Towards Improved Random Forest based Feature Selection for Intrusion Detection in Smart IOT Environment

Suresh B, Venkatachalam M and Saroja M

Abstract*: Internet of Things (IoT) is raised as most adaptive technologies for the end users in past few years. Indeed of being popular, security in IoT turned out to be a crucial research challenge and a sensible topic which is discussed very often. Denial of Service (DoS) attack is encountered in IoT sensor networks by perpetrators with numerous compromised nodes to flood certain targeted IoT device and thus resulting in vulnerability or service unavailability. Features that are encountered from the malicious node can be utilized effectually to recognize recurring patterns or attack signature of network based or host based attacks. Henceforth, feature extraction using machine learning approaches for modelling of Intrusion detection system (IDS) have been cast off for identification of threats in IoT devices. In this investigation, Kaggle dataset is measured as benchmark dataset for detecting intrusion is considered initially. These dataset includes 41 essential attributes for intrusion identification. Next, selection of features for classifiers is done with an improved Weighted Random Forest Information extraction (IW-RFI). This proposed WRFI approach evaluates the mutual information amongst the attributes of features and select the optimal features for further computation. This work primarily concentrates on feature selection as effectual feature selection leads to effectual classification. Finally, performance metrics like accuracy, sensitivity, specificity is computed for determining enhanced feature selection. The anticipated model is simulated in MATLAB environment, which outperforms than the existing approaches. This model shows better trade off in contrary to prevailing approaches in terms of accurate detection of threats in IoT devices and offers better transmission over those networks.

Keywords: Internet of Things; DoS attacks; Security; feature selection; improved weighted random forest information.

Revised Manuscript Received on September 2, 2019.

Mr.B.Suresh*, Research Scholar, Department of Electronics, Erode Arts and Science College (Autonomous), Erode, Tamilnadu, India and Assistant Professor, Department of Electronics and Communication Systems, VLB Janakiammal College of Arts and Science College (Autonomous), Coimbatore, Tamilnadu, India.

Dr.M.Venkatachalam, Associate Professor and Head, Department of Electronics, Erode Arts and Science College (Autonomous), Erode, Tamilnadu, India.

Dr.M.Saroja, Associate Professor, Department of Electronics, Erode Arts and Science College (Autonomous), Erode, Tamilnadu, India.

I. INTRODUCTION

In general, owing to the faster progressing technologies of network communication, Internet facilitates connectivity of people from everything to everywhere. With this, a novel idea of IoT came into existence and related with upcoming 5G connectivity [1]. The ultimate target of IoT is to connect huge amount of heterogeneous devices like wireless sensor networks (WSN), instance cameras, vehicles, smart meters by offering open access to data produced by these devices to give services to companies and citizens [2]. Moreover, IoT resources are front end constrained, numerous security methods are complex to defend IoT [3]. Certain lightweight encryption approaches are assumed as core technology to build security mechanisms for IoT devices, as well considers the rising amount of hackers computation ability (with cloud computing, distributed computing and Quantum computing and so on), these cryptographic approaches turns to ne crushed in future [4]. Alternative kind of security methods are intrusion detection system which is considered for protecting IoT networks.

In general, IDS is considered as an effectual approach to recognize attackers while encountering drawbacks in cryptography [5]. It will identify the policy violations or malicious activities by observing system activities or network traffics [6]. IDS are generally third party software or stand by device which does not inquire numerous modifications in present system [7]. It is appropriate for inherited system or resource constraint to guard its network security. Numerous recent investigations have observed security issues in IoT systems, and many intrusion detection approaches as in fig 1 are developed and anticipated [8]. Moreover, most of the anticipated approaches are still restricted to data mining and just provides an intrusion perspective of MANET, WSN, Zigbee and other IoT subnets [9]. As well, a uniform IDS for entire IoT is discussed infrequently. In mean time, packets are excavating and usually training of statistic features needs numerous computational resources [10], where these methods are very complex to execute in certain IoT environments.

Here, a novel feature selection approach for making effectual classification of attacks in IDS is anticipated and it is termed as a Weighted Random Forest Information extraction (IW-RFI). Here, the method considers an IoT network model and Kaggle dataset for pre-processing purposes.

The essential features needed for classifiers are attained using the proposed model and measured with performance metrics like accuracy, sensitivity, specificity. The feature selection model and algorithm can have the ability to identify intrusion by comparing the level of extracted features from other approaches and computes the complexity of identification.

a. Paper Contribution

With the use of proposed model, numerous complexities in extracting the highly influencing features of attacks are considered and solved. In this work, machine learning based random forest scheme is considered to solve security issues with mathematical estimations in heterogeneous IoT networks and model an essential part of intrusion detection method for IoT network, anomaly detection and features are determined and proposed. In contrary to prevailing approaches, advantages of the anticipated model are given below:

1. To the best of our knowledge, this approach is considered as a primary model to determine the highly influencing attack features to identify intrusions in IoT networks. With this anticipated feature selection approach, IoT system can be stabilized during the changing network characteristics. This also facilitates efficiency for classifiers.
2. This work defines a pre-processing phase of an intrusion detection system using anticipated IW-RFI method.
3. The attained features are given to classifier for enhanced classification in next phase of work. This also shows action flows and possible detection rate of false alarm.
4. The extracted features are analyzed with respect to performance metrics like accuracy, specificity and sensitivity. As well, it's influencing characteristics in detecting intrusion in IoT devices.

The following levels of work is categorized as below: In section II, background of IDS in IoT, problems in implementing it and related works for modelling an IDS for IoT devices are considered. In section III, complete proposed model termed IW-RFI is discussed in detail. In section IV, numerical results associated with feature selection and its related parameters are provided for analyzing IoT system. Section V depicts the conclusion of the proposed model, as well as the future direction for progress the proposed model.

II. Related works

This section explains in detail about existing IDS approaches for IoT and its corresponding merits and demerits.

a. Intrusion detection system for IoT

In current times, progression of Virtual reality, intellectual hardware and IoT, IDS under IoT has also turns to trending in information technology growth [11]. Moreover, investigators on those problems are still in infancy. Generally, IoT has been considered as an extensive heterogeneous network, most prevailing approaches commence to initiate IoT elements to determine appropriate intrusion detection technique.

In [12], Jover et al. anticipated a hybrid intrusion detection approach sourced on game theory utilization, which mixes the anomaly and signature utilization for IoT devices intrusion detection. By developing gaming model to normal users and intruders, Nash equilibrium was computed and utilized to determine IDS based anomaly model.

In [13], Li et al. depicted a real-time pattern matching pattern with CEP utilization. Benefits of this approach is utilizing this event flow features to identify the intrusions type, which can diminish false alarm rate in contrary with conventional intrusion detection approaches. Even though, these approaches will raises system computing resources and energy consumption, it is clear by decreasing feedback delay of corresponding IDS.

In [14], Yung et al. concluded certain IDS approaches which are an essential network structure that belongs to IoT. By examining and evaluating attack techniques and detection, this work evaluated prevailing GIDP, CRADS and other intrusion detection structures for MANET.

Usually, IDS is network/host based system. Host sourced IDS evaluates events significantly associated to OS, whilst network sourced IDS examines network associated events, like traffic, IP addresses. Eventually, in accordance to the intrusion detection manner, two significant IDS classification are elaborately depicted: anomaly IDS and misuse IDS [15]. Former IDS construct profiles of non-anomalous characteristics of computer systems' active subjects and latter utilizes templates or traces of known attacks.

In [16], Wu et al. executed a model that utilizes back propagation and feed forward algorithms to identify and categorize cyber attacks in desktop networks. Moreover, to measure their system, they cast off NSL-KDD dataset and tries to categorize U2R, Probe, DoS and R2U attack. Nonetheless, there is no proof that these systems are effectual while deployed in heterogeneous IoT environment, which comprise of numerous more protocols, network behaviours and devices.

In [17], Sherubha et al, utilized machine learning algorithms in IoT environment to identify Distributed denial of service (DDoS) attacks. They illustrate that it concentrates on IoT specific network characteristics (for instance: restricted amount of endpoint and regular time interval amongst packets) to demonstrate feature section outcomes with higher accuracy of DDoS detection in IoT network traffic with diversity of machine learning algorithms.

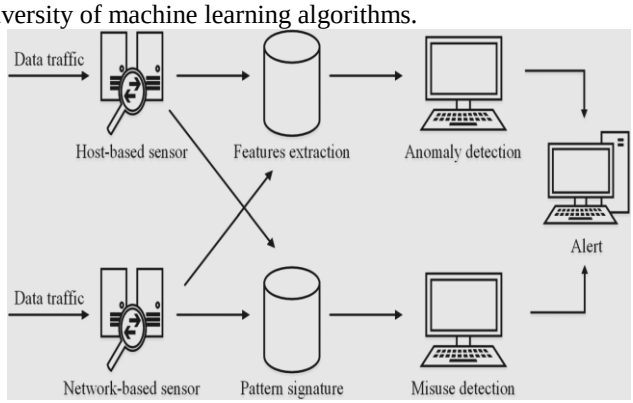

Fig 1: IDS for Smart IoT environment

In [18], Wang et al. analyzes security threats that are encountered in IoT networks and examines the potential security solution which utilizes machine learning to recognize and mitigate attacks with polymorphic hardware and software.

Moreover, no description of experimental setup, execution and consequent measurement of anticipated system is offered. In [19Tang et al. anticipated a distributed and a centralized framework for hybrid IDS, which they executed sourced on simulated networks and scenarios. It concentrates on recognizing routing attacks like wormhole attacks. In [20], Chen et al. provided an event processing based IDS for IoT environment. This model is specification based and it cast off complex event processing approached for attack identification. These models gathers all data from IoT devices, hauls out diverse events and carry out security event identification by performing match events with rules stored in Rule Pattern repository. Even though, it is more complex than conventional IDS, it is CPU intensive.

Even though these prevailing approaches can resolve IDS crisis from diverse level, an uniform IDS is essential to provide a complete intrusion based perspective of IoT.

III. Proposed method

This section discusses in detail about Intrusion detection in smart IoT environment. Initially, dataset is chosen for pre-processing. Next, feature extraction is carried out using weighted Mutual correlation information based random forest. Finally, performance parameters are evaluated.

IoT general concept

In general, the concept of IoT is utilized as objects that are interconnected to one another and facilitate people to communicate and produce IoT environment for cities, energy, health and so on [21]. IoT functions entirely in isolated circumstances, and never modelled to deal with threats, it is susceptible to attacks, moreover due to its' progress dense growth, limited resources and deployment. IoT's distributed and heterogeneous features lead it to be complex to apply standard security method, leading the system to make dangerous and wrong actions.

Therefore, IDS is one amongst appropriate approaches to offer security, while detecting attacks. Vulnerabilities are both economically costly and difficult. Moreover, IDS act as special purpose devices to identify anomalies is more crucial [22]. This work anticipates the effectual characteristics that influence IoT for detection to utilize scalable and security in IoT as in fig 2. The anticipated model is sourced on random forest based information generation. This work carry out an analysis of essential features for computation of threats in IoT, thus the IoT device performance is not degraded.

Dataset description

Usually, in network intrusion detection (IDS), anomaly based model in specific suffers from appropriate evaluation, deployment and comparison which commences from lack of adequate datasets. Numerous such datasets cannot be provided owing to its privacy concerns, others are extremely anonymized and do not reflect recent trends, or they deficient in some statistical features [23]. These deficiencies are essential cause for provide the reason for existence of dataset. Therefore, researchers has to resort datasets which they can acquire suboptimal outcomes.

As pattern change and network behaviours and intrusions evolve, it has extremely much turns to be essential to move from one time dataset and static datasets towards more dynamically produced datasets. It do not reflect intrusions and traffic compositions, however more extensible, modifiable and more reproducible.

Kaggle dataset

In general, dataset to be audited was offered with extensive intrusions generated over IoT environment. It generates a situation to attain TCP/IP based data by executing US based LAN Air force. Here, it is concentrated in real environment and blasted with numerous attacks. Connections are series of TCP packets that commences and terminates during certain time amongst data flows from source to target IP address under certain protocol. As well, every connection is labelled as normal or attack with specific type. Every record comprises of roughly 100 bytes. TCP/IP connection has 41 qualitative and quantitative features are acquired from data (38 quantitative features and 3 qualitative features). Variables comprises of two categories. One is Normal and other is anomalous.

Improved Random forest

After the selection of dataset, feature selection has to be done to recognize the most influencing feature that degrades the IoT device performance [24]. Here, RF approach is considered as meta-estimator that balances decision tree classifiers over diverse dataset sub-samples.

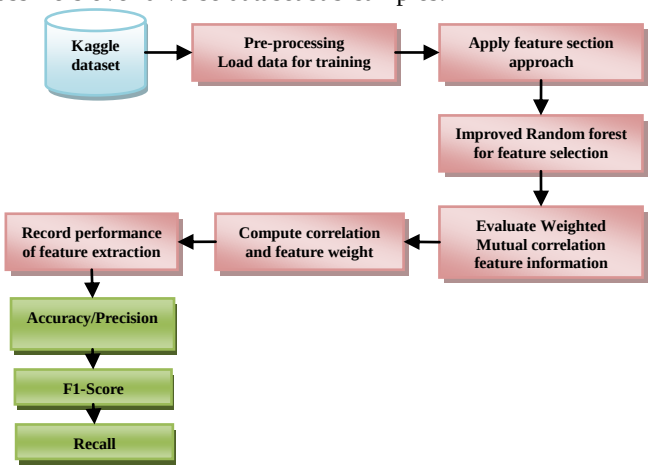

Fig 2: Flow diagram of proposed model

It utilizes moderating to enhance control over fitting and predictive accuracy. It is supervised classification approach which generates forest based decision trees. RF is generally an averaging numerous decision trees, trained over diverse fields of training set, which attempts to decrease variance. Classification approach is utilized to design tree with diverse bootstrap instances from primary data. When forest construction is performed, object to be classified is considered from every tree in forest. New objects are generated from the mutual correlated information attained from trees of forest.

Mutual correlation information (CI_M)

Here, Mutual correlation information (CI_M) is a measure of relation between two random variables, that is, context of selecting features amongst one single feature (S_F) and target feature (T_F) (for instance, labels). Lower case specified features realization between random variables. This is given as in Eq. (1):

$$CI_M(S_F,T_F) = \int_{S_F}\int_{T_f} p(S_F,T_f)\log\frac{p(S_F,T_f)}{p(S_F)p(T_f)}\, dtd_{S_i} \quad (1)$$

If CI_M is considered as zero, both correlated feature variables are independent and comprises no information regarding one another, therefore, features are considered to be irrelevant for target. Higher CI_M value specifies more information regarding target and therefore with higher correlation between variables. Clearly, simple feature ranking selects an appropriate amount of highest ranking features and entire features that are above threshold as appropriate learning machine features.

The concept of weighed Mutual correlation information is provided in Eq. (2), it is more suitable for numerous IoT applications. It is provided as:

$$wCI_M(S_F,T_F,W) = \int_{S_F}\int_{T_f} w(S_F)p(S_F,T_f)\log\frac{p(S_F,T_f)}{p(S_F)p(T_f)}\, dtd_{S_i} \quad (2)$$

For every class of samples S_F, input value $'t'$ and appropriate value $'t'$, weight $w(S_F) \geq 0$ is provided. It provides certain significance amongst every unique sample and not like certain combination of information. This is used to weighted influence of diverse samples.

Weighted Mutual Correlation Information of features (CI_{WM})

Significant problem associated with above equation is the evaluation of essential probability densities from available data. The appropriate method to evaluate CI_M utilizes histogram approximation, simplifies equation indeed of integrals to kernel density estimation approaches. Comparison amongst two sophisticated approaches for evaluating CI_M is found and specifically utilized in feature selection module of IoT based intrusion detection.

Computation of Weighted Mutual Correlation Information of features (CI_{WM}) is essential for features in dataset to recognize association amongst the data or attributes of IoT device classes. It is easier to compute the correlation and it is provided below as in Eq. (3):

$$wCI_M(S_F,T_F,W) = \int_{S_F}\int_{T_f} w(S_F)p(S_F,T_f)\log\frac{w(S_F)p(S_F,T_f)}{w(S_F)p(S_F)p(T_f)}\, dtd_{S_i} \quad (3)$$

In fact, this approximates weighted Mutual correlation Information by probability distribution manipulation. Every sample offers probability density function in accordance to weights (zero weight sample will not contribute much), which is evaluated with particle representation used in prevailing particle swarm optimization based filters.

With the utilization of adaptive histogram approach, that discretizes data in accordance to certain strategies and substitute integrals of Eq. (1) with sum in discrete histogram, utilization of feature weights are trivial. Every sample does not offer mutually to bins, however in accordance to weight. Similar to existing weighted KDE variant, sum of pair-wise interaction kernels is evaluated. Weighting is executed by sample's location manipulation.

Other existing approaches are k-NN estimators that are not sourced on formulation of Kullback-Leibler divergence, however entropy estimators are not effectually transformed to evaluate Mutual Correlation Information.

Weighted Mutual Correlation Information (WMCI)

Here baseline concept of anticipated feature selection is trailed as follows: An approximator is provided along with its error function, choice of determining the feature to enhance performance is sourced on errors that are made and not over all the available data of kaggle dataset. This is performed using weighting of appropriate or inappropriate classified samples diversely.

Algorithm 1:

Input: Kaggle Intrusion detection dataset for observation and its labels;

Output: Feature set

W→ sample data weight

While

Stopping criterion is not fulfilled do

$F_{max=}$ find feature with maximum weighted Mutual correlation information using Eq. (3)

$$wCI_M(S_F,T_F,W) = \int_{S_F}\int_{T_f} w(S_F)p(S_F,T_f)\log\frac{w(S_F)p(S_F,T_f)}{w(S_F)p(S_F)p(T_f)}\, dtd_{S_i}$$

S → accumulate subset feature

F → eradicate feature from candidate set

Classify → appropriate feature

W → residual each sample with new weight

Validate stopping criterion ( )

End while

Simple instance is that, if classifier generates discrete information. Subsequently, appropriate samples classified are left over for evaluation of Mutual correlation information amongst target and samples, as they have zero weight. Only wrongly classified data samples are utilized, they possess an equal weight.

Continuous predictor utilization is facilitated for diverse weighting of every sample in accordance to residual. For instance, sample that is classified as positive, however it is nearer to negative class boundary which offer non-zero residual indeed of being in appropriate class. However, influence is lower in contrast to sample over wrong side of decision boundary.

Therefore, weighted mutual correlation information (WMCI) is applied in this work for feature selection framework for IoT devices. This commences with empty feature, then MCI, to be specific, WMCI with equal sample weight is evaluated amongst feature variables and corresponding target. Feature provides maximal mutual correlation information as chosen in simple ranking procedure. Nonetheless, an approximator is trained using variable.

Residual outcome of every sample is with an interest, that it is utilized to determine weights for subsequent selection round using WMCI. Subsequent feature is selected based on maximal WMCI and approximator is re-trained by in co-operating new feature channel. This process is repeated till stopping criteria is fulfilled.

This is based on baseline concept Adaboost algorithm. Instances are misclassified are provided with higher significance for subsequent round, while appropriate samples are low significant. Reason is easy: appropriately classified instances are depicted feature subset and there is a necessity to determine features that projects misclassified samples.

Here, scaling values for targeted value are arbitrary, as absolute WMCI value is not significant however relative value to other features can be evaluated with similar weights. By considering the global function approximators, such as multi layer preceptors, this is not a huge problem. They can be competent to determine decision surface they found in sub-space of spanned features comprising novel chosen feature. New dimensions gather more options to determine superior decision surface, however similar result is achievable always.

As well, approximators with local activation functions such as nearest neighbour and RBF network classifiers, it is considered to be complex, specifically for low dimensional cases. Sample neighbourhood can be modified with addition of new features dramatically [25]. Clearly, this influence is less dramatic in higher dimensional space as new features affects neighbourhood. However, the entire performance of an approximator may reduce in preliminary stages as a result. The algorithm attempt to accurate when new residuals and selected features balances the newly generated errors, however as an outcome, error rates reduces drastically. Determining an optimal stopping criterion is complex and crucial, specifically for local approximators [26 – 36]. Anticipated algorithm is initiated with empty feature subset and adds correlated variables to determine subset in every step till pre-defined amount of features is attained, or approximation outcome does not enhanced in future.

If finest subset is raised more than threshold value, this is negative and leads to performance degradation. New subset will be determined for subsequent rounds, else algorithm terminates. Other probable stopping criteria with fixed round compensates selected features or approximation error in resulting approximator.

IV. Numerical results

This part depicts in detail about features selected for deploying effectual IoT and enhances performance of IDS. Simulation is carried out in MATLAB 2018a, for classification of correct and incorrect feature for IDS modelling.

Classification of normal and most influencing features

So as to measure the performance of Weighted Mutual Correlation Information based feature selection, Kaggel dataset is used. Kaggle dataset is considered as an IDS dataset with benchmark standard. This facilitates simulation over real time IoT data transmission. The outcome demonstrates that Weighted Mutual Correlation Information based random forest works effectually. Confusion matrix is cast off to measure performance. It is also assisting in recognizing which classes of algorithm are not classified appropriately. F1-score, recall and precision were evaluated as in Table II and Table III.

Table II: Performance metrics

Class	Precision	Recall	F1-score
0	100	97	80
1	87	100	94
Total	97	96	95

Table III: Confusion Matrix

	Normal	Attack
Normal	TP	FP
Attack	FN	TN

Classification of Intrusion features

Intrusion classes specified with 0-9 numbers appropriately to labels as in fig 3. For instance, normal class precision is specified using number 6. This is 1 roughly and recall is 98 roughly. Result demonstrates that most classes 2, 4 and 6 were classified.

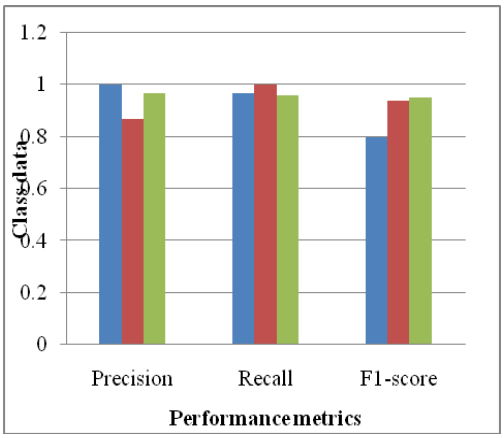

Fig 3: Performance metrics of class labels

Here, class 6 is classified appropriately recommends varying features that occurs in class 6. Approximation generally lies in class 2-4. Henceforth, classes 2 and 4 instances possess higher recall. Class 5-2 are same, data points come under class 2, and therefore precision is lesser as in Table IV.

Accuracy – is distinct as appropriately classified samples to total amount of samples as in Eq. (4):

$$Accuracy = \frac{samples\ appropriately\ classified\ with\ test\ data}{Total\ amount\ of\ test\ data\ samples} \quad (4)$$

Table V: Accuracy computation

S.No	Attack	Accuracy (%)
1	DoS	98.67
2	Probe	98.57
3	R2L	97.98
4	U2R	97.89

Table VI: Random tree performance (No of trees =100)

Class	Precision	Recall	F1-score
0	0	0	0
1	0	0	0
2	3	80	6
3	1	0	0
4	15	61	24
5	5	0	0
6	100	97	97
7	0	0	0
8	0	0	0
9	0	0	0
Total	88	86	87

This work attempts to integrate Random forest classifier with Weighted Mutual correlation Information to extract appropriate feature information to enhance performance as in Table V, it provides an impact over selection of features for classifier for detecting intrusion in IoT. The overall feature selection accuracy is higher as in fig 4. Diverse bypassing approaches are evaluated but those approaches do not provide any impact over increasing accuracy of feature selection.

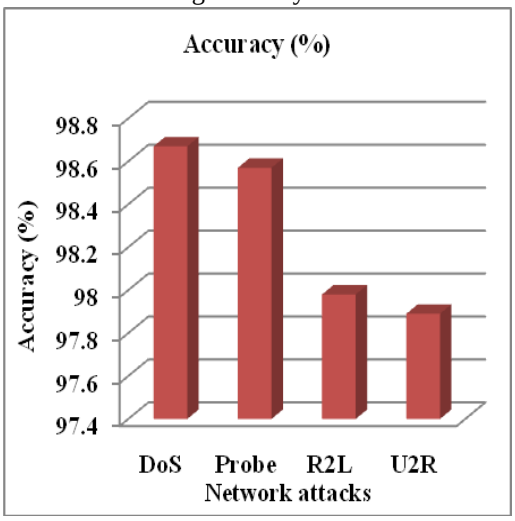

Fig 4: Accuracy computation

V. Conclusion

Here, IoT is considered as a distributed and a heterogeneous network to recognize the intrusion that degrades IoT device performance. There is a complexity encountered in deployment of IoT devices effectually due to security concern. In order to overcome this crisis, an improved Weighted Random Forest Information extraction (IW-RFI) approach is anticipated by estimating the weights and correlation characteristics of feature information with diverse IoT resource platforms. Accuracy enhancement and intrusion detection is extremely crucial in IoT application. To validate recital of anticipated model, this model is pre-processed with Kaggle dataset. The proposed model shows better trade off in contrast to conventional approaches and provides higher accuracy in selecting features for detection. Feature selection and extraction based on deep learning is will be done in future. Now, this investigation is based on validating the proposed model in IoT environment. As a future direction, IoT traffic based features are analyzed to offer superior accuracy. Pre-processing using detection algorithm will be offered in IoT. This work is extended in three different ways: 1) Feature dimensionality reduction for anomaly recognition 2) Feature extraction 3) recognizing anomalies as intrusions.

REFERENCES

1. Zheng, S.Q., Han, Y.J., Zhang, Q.: Architecture and application of IOT. Softw. Ind. Eng. 6(6), 27–31, 2010.
2. Mulligan, G.: The Internet of things: here now and coming soon.Internet Comput. 1, 36–37, 2010.
3. B. Arrington, L. Barnett, R. Rufus, and A. Esterline, "Behavioral modeling intrusion detection system (BMIDS) using internet of things (IoT) behavior-based anomaly detection via immunityinspired algorithms," in Proceedings of the 25th International Conference on Computer Communication andNetworks (ICCCN '16), pp. 1–6,Waikoloa, Hawaii, USA, August 2016.
4. Gendreau and M. Moorman, "Survey of intrusion detection systems towards an end to end secure internet of things," in Proceedings of the IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud '16), pp. 84–90, Vienna, Austria, August 2016.
5. Rayes and S. Samer, Internet ofThings—FromHype to Reality, Springer International Publishing, Cham, Switzerland, 2017.
6. H. Sedjelmaci, S. M. Senouci, and M. Al-Bahri, "A lightweight anomaly detection technique for low-resource IoT devices: a game-theoretic methodology," in Proceedings of the IEEE International Conference on Communications (ICC '16), pp. 1–6, IEEE, Kuala Lumpur, Malaysia, May 2016.
7. Pacheco, J., Hariri, S.: IoT security framework for smart cyber infrastructures. In: 1st International Workshops on Foundations and Applications of Self Systems, 2016.
8. Moustafa, N., Slay, J.: UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In: 2015 Military Communications and Information Systems Conference (MilCIS). IEEE, 2015.
9. Suo, H., Wan, J., Zou, C., Liu, J.: Security in the Internet of Things: a review. In: International Conference on Computer Science and Electronics Engineering (ICCSEE), vol. 3, 2012.
10. de Lima, I.V.M., Degaspari, J.A., Sobral, J.B.M.: Intrusion detection through artificial neural networks. In: Network Operations and Management Symposium NOMS 2008, pp. 867–870. IEEE, 7–11 April 2008.
11. Daniel Miessler. HP Study Reveals 70 Percent of Internet of Things Devices Vulnerable to Attack. http://h30499.www3.hp.com/t5/Fortify-Application-Security/HP-Study-Reveals-70-Percent//-of-Internet-of-Things-Devices/ba-p/6556284#.VH4faTHF9Zg, July 2014.
12. Murynets and R. Jover. Anomaly detection in cellular machine to-machine communications. In Communications (ICC), 2013 IEEE International Conference on, pages 2138–2143, June 2013.
13. GD Li, JP Hu, KW Xia, Intrusion detection using relevance vector machine based on cloud particle swarm optimization. Control Decision. 30(4), 698–702, 2015.
14. YH Yang, HZ Huang, QN Shen, et al., Research on intrusion detection based on incremental GHSOM. Chin. J. Comput.. 37(5), 1216–1224, 2014.
15. J Jiang, ZF Wang, TM Chen, et al., Adaptive AP clustering algorithm and its application on intrusion detection. J. Commun. 36(11), 118–126, 2015.
16. XN Wu, XJ Peng, YY Yang, et al., Two-level feature selection method based on SVM for intrusion detection. J. Commun. 36(4), 2015127-1–2015127-8, 2015.
17. Sherubha, "A detailed survey on security attacks on wireless sensor networks and its countermeasures", Int.Journal of soft computing, 2016.
18. B Wang, XW Nie, Multi-criteria mathematical programming based method on network intrusion detection. J. Comput. Res. Dev. 52(10), 2239–2246, 2015.

19. CH Tang, PC Liu, SS Tang, et al., Anomaly intrusion behavior detection based on fuzzy clustering and features selection. J. Comput. Res. Dev. 52(3), 718–728, 2015.
20. QA Wang, B Chen, Intrusion detection system using CVM algorithm with extensive kernel methods. J. Comput. Res. Dev. 49(5), 974–982, 2012.
21. B Liu, SX Xia, Y Zhou, et al., A sample-weighted possibilistic fuzzy clustering algorithm. Acta Electron. Sin. 40(2), 371–375, 2012.
22. Y Chen, An efficient feature selection algorithm toward building lightweight intrusion detection system. Chin. J. Comput. 30(8), 1398–1408, 2015.
23. M Ahmed, AN Mahmood, J Hu, A survey of network anomaly detection techniques. J. Netw. Comput. Appl. 60, 19–31, 2016.
24. P. Sherubha, N.Mohanasundaram, "An Adaptive FSCSO-RKPEM-based Feature Selection and Classification Techniques for Threat Identification in WSNs", Carribean Journal of sciences, Volume 53, ISSUE 2 (MAY - AUG), 2019.
25. P. Amudhavalli, sherubha and sasirekha, "Clone Attack Detection using Random Forest and Multi Objective Cuckoo Search Classification", International Conference on Communication and Signal Processing, April 4-6, 2019, India.
26. P. Sherubha, "An Efficient Intrusion Detection and Authentication Mechanism for Detecting Clone Attack in Wireless Sensor Networks", Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No. 5, 2019.
27. Rajendran T & Sridhar K P, "Epileptic seizure classification using feed forward neural network based on parametric features". International Journal of Pharmaceutical Research, 10(4): 189-196, 2018.
28. Hariraj V, Khairunizam W, Vikneswaran V, Ibrahim Z, Shahriman A B, Razlan Z M, Rajendran T, Sathiyasheelan R, "Fuzzy multi-layer SVM classification of breast cancer mammogram images", International Journal of Mechanical Engineering and Technology, 9(8): 1281-1299, 2018.
29. Muthu F, Aravinth T S & Rajendran T, "Design of CMOS 8-bit parallel adder energy efficient structure using SR-CPL logic style", Pakistan Journal of Biotechnology, 14(Special Issue II): 257-260, 2017.
30. Yuvaraj P, Rajendran T & Subramaniam K, "Design of 4-bit multiplexer using Sub-Threshold Adiabatic Logic (STAL)", Pakistan Journal of Biotechnology, 14(Special Issue II): 261-264, 2017.
31. Keerthivasan S, Mahendrababu G R & Rajendran T, "Design of low intricate 10-bit current steering digital to analog converter circuitry using full swing GDI", Pakistan Journal of Biotechnology, 14(Special Issue II): 204-208, 2017.
32. Vijayakumar P, Rajendran T & Mahendrababu G R, "Efficient implementation of decoder using modified soft decoding algorithm in Golay (24,12) code", Pakistan Journal of Biotechnology, 14(Special Issue II): 200-203, 2017.
33. Rajendran T & Sridhar K P, "Epileptic Seizure-Classification using Probabilistic Neural Network based on Parametric Features", Journal of International Pharmaceutical Research 46(1): 209-216, 2019.
34. Rajendran T, et al., "Recent Innovations in Soft Computing Applications", Current Signal Transduction Therapy, (Article in Press), 2019.
35. Emayavaramban G, et. al., "Identifying User Suitability in sEMG Based Hand Prosthesis Using Neural Networks", Current Signal Transduction Therapy, DOI: 10.2174/1574362413666180604100542 (Article in Press), 2019.
36. Rajendran T & Sridhar K P, "An Overview of EEG Seizure Detection Units and Identifying their Complexity- A Review", Current Signal Transduction Therapy, DOI: 10.2174/1574362413666181030103616 (Article in Press), 2019.

AUTHOR PROFILE

37. **MR.B.SURESH** has completed his B.Sc., M.Sc., M.Phil., and Pursing Ph.D., from Erode Arts and Science College (Autonomous), Erode. Affiliated to Bharathiar University. He has published more than 10 articles in National Journals, International Journals, and conference Proceedings. Presently he is serving as Assistant professor in the Department of ECS, VLB Janakiammal College of Arts and Science College (Autonomous), Coimbatore. His research interests are Microcontrollers, Embedded Systems and IoT

Dr.M.Venkatachalam has completed his B.Sc., M.Sc., M.Phil., and Ph.D., degrees from Bharathiar University. He has published more than 100 articles in National Journals, International Journals, and conference Proceedings. Presently he is serving as Associate professor and Head in the Department of Electronics, Erode Arts and Science College (Autonomous), Erode. He has served as Principal investigators for many funded projects and guided many scholars leading to the award of Ph.D. His research interests are Thin Film Technology, Nano technology, Embedded systems and IoT.

Dr.M.Saroja Has Completed Her B.Sc., M.Sc., M.Phil., And Ph.D., Degrees From Bharathiar University. She Has Published More Than 100 Articles In National Journals, International Journals, And Conference Proceedings. Presently She Is Serving As Associate Professor In The Department Of Electronics, Erode Arts And Science College (Autonomous), Erode. Her Research Interests Are Thin Film Technology, Nano Technology, Embedded Systems And Iot.

APPENDIX

Table I: Kaggle dataset test data

Duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent
0	tcp	private	REJ	0	0	0	0	0
0	tcp	private	REJ	0	0	0	0	0
2	tcp	ftp_data	SF	12983	0	0	0	0
0	icmp	eco_i	SF	20	0	0	0	0
1	tcp	telnet	RSTO	0	15	0	0	0
0	tcp	http	SF	267	14515	0	0	0
0	tcp	smtp	SF	1022	387	0	0	0
0	tcp	telnet	SF	129	174	0	0	0
0	tcp	http	SF	327	467	0	0	0
0	tcp	ftp	SF	26	157	0	0	0
0	tcp	telnet	SF	0	0	0	0	0
0	tcp	smtp	SF	616	330	0	0	0
0	tcp	private	REJ	0	0	0	0	0
0	tcp	telnet	S0	0	0	0	0	0
37	tcp	telnet	SF	773	364200	0	0	0
0	tcp	http	SF	350	3610	0	0	0
0	tcp	http	SF	213	659	0	0	0
0	tcp	http	SF	246	2090	0	0	0
0	udp	private	SF	45	44	0	0	0
0	tcp	private	REJ	0	0	0	0	0
0	tcp	ldap	REJ	0	0	0	0	0
0	tcp	pop_3	S0	0	0	0	0	0
0	tcp	http	SF	196	1823	0	0	0
0	tcp	http	SF	277	1816	0	0	0

IV: Feature selection for intrusion detection

Duration	protocol_type	service	flag	src_bytes	dst_bytes	count	srv_count	rerror_rate	srv_rerror_rate
0	tcp	private	REJ	0	0	229	10	1	1
0	tcp	private	REJ	0	0	136	1	1	1
2	tcp	ftp_data	SF	12983	0	1	1	0	0
0	icmp	eco_i	SF	20	0	1	65	0	0
1	tcp	telnet	RSTO	0	15	1	8	1	0.5
0	tcp	http	SF	267	14515	4	4	0	0
0	tcp	smtp	SF	1022	387	1	3	0	0
0	tcp	telnet	SF	129	174	1	1	0	0
0	tcp	http	SF	327	467	33	47	0	0
0	tcp	ftp	SF	26	157	1	1	0	0
0	tcp	telnet	SF	0	0	1	1	0	0
0	tcp	smtp	SF	616	330	1	2	0	0
0	tcp	private	REJ	0	0	111	2	1	1

0	tcp	telnet	S0	0	0	120	120	0	0
37	tcp	telnet	SF	773	364200	1	1	0	0
0	tcp	http	SF	350	3610	8	8	0	0
0	tcp	http	SF	213	659	24	24	0	0
0	tcp	http	SF	246	2090	16	16	0	0
0	udp	private	SF	45	44	505	505	0	0
0	tcp	private	REJ	0	0	204	18	1	1
0	tcp	ldap	REJ	0	0	118	19	1	1
0	tcp	pop_3	S0	0	0	1	1	0	0
0	tcp	http	SF	196	1823	17	17	0	0
0	tcp	http	SF	277	1816	17	18	0	0