iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 5 **Issue:** VI **Month of publication:** June 2017

**DOI:**

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

# Reduction of Rejection Instances Using Particle Swarm Optimization in Ids

Shilpi Goyal[1], Sushil Chaturvedi[2]
[1,2] *Computer Science, SRCEM*

***Abstract*: An intrusion detection system (IDS) is a tool or software function which examines a network or systems for malevolent action or policy infringements. An IDS mechanism by observing system action by probing vulnerabilities in the system, the sum of files and conducting an study of patterns based on already known attacks. It also directly monitors the Internet to explore for most recent threats which could result in a future attack. This paper introduces optimization techniques with advantages and algorithm. We proposed new technique to decrease the rejected instances of the dataset. This performed simulation in MATLAB on the given dataset then compute the various parameters like rejected instances, true positive rate(TPR), false positive rate(FPR) and false negative rate(FNR). The graph confirmed that the proposed work is healthier than the existing work by reducing rejection rate.***

**Keywords***: IDS(Intrusion Detection System),GA (Genetic Algorithm),ACO(Ant Colony Optimization),PSO(Particle Swarm Optimization), Clustering Algorithm.*

## I. INTRODUCTION

Intrusion detection system is system software which examine attack on a network (n/w) or PC. IDS are usually categorized as Anomaly detection and Misuse detection. In Misuse system the signature of known attacks are stored in database. Any data just like that data is classed as attacks. Anomaly detection refers to statistical expertise approximately ordinary activity. Semi-supervised anomaly detection method requires a collection of simply normal training information from that they found the profile of typical behavior. If the training info contains few attacks hidden inside it, the approach may not recognize future instances of these attacks. On the alternative hand, unsupervised anomaly detection approach set up the profile of ordinary conduct with unlabeled education data that contain of both ordinary as properly as anomalous samples. Intrusions correspond to deviations from the normal activity of system. The anomaly detection system has high false positive/ negative alarm rate compared to misuse detection systems [1].

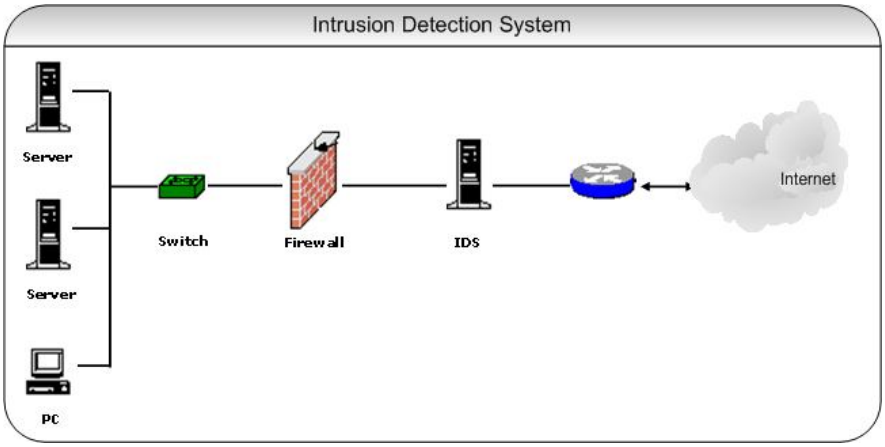



Fig .1 IDS

*A. Many drawbacks have in conventional Approach:*

*1)* Several signature-based IDSs have barely defined signatures which prevent them from detecting variant of general attacks.
*2)* Anomaly detection method typically makes a big no. of wrong alarms because of the random nature of consumer and n/w.
*3)* Anomaly detection method mostly necessitates broad “training sets” of system occurrence facts in sequence to characterize normal behavior patterns.
*4)* Application-based IDSs might be weaker than host depend IDSs to being attacks and incapable since they run like an usage on the host they are monitoring



©IJRASET: All Rights are Reserved

## II. LITERATURE SURVEY

Xingang Wang et al. [2016] This paper introduce allusion to the influence of the starting value of the K-Means algorithm on the optimal solution of the algorithm, a hybrid algorithm of K-Means based on SA-PSO is proposed. The new algorithm uses the gain of bouncing out of local minima to progress the execution of the PSO algorithm, with the global optimization, the new algorithm can conquer the lack of the K-Means algorithm which is simple to drop into the local best solution. The practical outcome shows that the K-Means calculation in light of the hybrid SA-PSO algorithm compared with the PSO algorithm in view of the K-Means calculation has been partially effective increase in Global convergence[2].

Gözde Karataş et.al. [2016] This paper introduce Information for the different detection algorithms using genetic algorithm, which are made of IDS algorithms is given and literature search's been made[3].

Tulasichandra Sekhar Gorripotu et. Al.[2015] This paper introduce The simulation comes about uncover that Proportional Integral Derivative (PID) controller confers better dynamic response compared to others. Finally, the power of PID controller is explored by incorporating Time Delay (TD) in the system and for the variation of Step Load Perturbation (SLP). It is monitor from the simulation outcomes that the proposed PID controller can withstand changes made in the method for the similar optimal values[4].

M. Hossein ahmandazdean et.al. [2015] This paper introduce An intrusion detection model was formed by GA and the construction of the classifying models in view of the training information observed from the GA by KNN (K-nearest neighbors) algorithm has been taken into consideration. In this algorithm, its support to reduced the FPR in comparison with previous method[5].

Tahir Mehmood et.al. [2015] This paper introduce Detection model, ant system with SVM, which makes use of ant system, a translation of Ant Colony Optimization(ACO), to filter the redundant and irrelevant functions for SVM category algorithm. KDD99, that is a benchmark dataset utilized for anomaly detection, has been taken after here. Every instance in KDD99 have been defined by 41 features which additionally has few redundant or irrelevant features. Ant system has been utilized to take away the ones redundant and nearby the point functions. The decided on characteristic subset using ant device is then confirmed the use of SVM. The practical outcome about confirmed that the achievement of the arrangement calculation, When trained with the reduced feature set, has been improved. The performance measures used in this contrast are real positive rate, FPR, and precision[6].

Xu Yang et.al. [2015] This paper introduce data mining technology is considered in this article that allows you to diminish the measure of the false alarms created with the guide of IDSs and in the interim enhance the detection accuracy, wherein such data mining is an unsupervised clustering strategy based on hybrid ant colony algorithm and can be. Used to determine prior knowledge. practices, without the require to diagnose the earlier learning. In the interim, we receive K- means clustering algorithm to help up the meeting rate of the Ant Colony calculation. Really, the trial result demonstrates that the technique proposed along these lines has higher detection rate yet bring lower false alarm rate[7].

Sadegh hesari et. al.[2014] This paper introduce PSO algorithm of rules has been utilized to imp and decrease the losses with the aid of the use of receiving electroma output speed in a web way. The contrasted for the two methods of and with The outcomes affirm the proposed strategy[8].

Naila Belhadj Aissa, Mohamed Guerroumi et.al.[2015] This paper introduce we offered a clustering-based detection technique applying a GA named Genetic Clustering for Anomaly-based Detection (GC-AD). GC-AD usesa dissimilarity measure to form k clusters. It, then, applies a genetic process where each chromosome represents the centroids of the k clusters. A two-stage fitness function is proposed. i) We establish a confidence interval to refine the clusters to obtain partitions that are more uniform. ii) We compute and maximize the inter-cluster difference above the generations. The correctness of our process is tested on various subset from KDD99 dataset. The results are discussed and compared to kmeans clustering algorithm[9].

## III.CLUSTERING ANALYSIS

### *A. K Means Clustering*

Fuzzy clustering allows each feature vector belongs to many cluster with different membership degrees (between 0 and 1), and imprecise or fuzzy boundaries among clusters. Fuzzy clustering is often used in modeling (fuzzy modeling, neural networks, rule-based systems), where the clusters are sought as a structural setup for the ensuing modelling activities.

### *B. Genetic K-Means Algorithm*

K. Krishna and M. Narasimha Murty proposed a novel hybrid genetic algorithm (GA) that search a global optimal partition of a known information into a predetermined amount of clusters. GAs used before in clustering, utilize either an costly crossover operator to create suitable child chromosomes from parent chromosomes, or a high fitness function or both. To go around these costly operations, they hybridized the GA with a classical gradient descent algorithm applied in clustering, viz., the K-means



©IJRASET: All Rights are Reserved

algorithm. Therefore, the name genetic K-means algorithm (GKA). They defined the K-means operator, one-step of the K-means algorithm, and utilized it in GKA as a search operator, rather than crossover. They also defined a biased mutation operator particular to clustering, called distance-based-mutation. Applying the finite Markov chain theory, they confirmed that the GKA merges to a global optimum. It is monitor in the recreations in which GKA meets to the best known optimum, comparing to the given information, in simultaneousness with the merging attacks. It is additionally observed that the GKA search faster than another evolutionary algorithms used for clustering. The benefits of the genetic k means clustering algorithm is that it is quicker than a portion of the another clustering algorithms [10].

## IV. OPTIMIZATION TECHNIQUES

### *A. Genetic algorithm*

GA is a circle of relatives of computational models based on ideas of evolution and herbal choice. These algorithms change the complexity in a specified area into a replica thru exploiting a chromosome-as information structure and evolve the chromosomes exploiting race, recombination, and change operators. The scope of the use which can create exploit of GA is quite broad. In PC security applications, it is mainly used for finding ideal answers for a specific problem. The methods of a GA commonly start with a randomly selected populace of chromosomes [11].

*1) Disadvantages*

*a)* GA need more computational time.

*b)* Certain optimization problem (they defined as variant problems) can't be solve with genetic algorithm. There is not full declaration that genetic algorithm will locate a global optimum.

*c)* Like other AI techniques, the genetic algorithm cannot guarantee consistent optimization reaction times.

### *B. Ant Colony optimization*

The ACO is a probabilistic technique for settling computational issues which can be decreased to finding legitimately ways by way of charts primarily based at the techniques of real ants. It was to start with proposed in 1992 through Colorni, Dorigo and Maniezzo. In ACO, every artificial ant is taken into opinion as a simple agent, speaking with other ants most effective not directly and by way of effecting modifications to a not unusual environment [12].

*1) Disadvantages*

*a)* Slower convergence than other Heuristics

*b)* Performed poorly for TSP problems larger than 75 cities.

*c)* No centralized processor to guide the AS towards good solutions.

### *C. Particle swarm optimization*

PSO turned out to be initially included by utilizing Dr. Russell C. Eberhart and Dr. James Kennedy in 1995. Particle Swarm has two essential administrators: Velocity redesign and Position overhaul. Amid every era every particle is extended toward the particles preceding first-rate function and the global great position. At every new release a new velocity value for several particle is calculated based totally on its current velocity, the distance from its preceding excellent role, and the distance from the global best position. The new velocity value is then used to calculate the following role of the particle in the seek area. This procedure is then iterated aset number of times or until a minimum error is achieved. The PSO algorithm of rules proceeds as follows [13]:

*1)* procedure PSO
*2)* repeat
*3)* for i = 1 to range of individuals do
*4)* if G(~xi) > G(~pi) then . G() evaluates goodness
*5)* for d = 1 to dimensions do
*6)* pid = xid . Pid is the excellent state determined so far
*7)* end for
*8)* end if
*9)* g = i . Arbitrary



©IJRASET: All Rights are Reserved

*10)* for j = indexes of neighbors do
*11)* if G(~pj) > G(~pg) then
*12)* g = j. g is the index of the exceptional performer inside the Neighborhood
*13)* end if
*14)* end for
*15)* for d = 1 to quantity of dimensions do
*16)* vid(t) = f(xid(t − 1), vid(t − 1), pid, pgd) . update velocity
*17)* vid 2 (−Vmax,+Vmax)
*18)* xid(t) = f(vid(t), xid(t − 1)) . Update role
*19)* end for
*20)* end for
*21)* until stopping criteria
*22)* end procedure
*23)* *Advantages*

*a)* Simple in implementation.
*b)* Easily parallelized for concurrent processing.
*c)* Derivative free.
*d)* Less number of algorithm parameters.
*e)* Efficient global search algorithm

## V. PROPOSED WORK

Intrusion Detection Systems (IDS) have been broadly arranged and many procedures to identify, detect and categorize attack have been planned, developed and tested either offline or online. An intrusion is described as “any group of activities that attempt to settlement the integrity, confidentiality or accessibility of an asset”. An intrusion detection system is a combination of components and associated methods that aim to observe network or computer-host activity to identify and react to any attempted attack or intrusion. Cluster analysis is the specialty of resulting groups in data. The purpose is to division a combination of unlabeled samples into homogeneous groups of same features based on some similarity process defined for the underlying samples space.

In the existing work, they performed anomaly based detection by using clustering and genetic for optimal partitioning of data into clusters. They performed clustering then genetic algorithm for finding the fitness function of the population. Selection, crossover and mutation performed on the clusters to get the better result. But with the amplify in the sum of dataset, the lot of rejected instances will also increase which is not suitable for large dataset.

To overcome the above problem, we performed the procedure to decrease the rejected instances to improve the process. Initially we defined the parameters on which we have to perform the operations. Then initialize the population and partition the dataset by using clustering and find the fitness function of each individual. Now optimal position and velocity of the particles can be computed by applying PSO Algorithm. Update the particle value until we get the Pbest and Gbest for the overall solution.

Proposed algorithm

| | |
|---|---|
| Step:1 | Set parameters like no. of max iterations, size of population |
| Step:2 | Normalized the dataset |
| Step:3 | Generate initial population |
| Step:4 | Put most useful attribute in variable called datam |
| Step:5 | Perform partition by using K-means clustering |
| Step:6 | Evaluate fitness function of each partition |
| Step:7 | Apply PSO for cost function |
| Step:8 | Initialize particle position, particle velocity |
| Step:9 | For iteration 1 to max_iterations |
| Step:10 | Evaluate cost function |
| Step:11 | If (cost function is best) |
| | Update Pbest |
| Step:12 | If (Pbest is best) |



©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653


Update Gbest

Step:13 Update particle velocity and position

Step:14 Output the optimal result as an output

Step:15 Stop

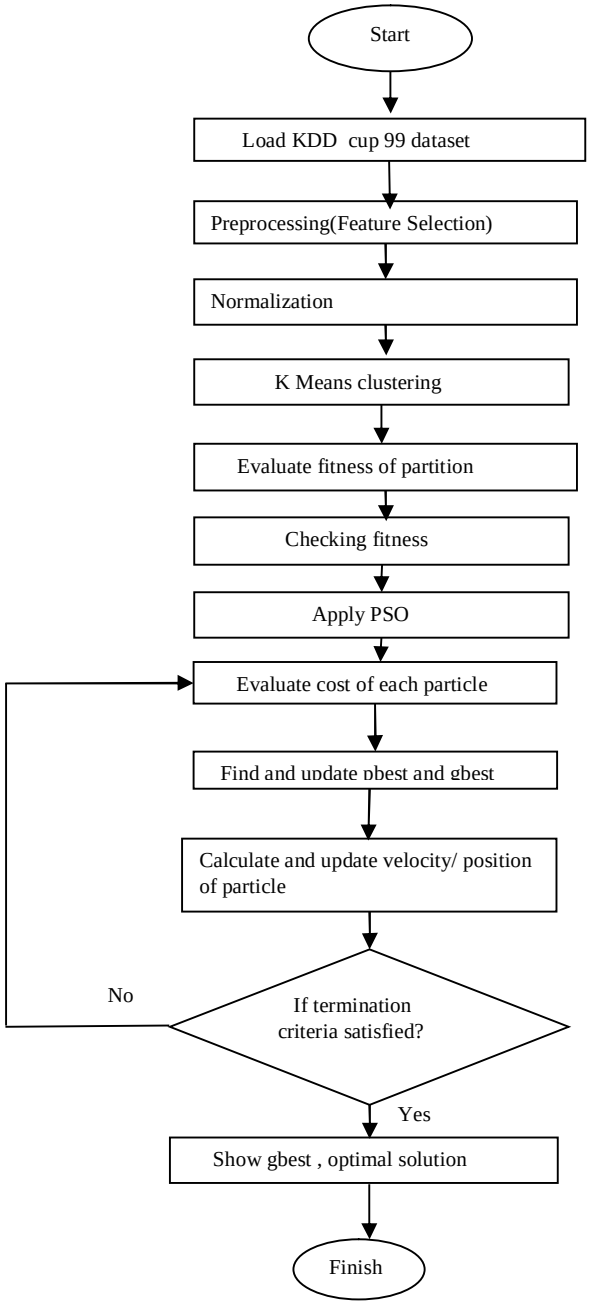



Fig. 2 Flowchart of proposed technique



©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

## VI. RESULTS ANALYSIS

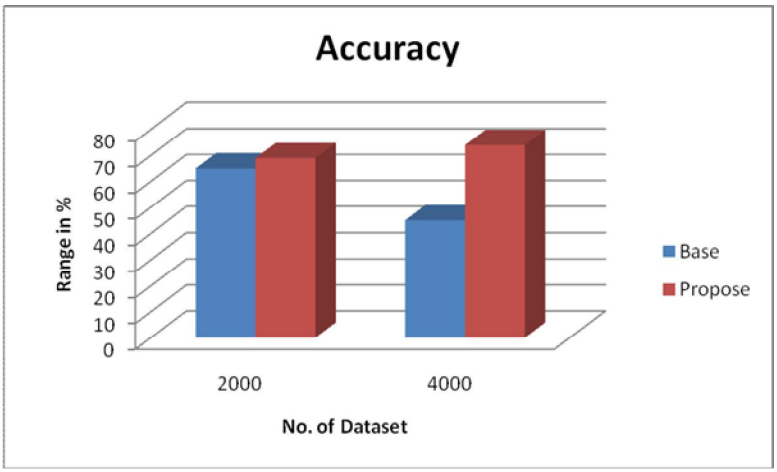



Fig. 3 comparison of Base and Propose accuracy

Fig. 4 comparison of Base and Propose rejected instances

Fig. 5 comparison of Base and Propose true positive rate


©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

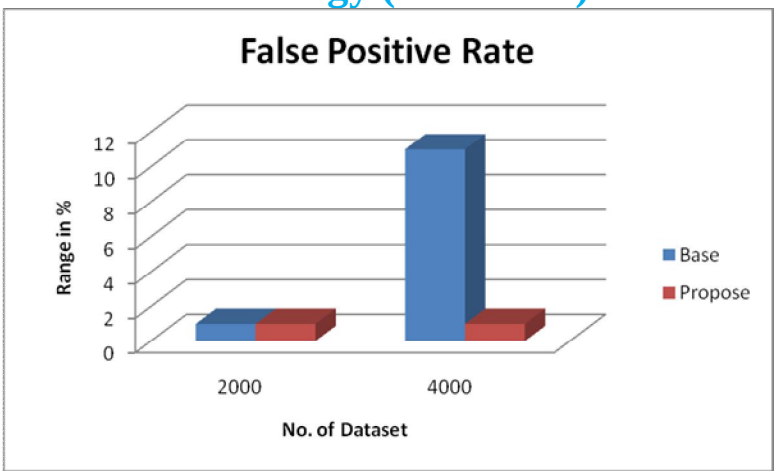



Fig. 6 comparison of Base and Propose false positive rate

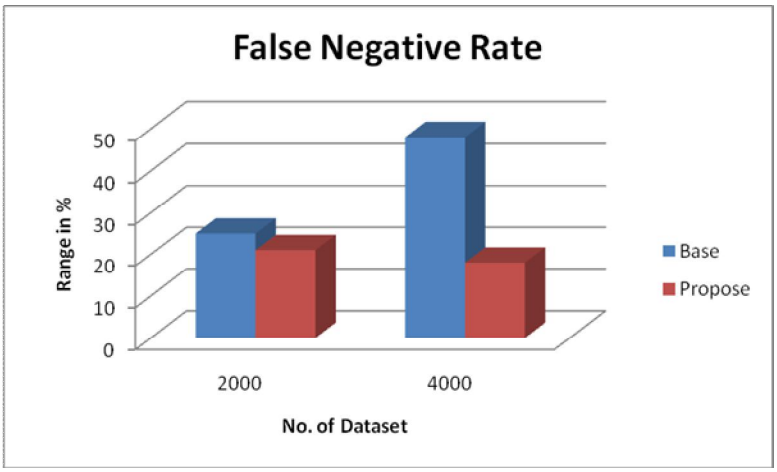



Fig. 7 comparison of Base and Propose false negative rate

## VII. CONCLUSION

An IDS is the system of detecting the unwanted activity by examining the traffic in the network. Clustering is the creation of group to perform some operations on the homogeneous data. Genetic algorithm is the method to find the optimal outcome from the dataset. PSO might sound complex, but it's really a very straightforward algorithm. In PSO, the position and velocity of population are calculated on the basis of cost function. We performed clustering then PSO algorithm to decrease the rejected instances of the dataset on the various records of data like 2000 and 4000.In future, we can work on feature selection algorithm.

## REFERENCES

[1] Harshit Saxena, Vineet Richaariya, Ph.D, "Intrusion Detection in KDD99 Dataset using SVM-PSO and Feature Reduction with Information Gain", International Journal of Computer Applications (0975 – 8887) Volume 98– No.6, July 2014
[2] Xingang Wang, Qi Sun, "The Study of K-Means Based on Hybrid SA-PSO Algorithm", 2473-3547/16 $31.00 © 2016 IEEE.
[3] Gözde Karataş, "Genetik Algoritma ile Saldırı Tespit Sistemi Genetic Algorithm For Intrusion Detection System", 978-1-5090-1679-2/16/$31.00 ©2016 IEEE.
[4] Tulasichandra Sekhar Gorripotu, Rabindra Kumar Sahu, Sidhartha Panda, "Comparative performance analysis of classicalcontrollers in LFC using FA technique", 978-1-4799-7678-2/15/$31.00 ©2015 IEEE.
[5] M. Hossein ahmandazdean,Ali asgar Khorshidvand, mahdi Ghalbi valiant, "Low –rate false alarm intrustion detection system with Genetic algorithm approach",978-1-4673-6506-2/15/2015 IEEE.
[6] Tahir Mehmood, Helmi B MD Rais, "SVM For Network Anomaly Detection Using ACO Feature Subset", 978-1-4799-7896-0/15/$31.00 ©2015 IEEE.
[7] Xu Yang, Zhao Hui, "Intrusion Detection Alarm Filtering Technology Based on Ant Colony Clustering Algorithm", 978-1-4673-9393-5/15 $31.00 © 2015 IEEE.
[8] Sadegh hesari, mohammad bagher naghibi sistani, "efficiency improvement by timely controlling power factor in permanent magnet synchronous motor using PSO algorithm", 978-1-4799-8021-5/14/2014 IEEE.
[9] Naila Belhadj Aissa, Mohamed Guerroumi "A Genetic Clustering Technique for Anomaly-BasedIntrusion Detection Systems"2015 IEEE

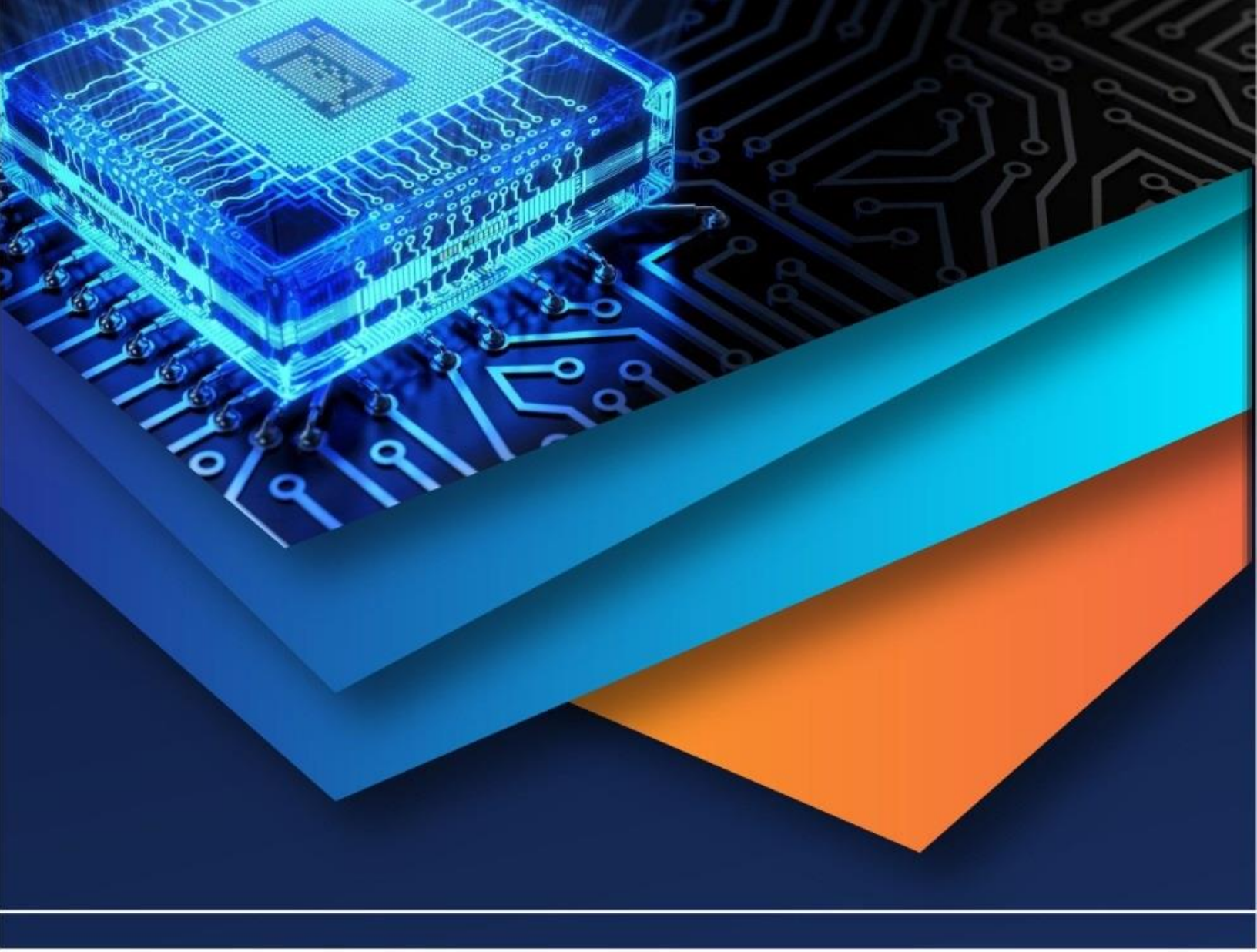


©IJRASET: All Rights are Reserved

Volume 5 Issue VI, June 2017
ISSN: 2321-9653


[10] P. IndiraPriya, Dr. D.K.Ghosh "A Survey on Different Clustering Algorithms in Data Mining Technique" International Journal of Modern Engineering Research (IJMER) Vol.3, Issue.1, Jan-Feb. 2013 pp-267-274 ISSN: 2249-6645.
[11] Xingang Wang, Qi Sun, "The Study of K-Means Based on Hybrid SA-PSO Algorithm", 2473-3547/16 $31.00 © 2016 IEEE.
[12] Gözde Karataş, "Genetik Algoritma ile Saldırı Tespit Sistemi Genetic Algorithm For Intrusion Detection System", 978-1-5090-1679-2/16/$31.00 ©2016 IEEE.
[13] M. Hossein ahmandazdean,Ali asgar Khorshidvand, mahdi Ghalbi valiant, "Low –rate false alarm intrustion detection system with Genetic algorithm approach",978-1-4673-6506-2/15/2015 IEEE


©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)