IMPROVED INTRUSION DETECTION SYSTEM TO DETECT SYBILL ATTACK USING HYBRID KNN AND EIUCLIDEAN DISTANCE APPROACH IN WIRELESS SENSOR NETWORK

Dr. Sandeep singh kang
Professor and Head of Dept. Computer Science
Global Institute of Management
And Emerging Technologies ,Amritsar(Punjab)

Amandeep kaur
Research Scholar, Dept. of Computer
Science Global Institute of Management And
Emerging Technologies, Amritsar (Punjab)

ABSTRACT

One of the important issues in network security is Intrusion Detection System (IDS). The intrusion detection system is the one in which the detection of malicious attacks that are known or unknown can be done. One of the intrusions is Sybil attack in which a sender node send data to another node but it is received by some another malicious node that may be not original user. In this paper, we use the KNN algorithm to find the location of Sybil node in the network and Euclidean distance algorithm to find the distance between them. This will result in identifying the nearest Sybil node so that we can prevent from sending data to malicious node. Result in terms of time consumption for proposed approach is better as compared to existing approach without Euclidean distance approach.

KEYWORDS : KNN, Euclidean Distance, IDS, Malicious Node.

I.INTRODUCTION

AN Intrusion Detection System (IDS) in terms of Sybil attack is an apparatus for recognizing vindictive exercises or irregular practices in a framework.[1] Since the sort of assaults has expanded significantly in the system, IDS have turned into a vital expansion to security foundation. IDS enables the association to shield their frameworks from the dangers that accompanied expanding system network and dependence on the data frameworks. There are two principle procedures of IDS are called abuse or mark based identification and inconsistency recognition. Signature-

based location uses to recognize known assault designs. Furthermore, irregularity identification has utilized the likelihood of distinguishing examples that are not typical.[1]Most strategies utilized as a part of today's IDS are not ready to manage the dynamic and complex nature of digital assaults on PC systems. Thusly, the efficient versatile strategy like different methods of machine learning can bring about higher discovery rates, bring down false alert rates and sensible calculation and correspondence costs. There are a few customary strategies to identify vindictive like get to control instruments, firewalls, encryptions so on. In any case, these techniques have a few impediments in completely securing systems. Specifically, when the frameworks are expanded in assaults like refusal of administration subsequently they can get high false positive and false negative recognition rates. In as of late, Machine Learning strategies have been utilized to this field with the desire of enhancing recognition rates. There are numerous scientists broaden and apply generally this procedure on IDS. In this paper, we concentrate a few papers that utilization Machine Learning procedures like neural systems for identifying interruption. At that point, we keep on giving another way to deal with enhance execution on this field. Also, we think about the outcomes from these strategies that are connected on IDS.[2] The way toward directing the occasions happening in a PC framework or arrange and investigating them for any plausibility of episodes, or any sort of vindictive assaults, defiled information or any sort of interruption that can posture danger to our frameworks is called interruption recognition. IDPS innovations can fall in different classifications that are: system based, remote, have based and arrange conduct investigation based. It should likewise be possible on different levels, for example, arrange layer, transport level and application layer. In our paper, we might want to focus on system layer and we might want to execute arrange conduct investigation based IDPS. An IDS by and large needs to manage a few issues, for example, tremendous system clog very unpredictable examples based information, qualification between general movement and interruptions and getting acclimated to the always showing signs of change system conditions. Organize conduct can be named abuse location and irregularity discovery. Abuse location procedures chips away at a known arrangement of prepared information or examples of understood assaults to the

framework to identify interruptions. Peculiarity identification strategies manages those information which demonstrates a deviation from regular arrangement of typical conduct or examples which are obscure to the framework to identify interruptions . [3]In the ancient times, there were couple of gatecrashers thus the client can oversee them effectively from the known or obscure assaults. As of late the security turns into the most difficult issue in issues of securing information or data year over year. Since the gatecrashers present another assortment of interruptions in the market, with the goal that client can't deal with their PC framework or system. Interruption recognition assaults can be ordered into two gatherings: abuse or mark based and peculiarity based interruption location. The abuse or mark based interruption location framework identifies the interruption by contrasting and its current marks in the database. The recognizing assaults and marks are coordinating, it's an interruption. The mark based interruptions are called known assaults at whatever point the clients are identifying the interruption by coordinating with the marks log records.[4] The log document contains the rundown of known assaults identifying from the PC framework or systems. The irregularity based interruption recognition is called as obscure assaults and this assault is seen from system as it goes astray from the typical assaults. The interruption discovery frameworks are named Network based or Host based assaults. The system based assault might be either abuse or oddity based assaults. The system based assaults are identified from the bury association of PC frameworks. The framework can speak with each other, so that the assault is sent starting with one PC framework then onto the next PC framework by the method for switches and switches. The host based assaults are identified just from a solitary PC framework and is anything but difficult to keep the assaults. These assaults for the most part happen from some outside gadgets which are associated. The outside gadgets are pen drive, CD, VCD, Floppy and so forth. The electronic assaults are conceivable when frameworks are associated over the web and the assaults can be spread into various frameworks through the email, talking, downloading the materials and so forth.

II.RELATED WORK

Intrusion Detection System (IDS) is one of the essential issues in system security. IDSs are worked to recognize both known and

obscure noxious assaults.[5] A few machine learning calculations are utilized broadly in IDS, for example, neural system, SVM, KNN and so forth. Be that as it may, these calculations have still a few restrictions, for example, high false positive and false alert rate. In this paper, our commitment is to assemble a classifier of IDS taking after profound learning approach. We find the most reasonable streamlining agent among six improves for Long Short-Term Memory Recurrent Neural Network (LSTM RNN) model are utilized to IDS. Through our tests, we found that LSTM RNN display with Nadam analyzer beats to past works.[6] We show our approach is truly efficiency to interruption location with exactness is 97.54%, discovery rate is 98.95%, and the false alert rate is sensible with 9.98%.Intrusion detection frameworks are frameworks that can identify any sort of malevolent assaults, undermined information or any sort of interruption that can posture risk to our frameworks. In our paper, we might want to introduce a novel way to deal with fabricate a system based interruption identification framework utilizing machine learning approach. We have proposed a two-level engineering to identify interruptions on system level. Arrange conduct can be delegated abuse identification and inconsistency recognition. As our investigation relies on upon the system conduct, we have considered information parcels of TCP/IP as our info information. After, pre-handling the information by parameter separating, we assemble a self-governing model on preparing set utilizing various leveled agglomerative grouping. [7]Further, information gets named customary movement example or interruptions utilizing KNN arrangement. This lessens cost-overheads. Abuse identification is directed utilizing MLP calculation. Peculiarity identification is directed utilizing Reinforcement calculation where arrange operators gain from the earth and take choices in like manner. The TP rate of our design is 0.99 and false positive rate is 0.01. Along these lines, our design gives an abnormal state of security by giving high TP and low false positive rate. Furthermore, it additionally breaks down the standard system designs and adapts incrementally (to construct independent framework) to separate ordinary information and dangers KDDCUP 1999 Dataset broadly utilized dataset of information mining in the field of interruption discovery by different analysts.[8] This dataset are freely accessible for the clients. Interruption

identification is the key difficulties for the clients on the grounds that the interruption may degenerate or demolish the system administrations. The interruption location framework is grouped into two classifications: Network based interruption identification framework and Misuse interruption recognition framework. In this paper, novel strategy is for interruption recognition with highlight diminishment utilizing in part ID3 calculation to discover higher data pick up for property determination and KNN based GA (hereditary calculation) is connected for order and discovery of interruptions on KDD dataset. The reenactment and investigation of the strategy is done on MATLAB2012A. [9]The test situation of proposed technique creates better outcome when it contrasted and some current methodologies, for the estimation of the outcome contrasting and the diverse execution measurements parameters, for example, affectability, specificity and precision. This paper induces the unmistakable quality of variegated machine learning strategies adjusted so far for the recognizing distinctive system assaults and proposes an ideal Intrusion Detection System (IDS) with the accessible framework assets while improving the speed and exactness. With blasting number of gatecrashers and programmers in todays huge and advanced automated world, it is endlessly testing to distinguish obscure assaults in promising time with no false positive and no false negative. Essential Component Analysis (PCA) diminishes the measure of information to be contrasted by lessening their measurements earlier with classification that outcomes in decrease of identification time. In this paper, PCA is embraced to decrease higher measurement dataset to lower measurement dataset. It is refined by changing over system parcel header fields into a vector then PCA connected over high dimensional dataset to lessen the measurement. [10]The diminished measurement dataset is tried with Support Vector Machines (SVM), K-Nearest Neighbors (KNN), J48 Tree calculation, Random Forest Tree classification calculation, Adaboost algorithm, Nearest Neighbors summed up Exemplars calculation, Navebayes probabilistic classifier and Voting Features Interval classification calculation. Acquired outcomes exhibits discovery exactness, computational efficiency with negligible false cautions, less framework assets usage. Trial results are contrasted with deference with recognition rate and discovery time and

found that TREE classification calculations accomplished better outcomes over different calculations. The entire investigation is directed by utilizing KDD99 informational index. With expanding dependence on Internet of Things (IoT) gadgets and administrations, the capacity to distinguish interruptions and noxious exercises inside IoT systems is basic for versatility of the system foundation. In this paper, we introduce a novel model for interruption identification in light of two-layer measurement diminishment and two-level grouping module, intended to recognize pernicious exercises, for example, User to Root (U2R) and Remote to Local (R2L) assaults. The proposed model is utilizing part examination and direct segregate investigation of measurement diminishment module to spate the high dimensional dataset to a lower one with lesser elements. We at that point apply a two-level grouping module using Naïve Bayes and Certainty Factor variant of K-Nearest Neighbor to distinguish suspicious practices.[11] The examination comes about utilizing NSL-KDD dataset demonstrates that our model outflanks past models intended to identify U2R and R2L assaults. [10]Advanced web is trickster of the critical system assaults because of intemperate utilization and enormous network requests. Machine learning is an effective way to deal with keep the interruption and characterize the system assaults. This review highlights the joined energy of channel methodologies in interruption identification structure. Include choice procedure expels the repetitive components and assembles a tedious better-performed interruption identifier structure. This review introduces a cross-breed sort highlight choice approach utilizing couple channel plans for interruption recognition. In this system include choice procedure take out the superfluous elements to decrease the time intricacy and construct a superior model to foresee the outcome with a more prominent precision and Bayesian system based grouping model has been developed to anticipate the sorts of assaults.[12] The trial demonstrates that the proposed system shows an unrivaled general execution regarding exactness which is 97.2746% and keeps the false positive rate at a lower rate of 0.008. The model shows better execution as far as precision than other driving state-of the-expressions systems like Boosted DT, Hidden NB, KNN and Markov chain. The NSL-KDD is utilized as benchmark informational collection with Weka library works in the exploratory setup.

III.PROPOSED SYSTEM

The proposed system uses hybridization of KNN and Euclidean distance in order to detect the intrusion at fast rate. The implication of KNN and Euclidean is describes as under

IV.IMPLICATION OF KNN

The dataset can be dissected by the utilization of K-closest neighbor system.[13] Let the estimation of K=2 and suggestion is doled out on X,Y and Z. The test purpose of x is thought to be 120. Test guide y is considered toward be 143.The test purpose of z is 99. At that point KNN creates

Id	X= $\sum \frac{X-120}{N}$	Y=$\sum(Y-143)/N$	Z=$\sum(Z-99)/N$
1	2.16	2.66	0.66
2	2.16	4.83	2

Table 1: KNN deviation results.

The choice limit is made for identifying and estimating reason if there should be an occurrence of KNN. The choice limit is built up by the utilization of guidelines. These tenets are as IF-THEN shape. The limit if thought to be 2.5 for X, 3 for Y and 1.5 for Z then X is misrepresented, individual 2 is recognized with ailment comparing to Y and Z.

V.IMPLICATION OF EUCLIDEAN DISTANCE

Euclidean separation is utilized to compute the separation of noted dataset vales from the test point. The limit esteem is contrasted with gotten an incentive with decide peculiarities. [14] The test purpose of x is thought to be 120. Test direct y is considered toward be 143.

Id	X= $\sqrt{\frac{(X-X_i)^2}{Total_{observatio}}}$	Y= $\sqrt{\frac{(Y-Y_i)^2}{Total_{observatio}}}$	Z= $\sqrt{\frac{(Z-Z_i)^2}{Total_{observatio}}}$
1	2.7	3.5	0.93
2	2.7	5.53	3.12

Table 2: Euclidean Distance Results

The limit if thought to be 2.5 for X, 3 for Y and 1.5 for Z then Person 1 is recognized with Z, individual 2 is identified with sickness relating to X, Y and Z. The hybridization of KNN + Euclidean separation can be utilized as a part without bounds look into for enhancement.

Proposed technique uses KNN to determine Sybil attack and Euclidean distance mechanism is used to determine location of the intruder.[15] Proposed algorithm is listed as under

Algorithm

The methodology to achieve the objectives is listed as follows

1. Input the number of nodes in the networks.
2. Enter the threshold coverage area(Ct) associated with node.
3. Initialize count=0
4. Check the neighbourhood of nodes in terms of coverage area(Ci)
   4.1 if Ct>Ci then
   Count=count+1
   End of if
5. if count=1 then
   5.1 Apply Euclidean distance to determine location of attacking node
   5.2 If Ct>Ci then
   5.3 Declare Sybil attack along with its location
   End of if
6. Repeat the above steps for all the nodes
7. Calculate lifetime, packet drop ratio and energy consumed
8. Stop

VI.RESULT AND PERFORMANCE ANALYSIS

Sybil attack will be the one in which one node takes the identity of other node. The overall performance goes down by the application of Sybil attack. In order to resolve the problem Euclidean distance mechanism is merged along with KNN approach. KNN used to find the neighbors of the node being analyzed. Incase their exist only one neighbor of current node then Sybil attack is detected the Euclidean distance is used to check the location of the Sybil node. The overall time consumption of simulation is achieved to be better as compare to existing approach. This is shown as under

PROPOSED KNN+EUCLIDEAN	EXISTING KNN
12.5357	22.4715
36.6243	44.4277
48.6805	64.4345
46.7414	60.4107
73.0829	98.9666
101.205	113.473

Table 3:- Showing time consumption of existing and proposed system

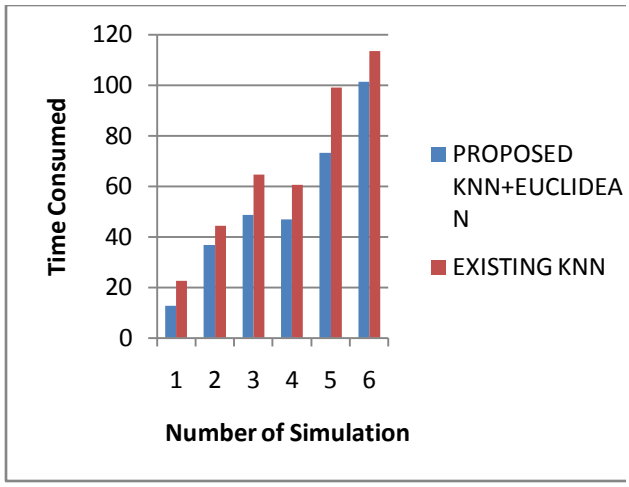
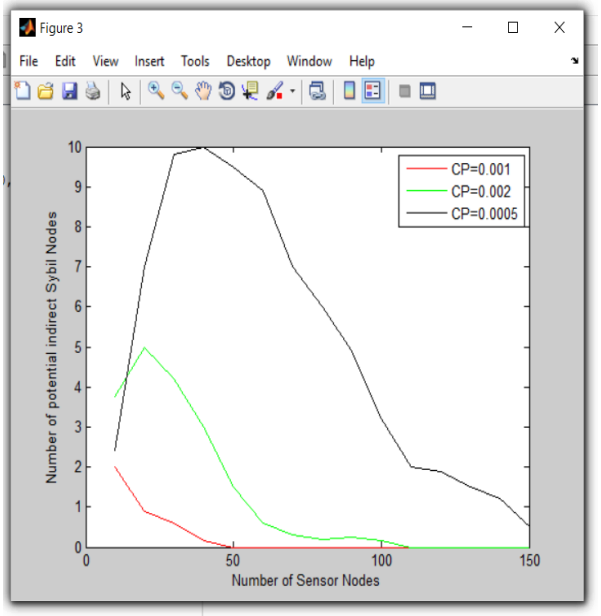
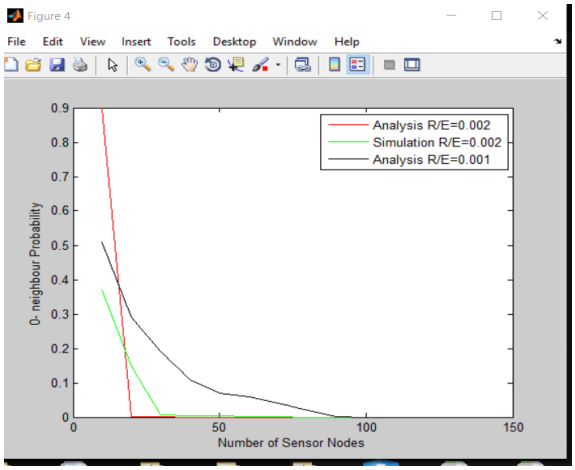


Figure 1:- Showing time consumption of existing and proposed system

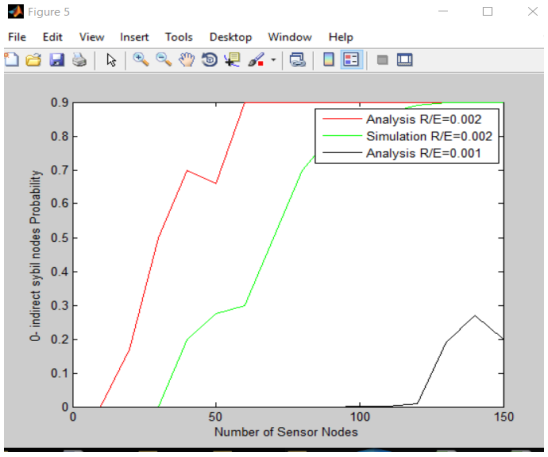
The simulation is conducted in MATLAB and Sybil nodes are recorded . the no of nodes are varied from 100 to 200 and result is recorded. The snapshot generated from proposed system is as under



This MATLAB plot indicates number of potential Sybil attack nodes.



Nodes having 0 neighbors are indicated through the proposed system.



O probability neighbor node attacks are predicted through this graphs. As the detection is more accurate hence less chances of attack and indirect attack probability decreases.

The result obtained from MATLAB simulation is given as under

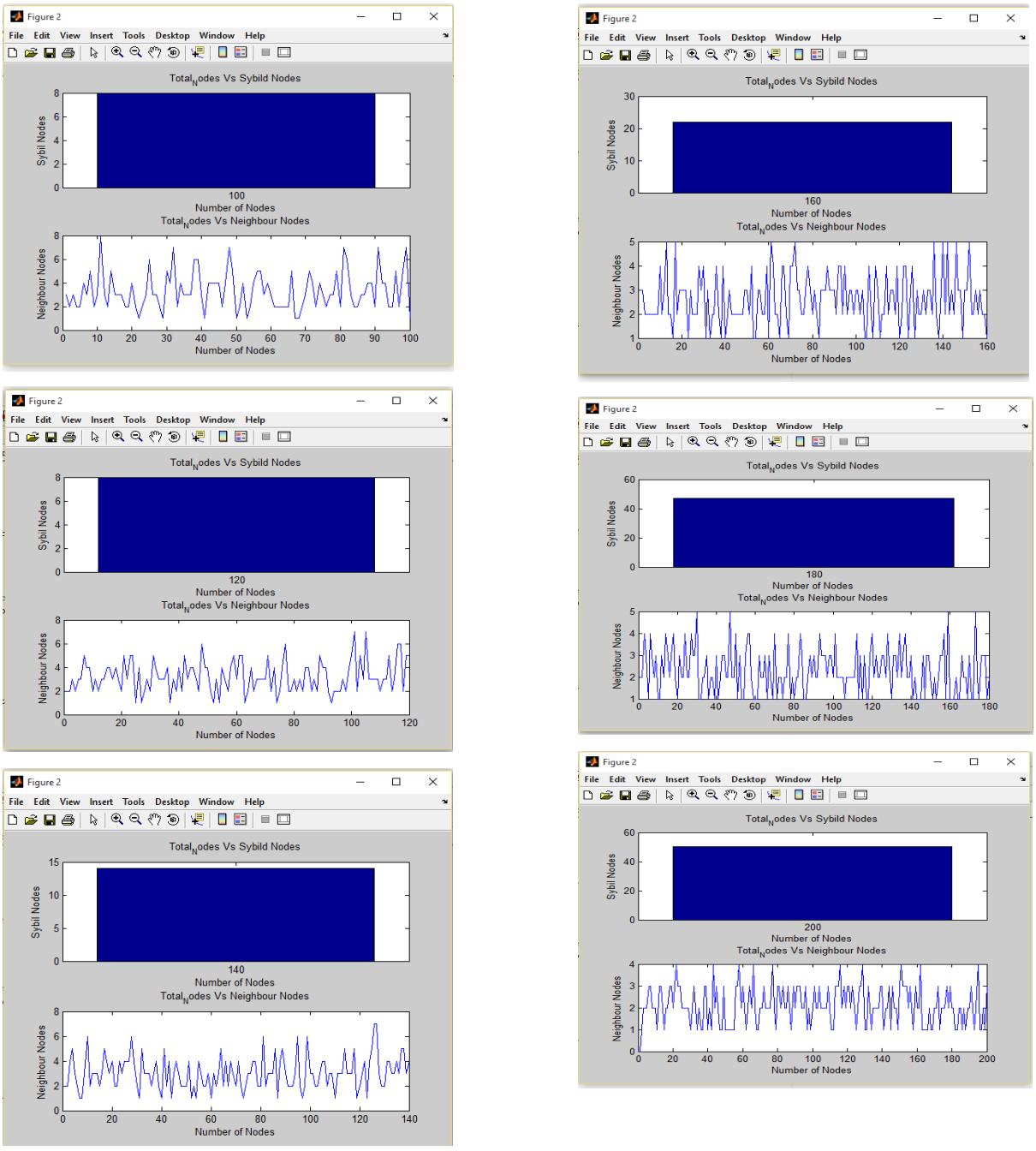


Figure 2: Detection results in terms of 100,120,140,160,180 and 200 Nodes

As the number of nodes increases sybil attack is also enhanced. The detection process shows time consumption is greatly reduced in determining location of sybil nodes.

VII.CONCLUSION AND FUTURE SCOPE

Sybil attack

Sybil attack is multiple identity attack. Which is used in order to introduced congestion with in the network. Individual KNN approach can detect the sybil attack ,however can not detect the location of the sybil node. The proposed approach however can detect both sybil node as well as location in less time. The result show betterment as compare to existing approach proving worth of the study.

In future clustering mechanisms k-means clustering can be merged with euclidean distance to achieve better result

VIII.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.