IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Issue-2, Feb 2019 E-ISSN: 2347-2693

Position Depended Sybil Attack Detection using Efficient KNN technique with Clustering

Rajeev Bedi[1], Baljinder Singh[2], Meenakshi Devi[3*]

[1,2]Dept. of Computer Science Beant College of engineering and Technology, India
[3]Mtech Scholar Beant College of engineering and Technology, India

**Corresponding Authors: mguleria6@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i2.266272 | Available online at: www.ijcseonline.org

Accepted: 10/Feb/2019, Published: 28/Feb/2019

Abstract- In today's world the wireless sensor network has great significant in application like defense surveillance, patient health monitoring, traffic control etc. As WSN utilize radio frequencies so there is threat of interference in network. These threats also include distributed denial of service in which the messages that are sent over the network may be attacked by unauthorized user. It would harm the confidentiality of the network user and the services of network. There are various algorithm that are utilized to detect Sybil attack in WSN but these schemes only stress on prevention of attack after it is occurred. This would leads to the loss of data and more consumption of limited network resources. So in this work we introduce a new algorithm that is based on clustering based KNN along with Euclidean distance. It would detect earlier the Sybil attack in WSN and prevent the data loss. The parameters like throughput, energy consumption etc are utilized to analyze the performance of this technique.

Keywords: *KNN, WSN, Sybil detection*

I. INTRODUCTION

Wireless sensor networking stays a standout amongst the most requesting and rising exploration territories of our chance. A Wireless Sensor Network (WSN) is a gathering of self-ruling nodes, which transmits information in wireless channel with little transmission capacity utilization and recurrence. [1]

The various applications such as military application, data collection and monitoring utilize the sensor network because it gives minimal effort solution. Every hub can discover their neighbor nodes in network and this give assistance in courses arrangement in the gathering.[2] These kinds of assaults lessen the ability of WSN, with the goal that they can't work for a drawn out stretch of time. It has often consequences for utilization assets in the network and expands the energy utilization, delay, and decreases the throughput. [3]

The un-ability of authorized user to access network resources that may be website or whole system is known as Sybil attack. A Distributed Sybil attack is a synchronized assault which is done on the accessibility of services of some specific network with the assistance of traded off processing frameworks in a roundabout way, so tracking the cloned packets turns out to be more troublesome [3] The principle point of this paper is to shield the Wireless Sensor Network from flooding, a kind of Sybil assault. Flooding can deplete all network assets, for example, data transfer capacity, energy and processing power and so on and plan another location plot named early identification of Sybil assault utilizing distributed method. This plan recognizes the attacker based on the quantity of transmissions relating to the quantity of neighbors of a hub and these transmissions are contrasted and the limit esteem registered and PDR of different nodes in the network. [4]

Sybil attack (additionally called hub replication attack) is a serious attack in WSNs. In this attack, a foe catches just a couple of hubs, duplicates them and afterward conveys subjective number of imitations all through the system. The catch of hubs is conceivable in light of the fact that sensor hubs are typically unprotected by physical protecting because of cost contemplations, and are frequently left unattended after deployment. On the off chance that we don't distinguish these reproductions, the system will be helpless against a vast class of inner attacks.[5] For instance, the foe presently can catch the movement passing the reproductions (which may contain the previously mentioned areas of troopers), infuse false information into the system (which might be false summons), slander different hubs and even disavow true blue hubs. Hitherto, most

conventions for identifying hub replication have depended on a put stock in base station to give worldwide location. Additionally a portion of the current verification strategies [4, 5] can't identify such attacks, since every one of the reproductions hold real keys. The current methodologies fall into following two classes:

A. Brought together Detection The clearest recognition conspires requires every hub to send a rundown of its neighbors and their guaranteed areas to the base station. The base station would then be able to analyze each neighbor rundown to search for imitated hubs. On the off chance that it finds at least one copy, it can repudiate the imitated hubs by flooding the system with a confirmed renouncement message. [6]

B. Nearby Detection: To abstain from depending on a focal base station, we could rather depend on a hub's neighbors to perform replication identification. Utilizing a voting system, the neighbors can achieve an agreement on the authenticity of a given hub. Sadly, while accomplishing recognition in a disseminated design, this technique neglects to distinguish circulated hub replication in disjoint neighborhoods inside the system. For whatever length of time that the duplicated hubs are no less than two bounces from each other, a simply neighborhood approach will fail. [7]

A clear answer for protect against Sybil attacks is to give the base station a chance to gather the area data (e.g. area, neighbor list, and so forth.) from every sensor and screen the system centralized. This approach experiences high correspondence overhead by asking for excess data from the system. Further, a "shrewd" Sybil may report the area of the first hub, influencing the base station to flop in distinguishing the imitation. In [8], propose for one-jump networks that the base station (BS) can store the one of a kind flag trademark for every gadget, and in this way gadget cloning can be distinguished as needs be. Nonetheless, in a multi-bounce sensor organize; it is unreasonable for BS to track the flag attributes of sensors multi-jumps away. In restricted voting/trouble making identification plans [8], hubs inside an area concur/vote on the authenticity of a given hub in view of their nearby perceptions. By the by, these plans are not fit for identifying clones with typical conduct, and may fizzle when various clones in closeness intrigue. Moreover, limited voting/trouble making identification plots intrinsically do not have the capacity to identify dispersed clones that may show up at wherever in the system.

II. LITERATURE REVIEW

Sybil attack detection methodology is proposed by [9]. The framework employed by Kontaxis et. Al can be used by the users to determine whether they are under Sybil attack or not. The components employed in this framework involves

a. Information distiller
This component is used in order to extract the information from legitimate social networking site. Information that could be used to identify the user is extracted by this component and maintained within the buffer.

b. Profile Hunter
Profile hunter used to locate the profile of the users. In case multiple records corresponding to single user is fetched then Sybil attack is detected.

c. Profile verifier
This component verifies the records filtered by profile hunter. The filtered information is compared against the profile of the user to find the nearest matches. In case matches do occur, profile Sybil attack is detected. User footprint analysis is proposed by [10]. User may have multiple accounts over the various services over the internet. All the services over the internet uses digital mechanisms.

Topological feature extraction mechanism is proposed by[11] for Sybil attack detection. In Sybil attack detection, earliest techniques assume that distinguished keywords are used by malicious users. But this may not be the case all the time. in order to tackle the situations, features like images, topological features etc. must be analysed. Topological analysis allow the user to construct the profile on the basis of heterogeneous features hence producing accurate result associated with the Sybil attack.
The Sybil attack detection techniques as proposed by [12] can be considered for such attack resolution. According to Dave et. Al., attack can either be on the access restricted information and anonymous data attacks. To tackle the situations attributes similarity based privacy preservation solutions are proposed. Several techniques corresponding to attribute similarity are used in order to determine the Sybil attacks.

Social networking is one of the most widely used internet activity as proposed by [9]. it is prone to profile Sybil attacks and its preservation is compulsory. Kontaxis et al proposed mechasnism for detection of profile Sybil attacks by the use of

archituctural design and prototype system for detecting similarity of attributes in case profile of the user is copied. Experiment result shows better result of Sybil attack detection hence proving worth of the study.

Sybil attack is a problem over the online social media. Detecting and preserving the state of the online social media is a need of the hour. Online social media plays a role of complex network. To detect the profile cloning attacks from such a network technique has been proposed by [13]**.** Entire social media is diided into tow parts. First part considered and dra the social network as a graph. In the second part, graph is divided into subparts based on the similarity of profile. The modular approach considered ultimatley led to the formation of smaller networks consisting of only those nodes having similar characterstics or properties thus facilitate detection of Sybilattacks. Online social media is a huge network of users. As the uers of the online social media grows, so does the chances of Sybil attack. To detect the Sybilattack a new approach for Sybilattack detection is proposed by [14]. Sybilattacks causes the similar profiles from one or more users. In order to determine the similarity, strength of suers profiles matching is determined. The strength determines profile Sybil attack by the said mechanism. degree of modularity achieved through this technqiue is not perfect and required certain degree of modifications.

III. PROPOSED SYSTEM

Sybil Algorithm is a range based algorithm. In range based algorithm only use range measurement whereas range free algorithm consider content of the message. Sybil Attack algorithm is created for detecting and removing wormhole attack. In our algorithm we have included NCA also. NCA means node capture attack. In Node capture attack, a node is captured and then falsifying information is given about the node. The node capture attack will make the attacker grab the information about the particular node and replace the existing node with the malicious node. The malicious node then act in place of the other node. The malicious activity performed by the node will make the actual node to be accounted for and punished. In order to resolve the problem random key will be utilized. The KNN with random key hence is proposed.

Algorithm

The existing algorithm will be as follows

a) Nodes are assigned the unique Ids.
b) Apply the Euc-dist to determine the position of nodes.
c) Compare the distance with the threshold
d) Add the node into the cluster
e) If More nodes with same id in cluster exist then
f) Node must be added in black list
   Else
g) Move onto next step in sequence
   End of if
h) Calculate falsifying information by the use of location error
i) Stop

The security is the big issue in the existing system. The security concern causes the use of random keys in the proposed approach. the random keys along with permanent blockage mechanism is employed in the proposed system. The Sybil attack with multiple identity nodes are blocked by the use of proposed system.

a) Nodes are assigned the Random Ids.
   b) Apply the Euc-dist to determine the position of nodes.
   c) Compare the distance with the threshold
   d) Add the node into the cluster
   e) If More nodes with same id in cluster exist then
   f) Node must be added in black list
      Else
   g) Move onto next step in sequence
      End of if
   h) Calculate falsifying information by the use of location error
   i) Stop

The main difference between the existing and proposed mechanism is that the proposed mechanism not only identify the attack but also block the attack. In addition localization error in case of existing approach is high but in case of proposed system it is low. Euclidean distance mechanism is employed in the proposed approach with the random keys. The random keys generated and assigned to the nodes help in improving the localization process.

Number of malicious nodes identified in the existing system are less and number of malicious nodes identified in the proposed system are more.

IV. RESULT AND PERFORMANCE ANALYSYS

Attacks on the sensors within wireless sensor network are common and must be prevented for increasing lifetime of the network. Sensor energy depleted quickly in case of attacks. Most common type of attack is known as Sybil attack. In case of Sybil attack, nodes with same identity exists with in the cluster and sender cannot identify the actual destination. Thus critical information is leaked and transmitted to the malicious node.

The proposed mechanism tries to resolve the problem of Sybil node by identifying malicious nodes and also blocking the nodes. The execution time is considerably reduced by the use of proposed mechanism. In addition, number of nodes detected as malicious are more in proposed system as compared to existing system.

Table 1: Showing execution time of existing without euc and proposed system with euc in milli seconds

Clustering using Hybrid_KNN(KNN_EUC)	Clustering without EUC
12.4357	21.4715
36.3243	43.4277
48.2805	62.4345
46.1414	61.4107
73.1829	97.9666
101.005	103.473

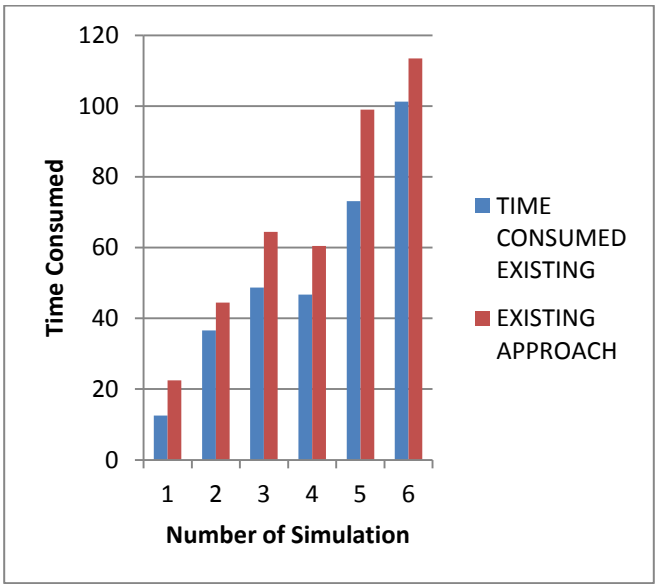


Figure 1: Showing execution time in milliseconds with and without Euclidean distance

The simulation takes place in MATLAB 2017 and number of nodes are varies. The probability of malicious nodes varied greatly as number of nodes varies in the existing and proposed system. In the existing approach manhatton distance approach is followed but in the proposed approach random key with Euclidean distance mechanism is employed and probability of indirect Sybil node is determined in advance. The crossover probability(CP) of nodes indicates the nodes crossing each other. Higher the probability more chance of Sybil nodes as predicted in the following result.

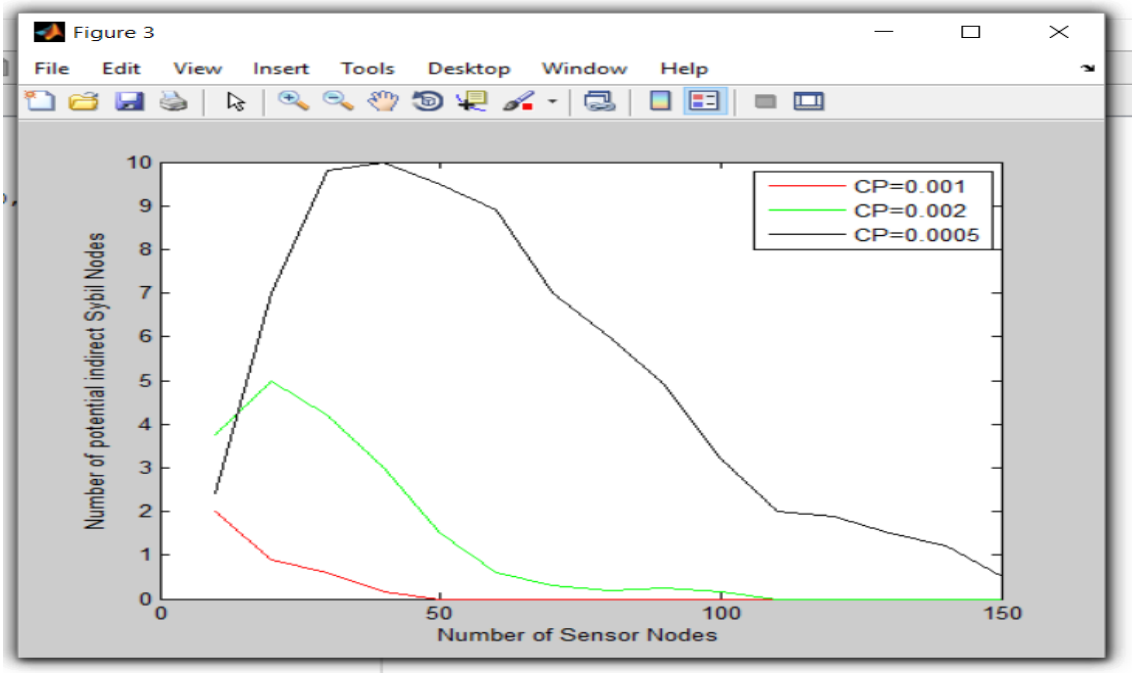


Figure 2: Nodes that could be termed as Sybil nodes

'R' indicates the random nodes without intersection presence. This means nodes are not intersecting. 'E' is the localization error. As the ratio of R/E increases the Sybil node probability also decreases. This means neighbor detected are less in this case. This is given through 0 neighbor.

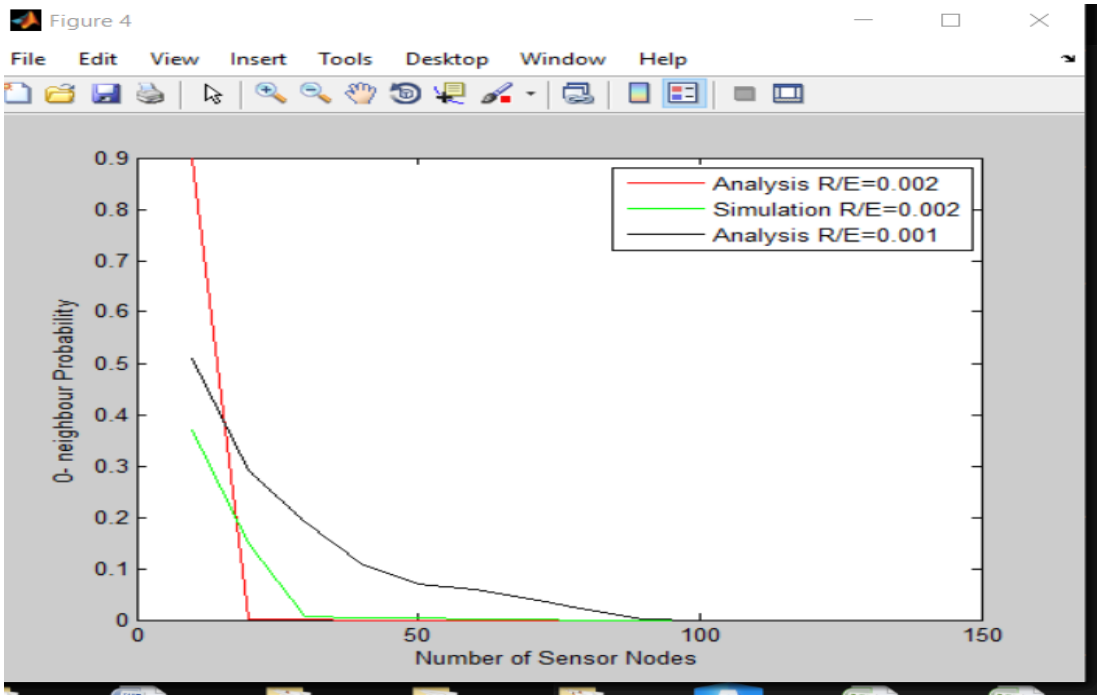


Figure 3: Probability of least Sybil nodes.

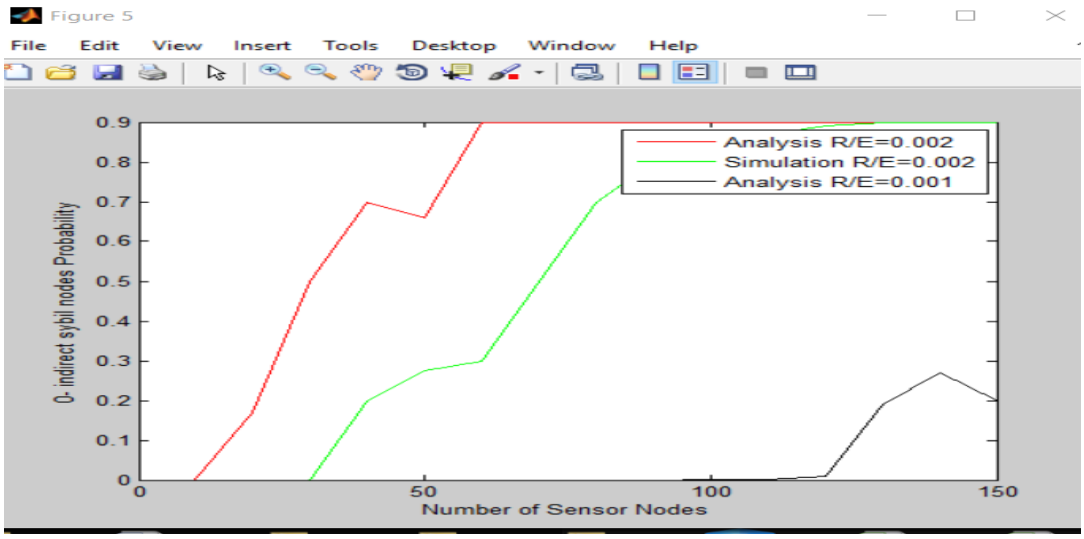
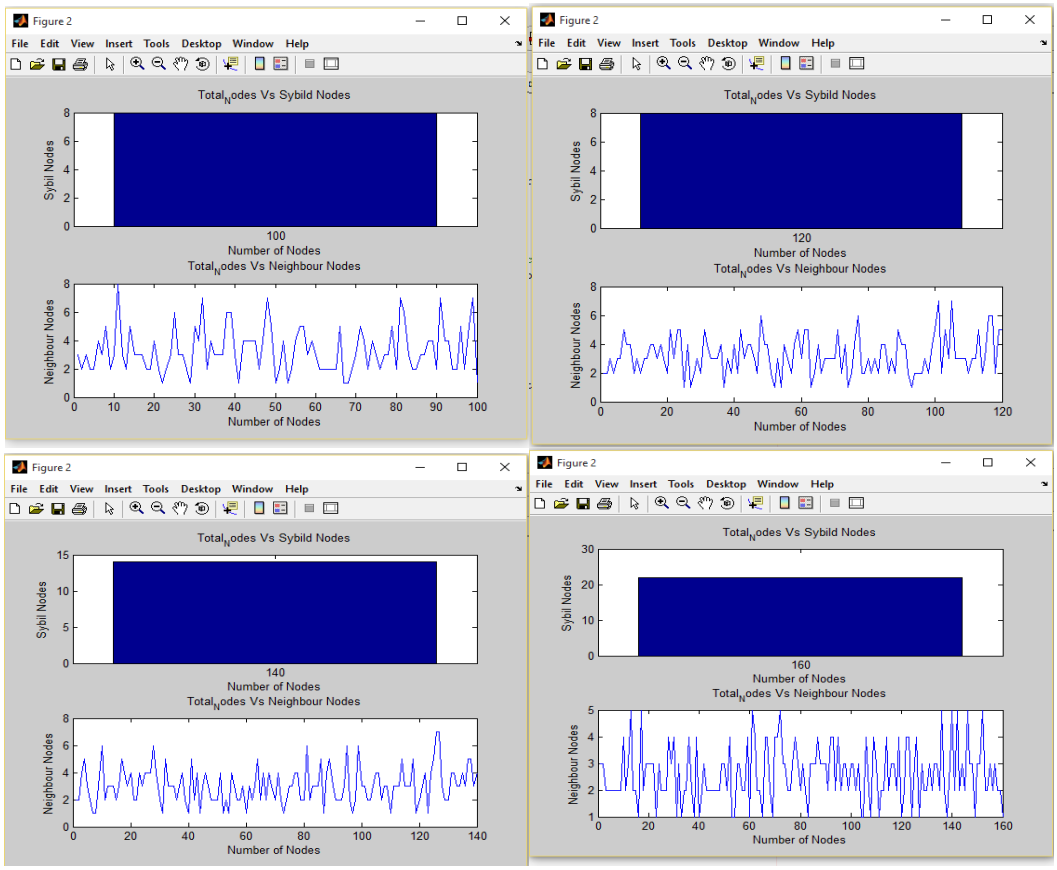


Figure 4: Converse of '0' node identified or threat indentified as Sybil node.

After obtaining the result from the proposed system it is clear that Sybil node detection through the proposed system are more and blockage process indicates localization error is reduced and hence classification accuracy is improved through the proposed system.

Simulation results are plotted in terms of bar chart and are highlighted through the snapshots as under



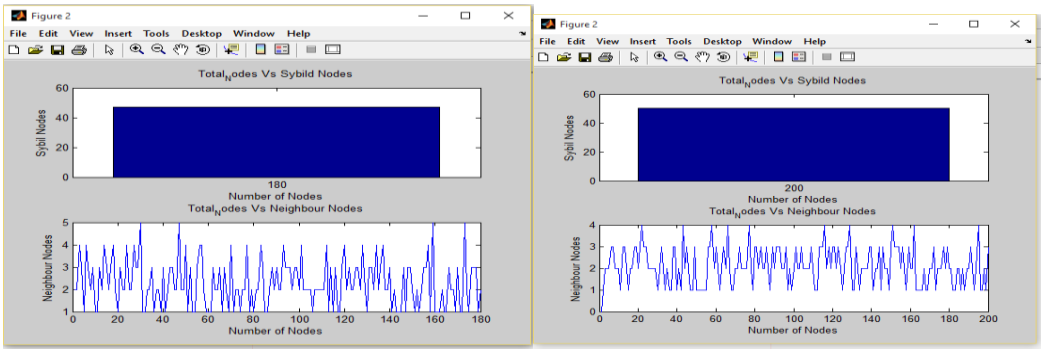


Figure 5: : Nodes variation and subsequent results where nodes=100,120,140,160,200

The number of nodes as increases, chances of sybil attack also increases. The detection process employed in the proposed system not only detect the attack but also block the current nodes that are malcious. The malcious nodes blockage is accomlished by varying the Ids of the nodes using random id generation mechanism.

V. CONCLUSION AND FUTURE SCOPE

Sybil attack is numerous personality assault. Which is utilized with the end goal to presented clog with in the system. Individual KNN approach can distinguish the sybil attack ,anyway can not identify the area of the sybil hub. The proposed methodology anyway can identify both sybil hub and in addition area in less time. The outcome demonstrate improvement as contrast with existing methodology demonstrating worth of the investigation.
In future clustering instruments k-implies clustering can be converged with euclidean separation to accomplish better outcome

REFERENCES

[1] C. Science and K. Mangalore, "A Two-tier Network based Intrusion Detection System Architecture using Machine Learning Approach," pp. 42–47, 2016.
[2] P. Singh and A. Tiwari, "An Efficient Approach for Intrusion Detection in Reduced Features of KDD99 Using ID3 and Classification with KNNGA," *Proc. - 2015 2nd IEEE Int. Conf. Adv. Comput. Commun. Eng. ICACCE 2015*, pp. 445–452, 2015.
[3] K. J. Chabathula, C. D. Jaidhar, and M. A. Ajay Kumara, "Comparative study of Principal Component Analysis based Intrusion Detection approach using machine learning algorithms," pp. 1–6, 2015.
[4] H. Haddad Pajouh, R. Javidan, R. Khayami, D. Ali, and K.-K. R. Choo, "A Two-layer Dimension Reduction and Two-tier Classification Model for Anomaly-Based Intrusion Detection in IoT Backbone Networks," *IEEE Trans. Emerg. Top. Comput.*, vol. 6750, no. c, pp. 1–1, 2016.
[5] A. R. Onik, N. F. Haq, and W. Mustahin, "Cross-breed type Bayesian network based intrusion detection system (CBNIDS)," *2015 18th Int. Conf. Comput. Inf. Technol.*, pp. 407–412, 2015.
[6] Y. Canbay and S. Sagiroglu, "A Hybrid Method for Intrusion Detection," *2015 IEEE 14th Int. Conf. Mach. Learn. Appl.*, pp. 156–161, 2015.
[7] M. Xie and J. Hu, "Evaluating host-based anomaly detection systems: A preliminary analysis of ADFA-LD," *Proc. 2013 6th Int. Congr. Image Signal Process. CISP 2013*, vol. 3, no. Cisp, pp. 1711–1716, 2013.
[8] C. Huijun, S. Hong, and Z. Hong, "Early recognition of Internet service flow," *Proc. - 2013 Wirel. Opt. Commun. Conf. WOCC 2013*, pp. 464–468, 2013.
[9] S. Behrozinia, R. Azmi, M. R. Keyvanpour, and B. Pishgoo, "Biological inspired anomaly detection based on danger theory," *IKT 2013 - 2013 5th Conf. Inf. Knowl. Technol.*, pp. 102–106, 2013.
[10] A. Daneshpazhouh and A. Sami, "Semi-supervised outlier detection with only positive and unlabeled data based on fuzzy clustering," *5th Conf. Inf. Knowl. Technol.*, pp. 344–348, 2013.
[11] T. Weiming and C. Hongzhi, "An Improved Feature Selection Algorithm Based on MAHALANOBIS Distance for Networl < Intrusion Detection," pp. 69–73, 2013.
[12] S. Gopal, Y. Yang, K. Salomatin, and J. Carbonell, "Sctatistical learning for file-type identification," *Proc. - 10th Int. Conf. Mach. Learn. Appl. ICMLA 2011*, vol. 1, no. DiiD, pp. 68–73, 2011.
[13] P. M. Mafra, V. Moll, J. Da Silva Fraga, and A. O. Santin, "Octopus-IIDS: An anomaly based intelligent intrusion detection system," *Proc. - IEEE Symp. Comput. Commun.*, pp. 405–410, 2010.
[14] H. Yu, P. P. K. Chan, W. W. Y. Ng, and D. S. Yeung, "Apply randomization in KNN to make the adversary harder to attack the classifier," *2010 Int. Conf. Mach. Learn. Cybern. ICMLC 2010*, vol. 1, no. July, pp. 179–183, 2010.
[15] Z. Wang *et al.*, "Detecting Malicious Server Based on Server-to-Server Realation Graph," *2016 IEEE First Int. Conf. Data Sci. Cybersp.*, pp. 698–702, 2016.

