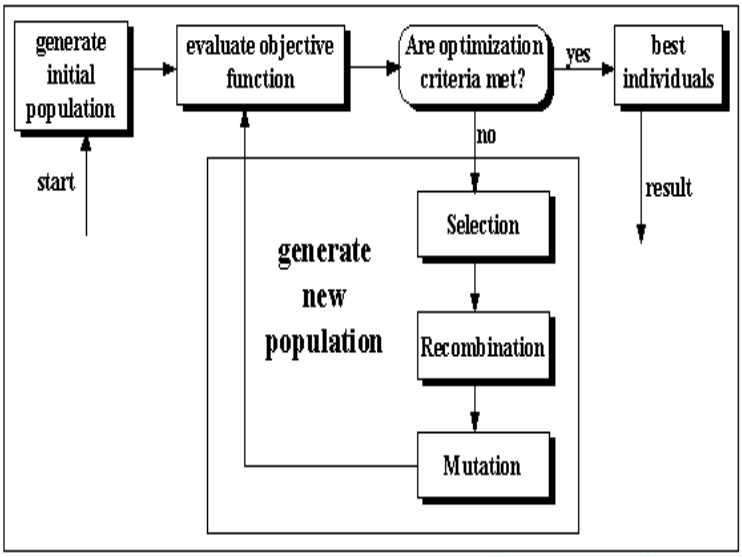


Figure 2: Architecture of IDS Using Genetic Algorithm ([SM07])

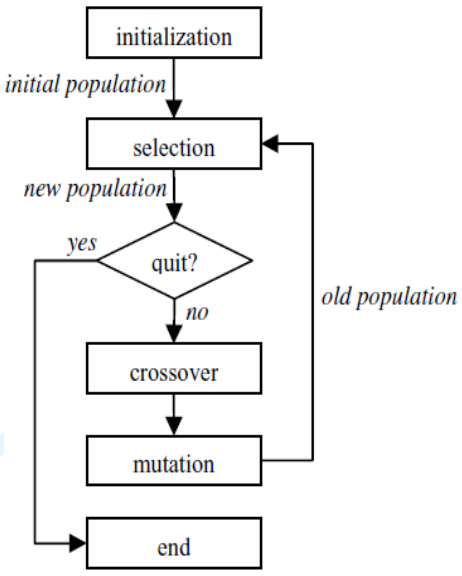


Figure 3. Operation of a Genetic Algorithm ([GZA05])

5. TABU SEARCH ALGORITHM

Tabu Search is a global heuristic approach (metaheuristic) that guides against falling into local optima by producing a special list referred to as tabu ([Mus06]). One of the major components of Tabu Search is its application of adaptive memory, which provides a more flexible search behaviour. Memory-based strategies are therefore the hallmark of tabu search techniques, founded on a quest for "integrating principles," by which alternative forms of memory are appropriately combined with effective strategies for exploiting them.

6. RELATED WORK

Genetic Algorithm can be employed in different ways in Intrusion Detection Systems. Several works have been done using GA for intrusion detection system. These include:

Gong et al.([GZA05]) presented GA for intrusion detection. For the derivation and encoding of rules seven network features including both categorical and quantitative data fields were used. To judge the quality of each rule, a simple but efficient and flexible fitness were applied in the study. Depending on the selection of fitness function weight values, the generated rules can be used to either generally detect network intrusions or precisely classify the types of intrusions.

Purushottam et al. ([PYM14]) proposed available studies on Genetic Algorithm based Network Intrusion Detection Systems (NIDS). The work evalauted several systems implemented using GA, Accuracy, detection rate and false alarm rate against KDD cup database. Implementation of intrusion detection system using Genetic Algorithm ([JG14]), GA was applied to develop set of classification rules from audit data and support confidence frame work was utilized as fitness function to judge the quality of each rule. The generated rules were used to detect and classify network intrusion. Experimental results

showed more effective detection rates on benchmark DARPA datasets.

Dhak and Lade ([DL12]) presented intrusion detection technique based on Geentic Algorithm for identification of malicious packets on the network and ultimately help to block the respective IP addresses. The training was done on the predefined data rules. The testing was done on the entries generated by the firewall system of machine in pfirewall.log file. The proposed system can be integrated with any of the IDS system to improve the efficiency and the performance of the same.

Ehab and Nasereddin ([EN10]) developed a Genetic Algorithm for Intrusion Detection System in a network security. The system considered temporal and spatial information of network connections in encoding network connection information into rules in IDS. The network traffic used for implementing GA was a preclassified data set which differentiated normal network connections from anomalous ones. Dataset was collected using network sniffers (a program used to record network traffic without doing something harmful). The dataset was manually classified based on the knowledge of experts. The rules generated were good enough for filtering new network traffic. The various attributes of network connections used for generating rules were: source IP address, destination IP address, source port number, destination port number, duration, state, protocol, number of bytes sent by originator, number of bytes send by responder. The proposed system contributed towards identification of complex anomalous behaviour.

Ojugo et al. ([O+12]) presented a Genetic Algorithm based approach which employed classification rules derived from network audit data for network intrusion detection. The system employed a set of classification rule derived from network audit data and support confidence framework, which utilized as fitness function to judge the quality of each rule. Effective intrusion detection system using Genetic Algorithm for MANETs ([TU16]). The study designed a three level hybrid framework for IDS/IPS for MANETs. The new design used evolutionary based scheme with GA to detect unknown types of attacks. The proposed system detected signature based attacks and unknown attacks in MANETs.

7. METHODOLOGY

The proposed system combined GA and Tabu Search for embedded hybridization method. Genetic algorithm was applied as a first step, Tabu Search technique was used at final step to enhance the solution quality. In the system implementation, each generated chromosome by GA was a solution which then passed through crossover and mutation operation. After every mutation, solution entered tabu list and among of all the solutions. The best chromosome was selected as a best solution with the better fitness value. The proposed IDS system was divided into two main phases: Pre-calculation Phase and the Detection Phase. Listing 1 depicts major steps in Pre-calculation phase, where a set of chromosome is created using training data. The chromosome set was used in the next phase for the purpose of comparison.

Listing 1. Significant Steps in Pre-calculation Phase

Algorithm: Initialize chromosomes for comparison
Input: Network audit data (for training)
Output: A set of chromosomes

1. Range = 0.125
2. For each training data
3. If it has neighboring chromosome within Range
4. Merge it with the nearest chromosome
5. Else
6. Create new chromosome with it
7. End if
8. End for

Listing 2: It shows major steps in Detection stage, a population was created for test data and going through evaluation analyses (crossover, mutation and selection). The precalculated set of chromosomes were used in this phase to find out fitness of each chromosome of the population.

Listing 2. Significant Steps in Detection

Algorithm: Predict data/intrusion type (using GA)
Input: Network audit data (for testing), Precalculated set of chromosomes Output: Type of data.

1. Initialize the population
2. CrossoverRate = 0.15, MutationRate = 0.35
3. While number of generation is not reached
4. For each chromosome in the population
5. For each precalculated chromosome
6. Find fitness
7. End for
8. Assign optimal fitness as the fitness of that chromosome
9. End for
10. Remove some chromosomes with worse fitness
11. Apply crossover to the selected pair of chromosomes of the population
12. Apply mutation to each chromosome of the population
13. End While

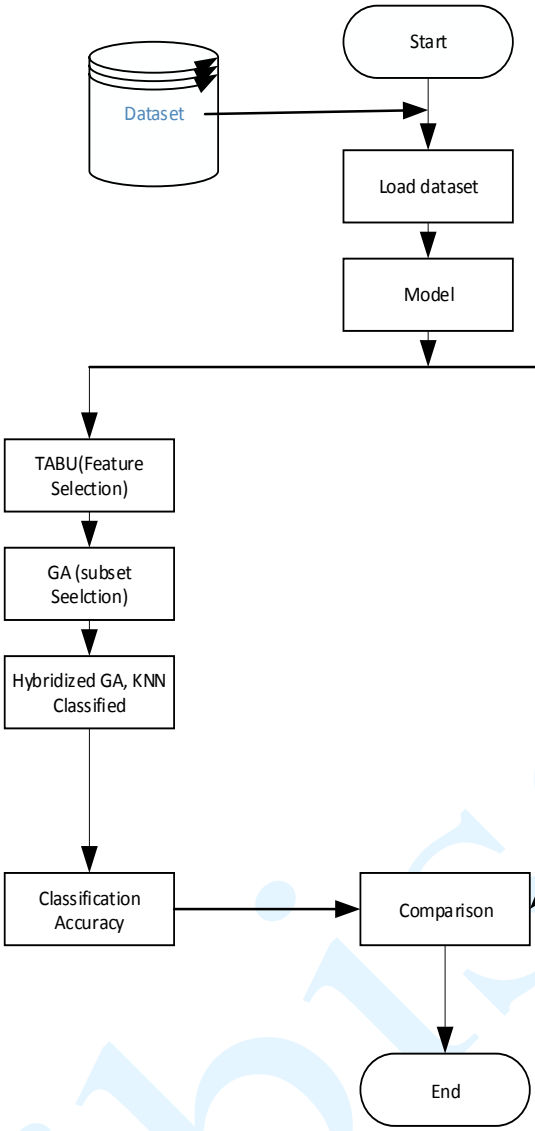


Figure 4. Hybridized Intrusion Detection System

8. DATASET

The dataset used in this work was divided into two groups. A total number of 1499 data extracted from the Standard dataset in KDD Cup 1999 Computer network intrusion detection with two class of attacks considered. The attacks forms the basis of the class category for which the dataset was grouped into DoS and Probe. The cross validation method was applied to validate and predict new instances. 10 folds were used to partition the data. The system was trained for each fold using all the data outside the fold. The performance of the system was tested by using the data inside the fold.

9. EXPERIMENTAL RESULTS AND ANALYSIS

(a) The loaded sample of database for selection using Tabu and GA Algorithm is shown in figure 5.

(b) Figure 6 indicates the loaded database in the tabu search which considers the first stage feature selection. The features with good near solution optimum solution were picked.

(c) Optimization of Genetic Algorithm which involves the GA subset feature and KNN Classification is illustrated in figure 7.

(d) Figure 8 presents a total number of ten relevant and optimal features selected by the Genetic Algorithm.

(e) The confusion matrix shows the prediction of the class based on the three class label, the result of the tabu search for feature selection, GA for subset feature selection and hybridized KNN as shown in figure 9.

% OF correctly classified=83.56%
% OF incorrectly classified=16.44%
TRUE POSITVE RATE=0.8356
FALSE NEGATIVE RATE=0.1644
RECALL=0.731

PRECISION=0.8356
DoS was replaced with a numeric value of 1 for class labelling
TRUE POSITIVE RATE=100%
FALSE POSITIVE RATE=0%
Normal was replaced with a numeric value of 2 for class labelling
TRUE POSITIVE RATE=100%
FALSE POSITIVE RATE=0%
Probe was replaced with a numeric value of 3 for class labelling
TRUE POSITIVE RATE=50.7%
FALSE POSITIVE RATE=49.3%

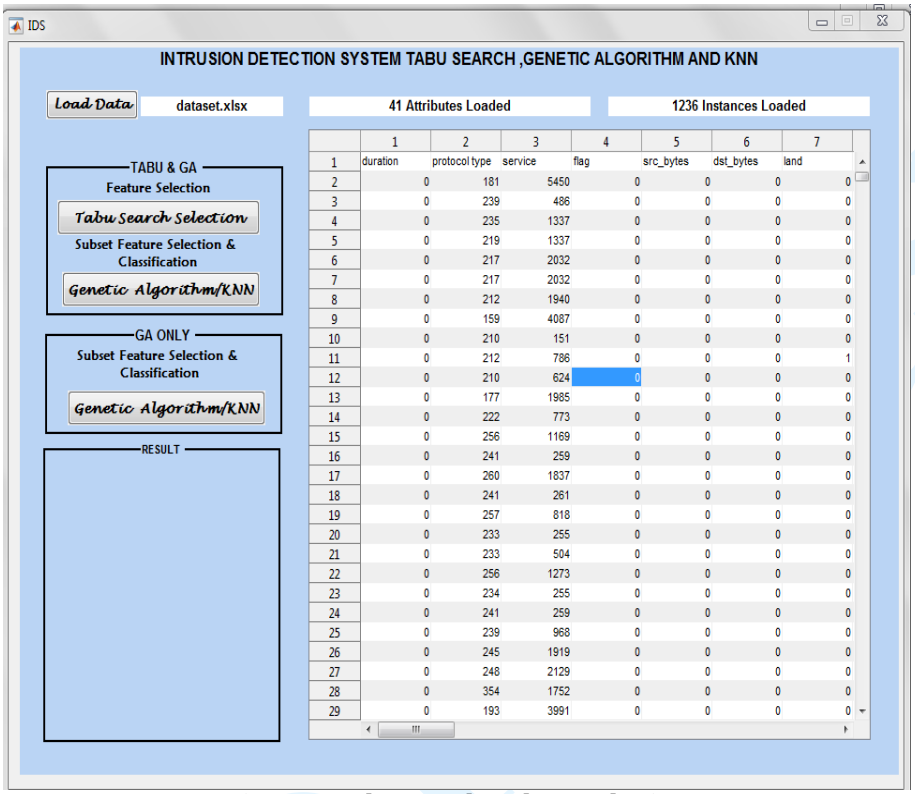


Figure 5. Tabu Search and GA Selection

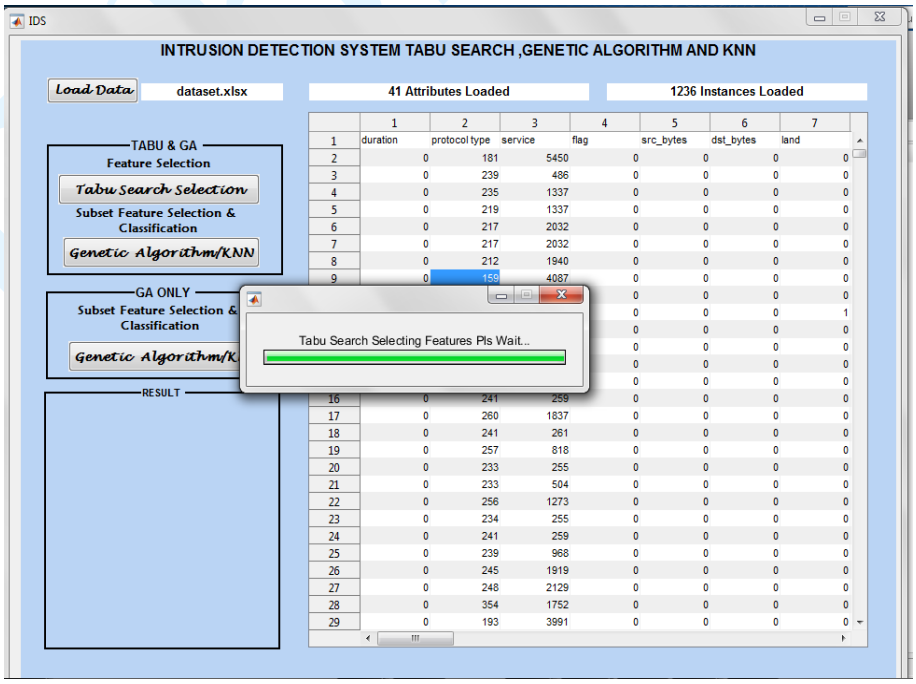
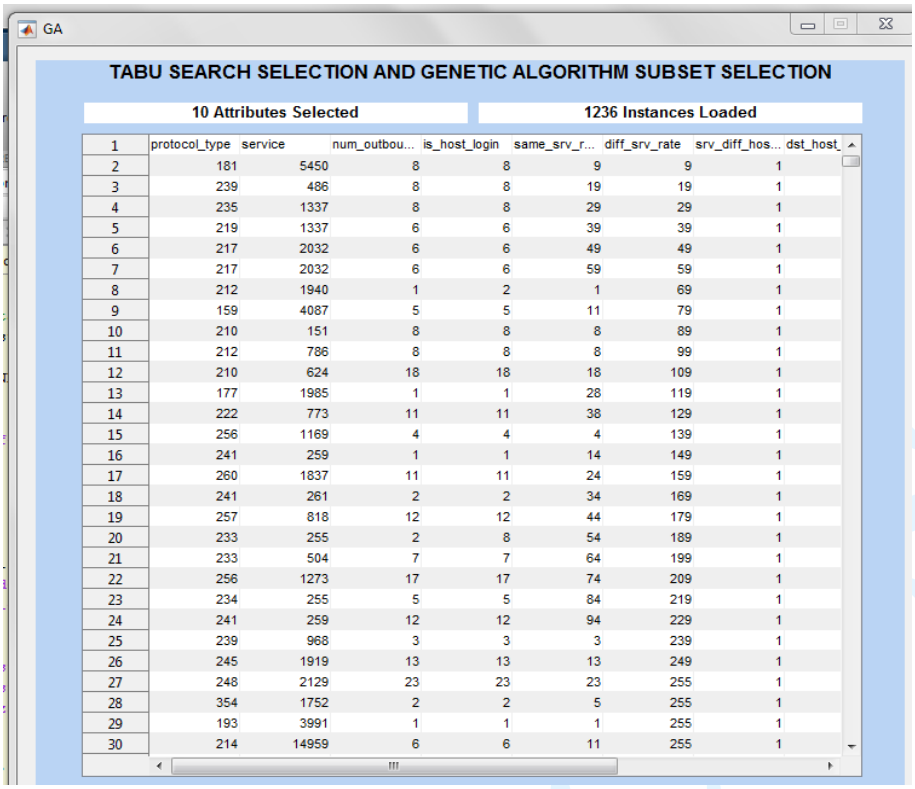


Figure 6. Tabu Search feature selection



	protocol_type	service	num_outbou...	is_host_login	same_srv_r...	diff_srv_rate	srv_diff_hos...	dst_host_
1								
2	181	5450	8	8	9	9	1	
3	239	486	8	8	19	19	1	
4	235	1337	8	8	29	29	1	
5	219	1337	6	6	39	39	1	
6	217	2032	6	6	49	49	1	
7	217	2032	6	6	59	59	1	
8	212	1940	1	2	1	69	1	
9	159	4087	5	5	11	79	1	
10	210	151	8	8	8	89	1	
11	212	786	8	8	8	99	1	
12	210	624	18	18	18	109	1	
13	177	1985	1	1	28	119	1	
14	222	773	11	11	38	129	1	
15	256	1169	4	4	4	139	1	
16	241	259	1	1	14	149	1	
17	260	1837	11	11	24	159	1	
18	241	261	2	2	34	169	1	
19	257	818	12	12	44	179	1	
20	233	255	2	8	54	189	1	
21	233	504	7	7	64	199	1	
22	256	1273	17	17	74	209	1	
23	234	255	5	5	84	219	1	
24	241	259	12	12	94	229	1	
25	239	968	3	3	3	239	1	
26	245	1919	13	13	13	249	1	
27	248	2129	23	23	23	255	1	
28	354	1752	2	2	5	255	1	
29	193	3991	1	1	1	255	1	
30	214	14959	6	6	11	255	1	

Figure 7. Genetic Subset feature and KNN Classification

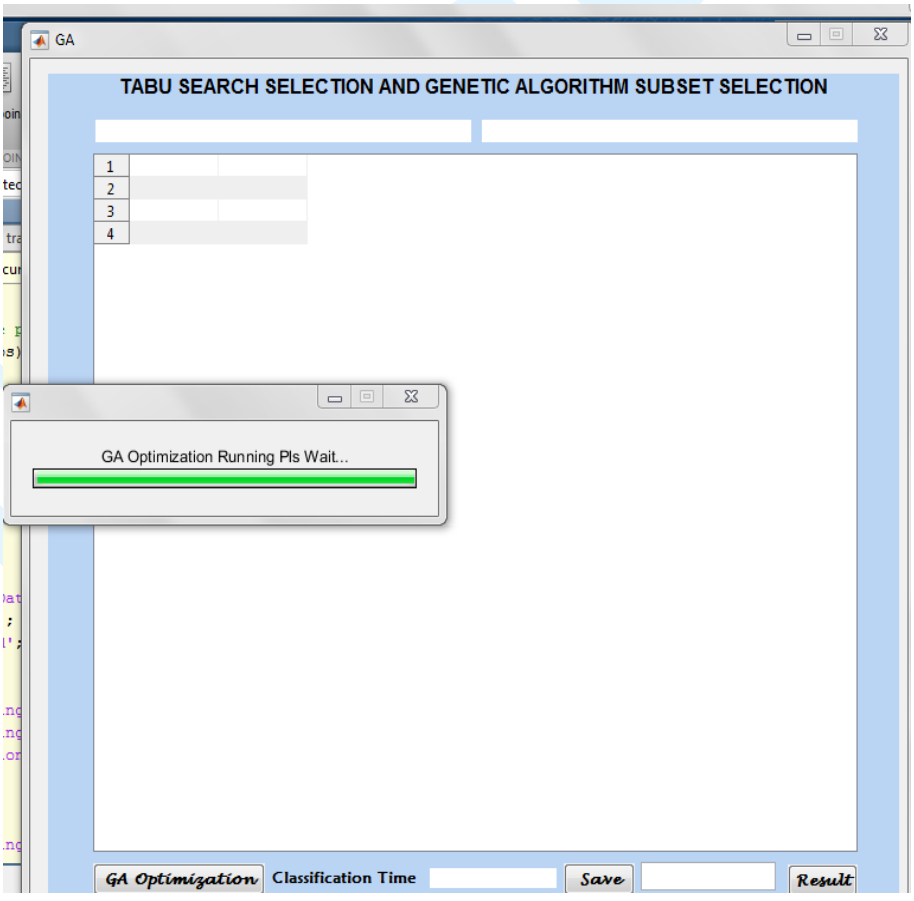


Figure 8. Tabu Search and KNN Selection

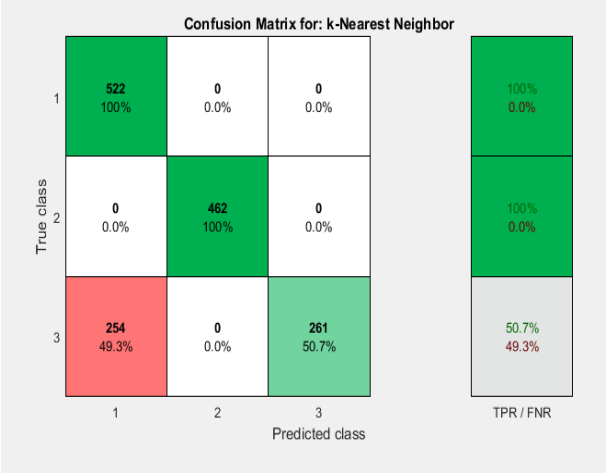


Figure 9. Confusion Matrix for KNN

(f) Figure 10 shows the hybridization of KNN with the Genetic algorithm by creating a fitness function handle that calls KNN classification algorithm so as to minimize the classification error from the KNN.
(g) The phase illustrates feature selection with Genetic Algorithm as illustrated in Figure 11.
(h) The total number of fifteen relevant and optimal features selected by the Genetic Algorithm. The results of the selected features is shown in figure 12
(i)Figure 13 shows the generation of fitness value with GA (variation between the best fitness and mean fitness).
(j) This stage shows result of the GA for feature selection and the hybridized KNN. The confusion matrix illustrates the prediction of the class based on the three class label is shown figure 14.

% OF correctly classified=77.17%
% OF incorrectly classified=28.83%
TRUE POSITVE RATE=0.7117
FALSE NEGATIVE RATE=0.2883
RECALL=0.6205
PRECISION=0.7117

k) Receive operating characteristic is indicated in figure 15.
(l) Comparative Analysis of Tabu Search, Hybridized Genetic Algorithm and KNN.
A comparative analysis was done using the classification time, attribute selected, classification accuracy and misclassification accuracy to evaluate the efficiency of the two states. The tabu search when combined with the genetic algorithm gave a more reduced features than the Genetic Algorithm only. A total number of 10 features were selected when the dataset was passed into tabu and GA for subset feature selection, while for Genetic Algorithm only total features of fifteen were selected. The combination of tabu search and GA also gave a better classification accuracy than the optimization of Genetic Algorithm only. A classification accuracy of 83.56% to 77.17% and misclassification accuracy of 16.44 % to 28.83% as illustrated in Table 1.

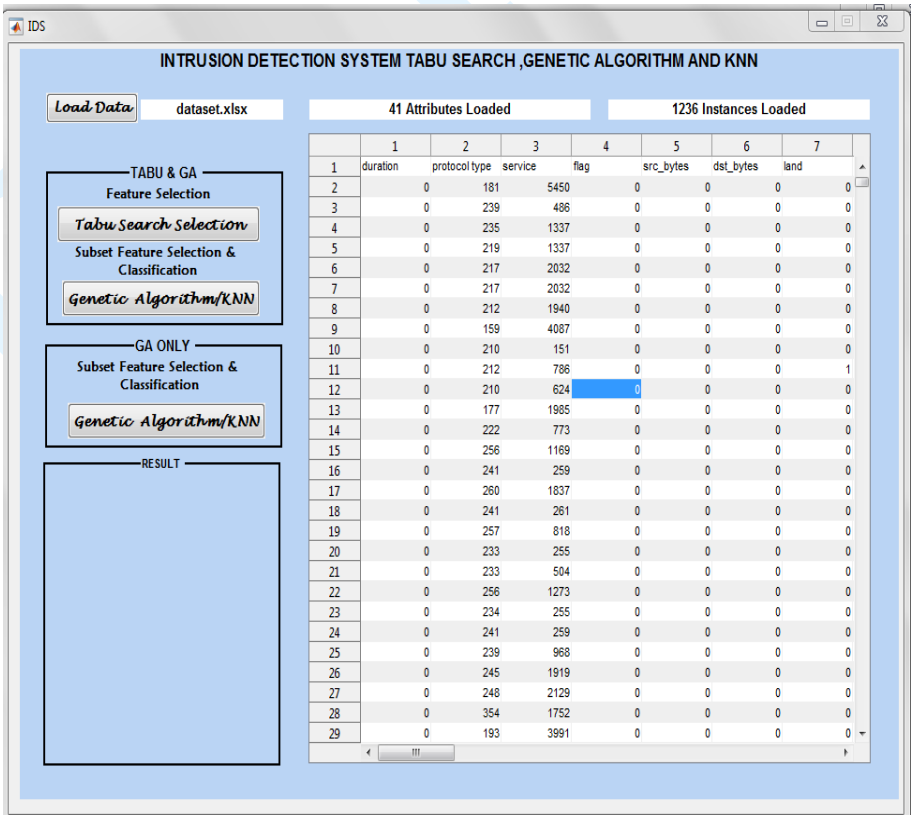


Figure 10. Genetic Algorithm and KNN classification

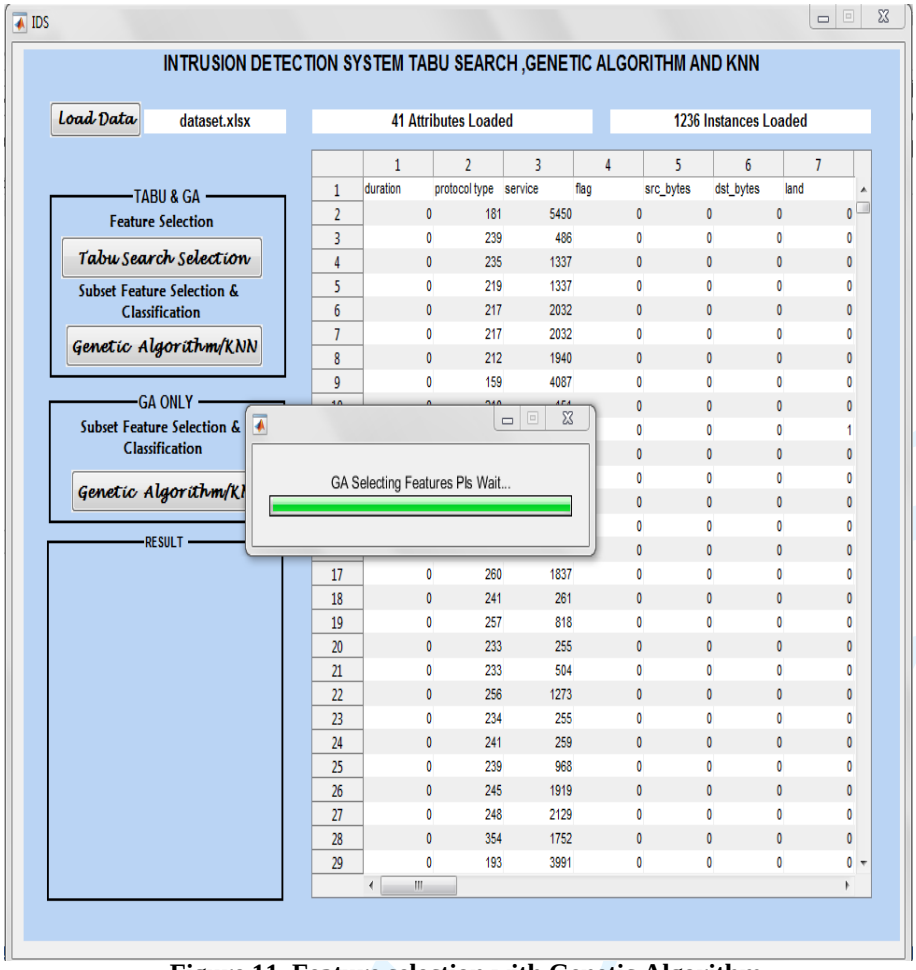


Figure 11. Feature selection with Genetic Algorithm

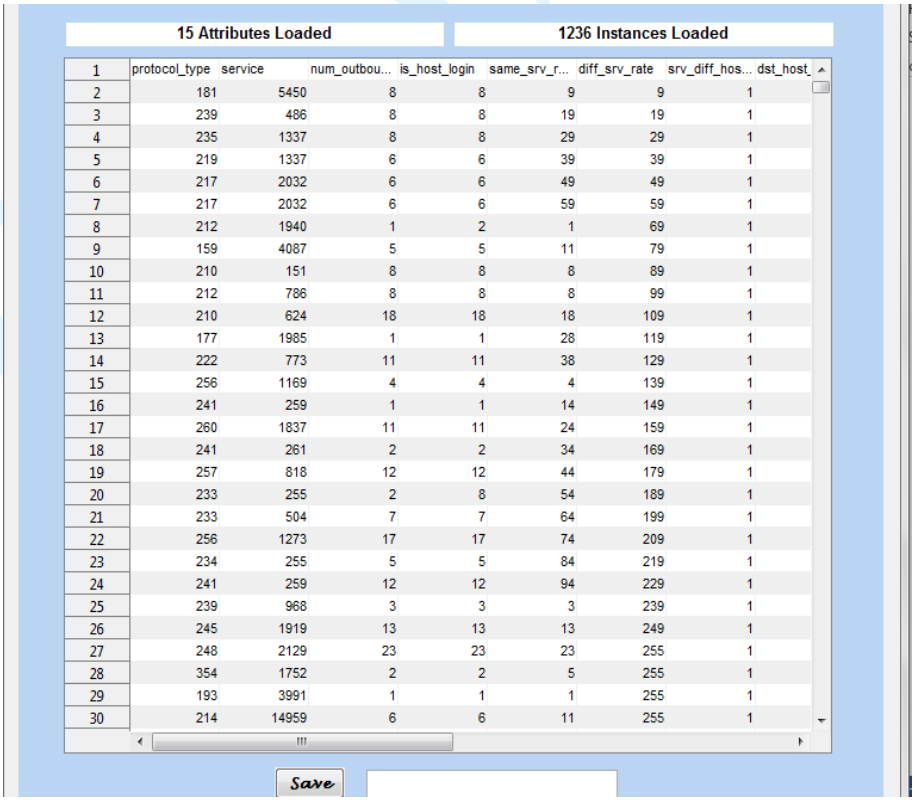


Figure 12. Selection of Features with Genetic Algorithm

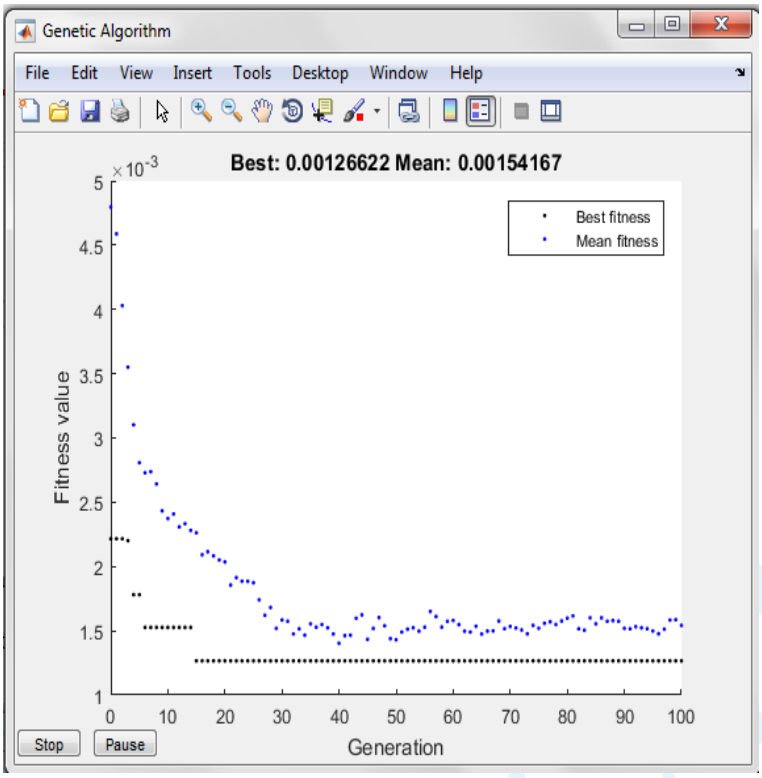


Figure 13. Generation of Fitness Value with GA

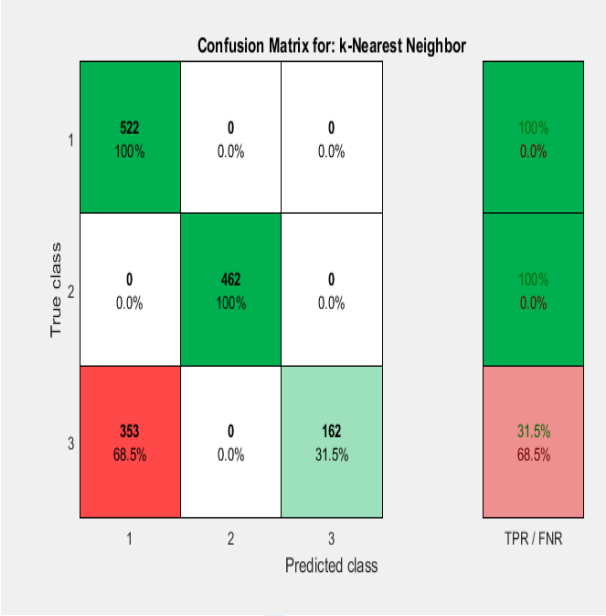


Figure 14. Confusion Matrix of GA with KNN

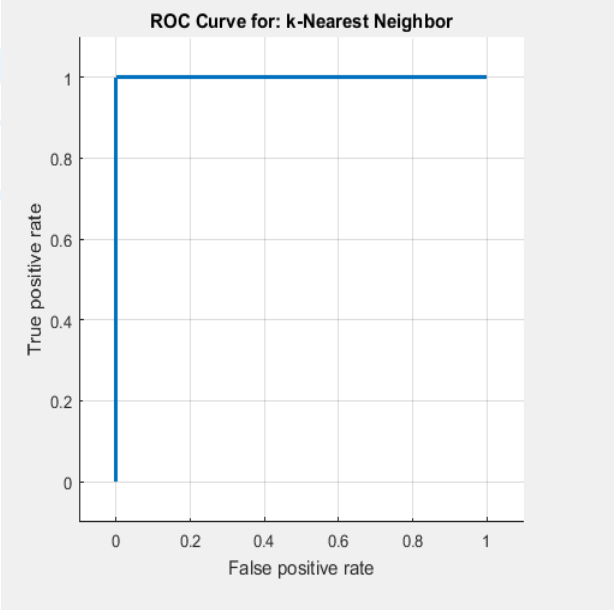


Figure 15. ROC Curve for GA and KNN

10. CONCLUSION

This paper presents a hybridized intrusion detection system that predicts network intrusion based on extracted data from standard intrusion detection dataset obtained from the KDD Cup 1999. The study compared the hybridization of Tabu search with GA and Genetic Algorithm only.

The proposed IDS system followed a data mining technique to select features so as to remove features with little or no predictive information from the dataset, this was carried out using combination of tabu search with GA and GA only. The first stage analysis presented a better result as tabu search was used to select features while the GA optimized the reduced dataset by tabu search to obtain a subset rather than GA only that was applied to select the discriminant features. The predictive accuracy was seen in hybridization of KNN with GA. The final

classification accuracy showed that combination of tabu search and Genetic algorithm gave a better result than the using Genetic Algorithm only.

Table 1. Comparative Analysis of Tabu, Genetic and Hybridized KNN

Algorithm	Time(s)	Selected Features	Classification (%)	Misclassification Accuracy (%)
Tabu Search & GA	2.316s	10	83.56	16.44
GA Only	2.544s	15	77.17	28.83

REFERENCES

[Bij16] **M. Bijone** - *A Survey on Secure Network : Intrusion Detection & Prevention Approaches*, Am. J. Inf. Syst., vol. 4, no. 3, pp. 69–88, 2016.

[Deb00] **H. Debar** - *An introduction to intrusion-detection systems*, Proc. Connect, pp. 1–18, 2000.

[Dha13] **K.Dhangar** - *A Proposed Intrusion Detection System*, vol. 65, no. 23, pp. 46–50, 2013.

[DL12] **B. S. Dhak, S. Lade** - *An Evolutionary Approach to Intrusion Detection System using Genetic Algorithm*, Int. J. Emerg. Technol. Adv. Eng., vol. 2, no. 12, pp. 632–637, 2012.

[DM15] **M. Durairaj, A. Manimaran** - *A Study on Securing Cloud Environment from DDoS Attack to Preserve Data Availability*, Int. J. Sci. Technol., vol. 3, no. 2, pp. 63–72, 2015.

[EN10] **T. Ehab, H. O. Nasereddin** - *Using genetic algorithm in network security*, IJRRAS, vol. 5, no. 2, pp. 148–154, 2010.

[GZA05] **R. H. Gong, M. Zulkernine, P. Abolmaesumi** - *A software implementation of a genetic algorithm based approach to network intrusion detection*, Sixth Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. First ACIS Int. Work. Self-Assembling Wirel. Netw., pp. 246–253, 2005.

[G+09] **P.García-Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández, E. Vázquez** - *Anomaly-based network intrusion detection: Techniques, systems and challenges*, Comput. Secur., vol. 28, no. 1–2, pp. 18–28, 2009.

[HMB12] **M. S. Hoque, A. Mukit, A. N. Bikas** - *An Implementation of Intrusion Detection System Using Genetic Algorithm*, International J. Netw. Security Its #Appl., vol. 4, no. 2, pp. 109–120, 2012.

[JG14] **M. L. Jadhav, P. C. M. Gaikwad** - *Implementation of Intrusion Detection System using GA*, Int. J. Innov. Res. Electr. Intrusmentation Control Eng., vol. 2, no. 7, pp. 1733–1736, 2014.

[KI10] **K. R. Karthikeyan, A. Indra** - *Intrusion Detection Tools and Techniques –A Survey*, Int. J. Comput. Theory Eng., vol. 2, no. 6, pp. 901–906, 2010.

[KV02] **R. A. Kemmerer, G. Vigna** - *Intrusion detection: a brief history and overview*, Computer (Long. Beach. Calif)., vol. 35, no. 4, pp. 27–30, 2002.

[KPJ12] **G.Kanagaraj, S. G. Ponnambalam, N. Jawahar** - *Trends in Intelligent Robotics, Automation, and Manufacturing*, Commun. Comput. Inf. Sci., vol. 330, no. February, pp. 491–501, 2012.

[K+13] **B. S. Kumar, T. C. Sekhara, P. Raju, M. Ratnakar, S. D. Baba, N. Sudhakar** - *Intrusion Detection System- Types and Prevention*, Int. J. Comput. Sci. Inf. Technol., vol. 4, no. 1, pp. 77–82, 2013.

[Mus06] **A. R. Mushi** - *Tabu Search Heuristic for University Course Timetabling Problem*, African J. Sci. Technol., vol. 7, no. 1, pp. 34–40, 2006.

[MIS12] **A. B. Mohamed, N. B. Idris, B. Shanmugum** - *A Brief Introduction to Intrusion Detection System*, Commun. Comput. Inf. Sci., vol. 330, pp. 491–501, 2012.

[O+12] **F. O. Ojugo, A. A, Eboka, A. O, Okonta, O. E, Yoro, R. E. AGhware** - *Genetic Algorithm Rule-Based Intrusion Detection System (GAIDS)*, J. Emerg. Trends Comput. Inf. Sci., vol. 3, no. 8, pp. 1182–1194, 2012.

[Paw13] **S. N. Pawar** - *Intrusion Detection in Computer Network Using Genetic Algorithm Approach: A Survey*, Int. J. Adv. Eng. Technol., vol. 6, no. 2, pp. 730–736, 2013.

[PYM14] **P. Purushottam, S. Yogesh, K. Manali** - *Using Genetic Algorithm : A Study*, Int. J. Emerg. Technol. Comput. Sci., vol. 3, no. 2, pp. 282–286, 2014.

[RS14] **R. Ragupathy, R. Sharma** - *Detecting Denial of Service by Attacks by Analysing Network Traffic in Wireless Networks*, Interntaional J. Grid Distrib. Comput., vol. 7, no. 3, pp. 103–112, 2014.

[Sri08] **S.Srivatsa** - *Detecting and preventing attacks using network intrusion detection systems*, Int. J. Comput. Sci. Secur., vol. 2, no. 1, p. 49, 2008.

[SM07] **K. Scarfone, P. Mell** - *Guide to Intrusion Detection and Prevention Systems (IDPS) Recommendations of the National Institute of Standards and Technology*, Nist Spec. Publ., vol. 800–94, p. 127, 2007.

[TU16] **R. Thanuja, A. Umamakeswari** - *Effective Intrusion Detection System Design Using Genetic Algorithm For MANETs*, APRN J. Eng. Alpplied Sci., vol. 11, no. 7, pp. 4696–4700, 2016.