Security Framework Connection Assistance for IoT Device Secure Data communication

Sarangam Kodati[1*], *Kumbala* Pradeep Reddy[2], *Thotakura* Veerananna[3], *S* Govinda Rao[4], *G* Anil Kumar[5]

[1]Associate Professor, Department of CSE, Teegala Krishna Reddy Engineering College, Hyderabad, Telangana.
[2]Associate Professor, Department of CSE, CMR Institute of Technology, Hyderabad, Telangana.
[3]Assistant Professor, Department of CSE, Sai Spurthi Institute of Technology,Sathupally, Telangana, India
[4]Professor CSE Department, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad.
[5]Assistant Professor CSE Department, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad.

Abstract: Today, Internet of Things (IoT) services has been increasing extensively because of their optimum device sizes and their developed network infrastructure that includes devices based on internet embedded with various sensors, actuators, communication, and storage components providing connection and data exchange. Presently number of industries use vast number of IoT devices, there are some challenges like reducing the risks and threats that exposure, accommodating the huge number of IoT devices in network and providing secure vulnerabilities have risen. Supervised learning has recently been gaining popularity to provide device classification. But this supervised learning became unrealistic as producing millions of new IoT devices each year, and insufficient training data. In this paper, security framework connection assistance for IoT device secured data communication is proposed. A multi-level security support architecture which combines clustering technique with deep neural networks for designing the resource oriented IoT devices with high security and these are enabling both the seen and unseen device classification. The datasets dimensions are reduced by considering the technique as auto encoder. Therefore in between accuracy and overhead classification good balancing is established. The comparative results are describes that proposed security system is better than remaining existing systems.

1. Introduction:

The Internet of Things (IoT) technology is widely spread around us because of its high level of security and provides best privacy to the system [1]. As much as the best, facilities of the IoT devices are used. If there is increment in connected devices in a network through internet then estimation is created by IoT as billions of users are crossed till 2020 [2]. Therefore security issues are raised by increasing the number of devices in IoT wireless and security devices. Number of devices is connected with internet through the Internet of Things (IoT). So there is chance of threats from unauthorized user on a large scale which can manipulate the data[3]. Therefore data confidentiality, privacy, authorization and authentication are IoT main security issues [16]. In the following mentioned layers attackers can enter into the communication as cloud layer, network layer and hardware layer[6]. The attacker was entered into the communication at hardware layer of IoT device and security parameters are retrieved or hacked which are stored in the IoT device [4]. By using these stolen security parameters virtual IoT device or duplicate one is recreated by the attacker. False data is uploaded to the server by this duplicate IoT device and users secure information is retrieved from network to which IoT device is connected [5].

Once the attacker starts to retrieve security parameters of the IoT device, there are some extra security issues are raised without being physical connection with device. ECC (*Elliptic Curve Cryptography)* and RSA (Rivest–Shamir–Adleman) based encryption keys are stolen by the side channel attacks based on electromagnetic which is exhibited by the researchers. From IoT devices AES encryption keys are stolen by using side channel attacks because all these IoT devices are connected to the internet so weak strength is acquired by IoT devices which causes to interferences in the form of attacks [17]. One example of such attacks is MIRAI malware in which most of the IoT devices outside of the network are attacked. Other internet services and websites are attacked by using network zombies which are from outside of the

Corresponding author: k.sarangam@gmail.com

network. The device performance is detected first in the proposed architecture and then securing operation is being processed to the IoT devices. The combination of clustering technique with supervised learning is proposed in this paper for enabling device seen and unseen type classification, hence the difference between secured IoT networks and unauthorized device accessing networks are detected[9]. Datasets dimensionality is reduced with proposed auto encoder technique which resulting the good accuracy and load balancing.

2. IoT Security Threats and Challenges

2.1 IOT security threats

An IoT technology non-standardization with weakness intensification is gives the IoT systems with great security [7]. Some generic threats brief discussion is described below.

2.1.1: Hardware Vulnerabilities:
The IoT products which are commercially developed are considering one main parameter as security while other devices which are functionality centric are not. So the addition of security features with devices is later. Hence, hardware vulnerability like open physical interfaces and boot process vulnerabilities remain in such devices, which can be exploited remotely [8].
2.1.2: Vulnerabilities of Social Engineering: IoT devices interactions with humans and socialization are maintain a great impact on user's life. Social engineering attacks are attracted by IoT users because of large amount of collected data. Smart TVs, Google Glasses, Fitbits and refrigerators are some smart devices which are also controlled by hackers [20].
2.1.3: Legislation Challenges: IoT data security is cannot guarantee by legislation so data misuse may results to damage of system then it can compensate. Till now there are not drafts for secure data policy and standardized legislation. Health Insurance Portability and Accountability Act (HIPPA) and General Data Protection Regulation (GDPR) are safety measures which are provided from different countries.
2.1.4: User Unawareness: Users are the conventional or traditional attack vectors for the network. Lacking of security awareness and training cause deficiency in security in phishing/spear-phishing or social engineering networks and in this end user as well as employees both are susceptible. Sensitive data transmission in public networks through mobile devices is also results the security degradation [10].

2.2 IoT security challenges

IoT is having different types of security issues or challenges. Three categories of challenges are divided as named as end applications, IoT data and communication related security [11]. The detail explanations related to these issues are mentioned below after layer and generic wise IoT threats discussion [18]. Confidentiality, integrity and availability are can be short formed as CIA. In any organization security of the information may follow the guidelines of CIA which are basic ones. So security of the system is defined by these three variables mostly.

2.2.1 Confidentiality: The information availability is limited by these set of rules. The sensitive data cannot handled by unwanted people and make it for selection of right owner of data for doing further actions with these set of measures. The IoT services trustworthiness such as societal, manufacturer and personal are greatly depends on data genuineness which has the output with its undeviating effect [13]. End nodes of IoT must be confidential and authentic for secure transmission of data among the IoT applications and services.

2.2.2 Integrity: trustworthiness and correct data is explained through integrity. Over the data complete life cycle trustworthiness, accuracy and information consistency are involved in this integrity parameter [19]. The data should be same during the transmission and make sure with different measures that this information is cannot be changed are break by any unauthorized participants.

2.2.3 Availability: The accessibility of data to authorized users is called as availability. This hardware is best practiced with strict maintenance. So, operating-system with proper working circumstances is provided which is free from software frays. Time to time up gradation of the system is also being done with the availability of data [15].

3. Security Framework Connection Assistance for IoT

The proposed architecture is shown in fig. 1 which is used for enabling the security operation for IoT devices without increasing the processing load. It is containing the IoT devices Gateways, clustering, a platform and applications (APP) and classification processing. End users are got the services from service providers by using IoT devices and APP. But service providers are connected with network infrastructures by the carrier with gateways and platform.

One of the extra network infrastructure advantages is enabling security-operation from the carrier standpoint which is called as "device assistance". The traffic in network can be captured and desired features are extracted by the data processing module when there is a connection of device with network. Each known device type uses the creation of one-vs-rest binary classifier and white list method is used in the Train module. Predict modules and labels used in the input as processed data when classifier ready for acceptance [12]. Classifier models are directly used by the Prediction module from Train for feature vector labeling and device type prediction. The labeling process of feature vector is observed by the discriminator [22]. If labeled then action module is receives the feature vector. Where the improvement

strategy is applied that module is called as action module. Several mitigation strategies are given to different categories and phases by the action module. If not labeled then clustering module is receives the feature vector and then continuously fed to the active module [14]. Clustering module detail explanation is described below.

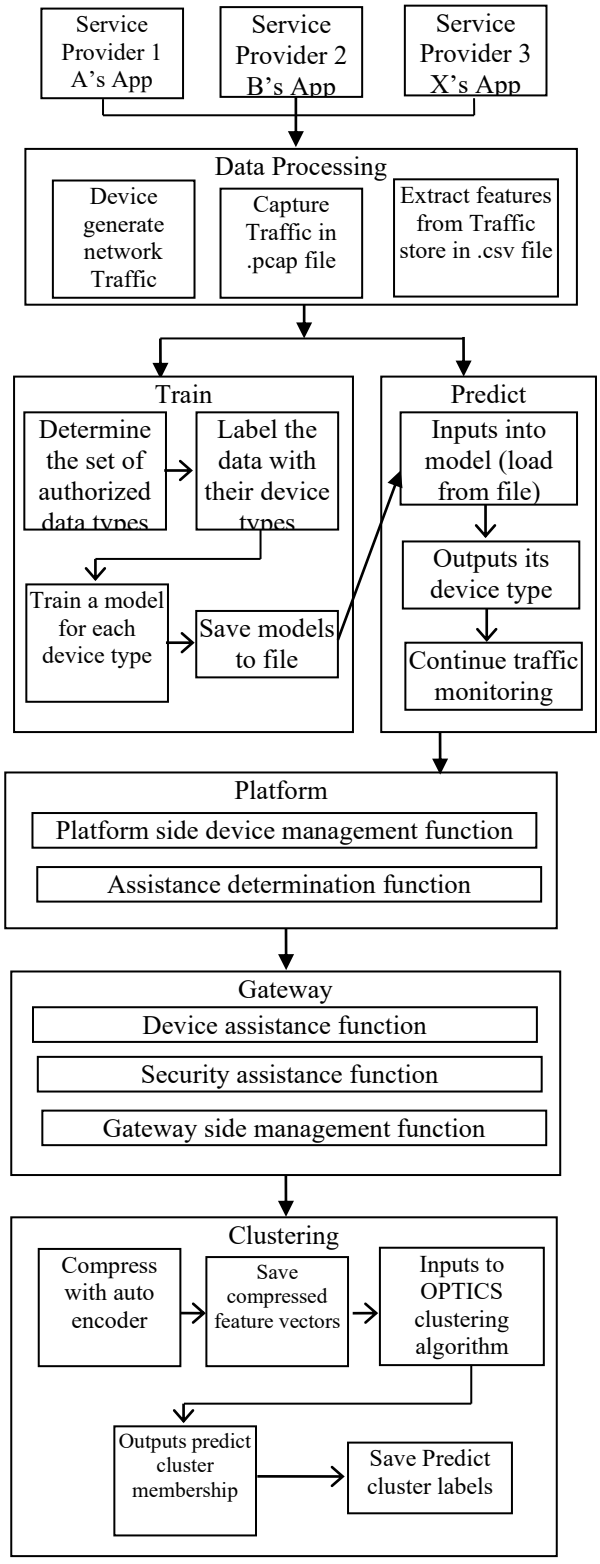

Fig. 1: Security Framework for IoT Devices

The proposed architecture is uses the referred assistant technology as a key which is not only deals with "device assistance" but also concentration on processing load on the gateway. Device management function is arranged on both sides of gateway sides and platform sides for achieving high response speed. All devices are can be managed by management function from platform side and connected devices are can be managed by the management function from gateway side. Assistance determination function is also arranged at platform side for simultaneously assisting the policy by enabling the service provider. Security assistance function is maintained at gateway side for doing the mechanism of session resumption. Estimation function for device performance is kept at platform side this is because it requires interaction with devices like sending packets to devices. Therefore less amount of traffic is achieved at wide area network.

3.1 Gateway

The requirement of assistance is determined by the first step of gateway and provides the assists if required. One request for gateway's processing load l_d t_d is for high performance devices, where, l_d processing time for device performance determination is represented with t_d and processing load for device performance determination is represented with l_d. One request for gateway's processing load is ($l_d\, t_d + l_a t_a$) for constrained devices. l_c is one request for processing load as $l_a t_a$, here for device assisting involved processing load is represented with l_a and for device assisting involved processing time is represented with t_a. Total number of devices are treated as N which access the gateway and assumed as n($0 \le n \le N$) high-performance devices. Therefore, the gateway's processing load of one request l_p is $l_d\, t_d$ $l_d t_d \frac{n}{N} + l_d t_d + l_a t_a) \frac{N-n}{N}$.

3.2 OPTICS

OPTICS can be abbreviated as Ordering Points To Identify the Clustering Structure. It is a data space density based clustering algorithm of unsupervised data. Density-based spatial clustering of applications with noise (DBSCAN) gives the fundamental idea for OPTICS. Outlier, border and core are three point classifications. Very less amount of points are existed in core point within its ϵ-neighborhood radius, including itself. Border point is core point neighborhood and within its ϵ-neighborhood radius, 'minimum number of points' less points are having by border point. Any cluster cannot reach the outlier point. $minPts$ and ϵ are two parameters used in cluster definition for DBSCAN. The neighborhood maximum radius is represented with ϵ and required number of minimum points for defining the cluster is represented with $minPts$ in the ϵ-neighborhood. DBSCAN uses the two parameters while OPTICS uses only one parameter as $minPts$. Therefore OPTICS

is less sensitive for parameters.

According to the distance between data points, a database is created in ascending density order from the idea of OPTICS. Density-based clustering structure is represented by the distance between the points. reachability distance and core distance are two distances used in storing the clustering order. At point o core distance is CD (o) and defined as:

$$CD(o) = \begin{cases} UNDEFINED\ if |\{x|d(x,0) \le \varepsilon_{max}\}| < minPts \\ minPts - dis(0) \quad \text{otherwise} \end{cases}$$

Where, the distance to the given $minPts$ nearest neighbor is denoted with $ts - dis$. The core distance is undefined when number of other points are sufficiently isolated by the point o within radius ε_{max} is less than $minPts$. If $minPts$ as the core distance otherwise. At point p, reachability distance is $RD(o,p)$ and defined as:

$$RD(o,p) = \begin{cases} UNDEFINED\ if |N_\epsilon(p)| < minPts \\ max\{d(o,p), CD(o)\} \quad \text{otherwise} \end{cases}$$

Where, p neighborhood is $|N_\epsilon(p)|$, in above equation all distances are referred as Minkowski distance. The first processed points are having the smallest reachable distances i.e. high density. The data points at OPTICS output are sorted according to their reachability distance and processed order.

3.3 Auto Encoder (AE)

The input x is reconstructed from output by the training of auto encoder (AE) which is a symmetrical artificial neural network. Two parts are existed in AE: one is encoder in which the features (bottlenecks) are mapped with input x and another one is decoder which is reconstructs the input from features. The reconstructed $\hat{x}$ is having the features same as the input x and this is possible by neural network parameters. Given a set of p input data vectors, $x = \{x_1, x_1, \cdots, x_p\}$, an input vector x_i feed-forwards to a bottleneck vector $\mathcal{L}_i = f_0(x_i) = \sigma(Wx_i + b)$, where activation function is denoted with σ , bias vector is denoted by b and weight matrix denoted by W. Weight and bias form the parameter set θ = {W, b}. By using $\hat{x}_i = g\theta(\mathcal{L}_i) = \sigma(W'^{\mathcal{L}_i} + b')$, the vector $\hat{x}_i$ is reconstructed from bottlenecks with the same dimensions of input vector. The decoder weights are considered from the transpose matrix of weights of encoder because of its symmetrical structure, i.e., $W' = W^T$. Now AE is back-propagated for parameters optimization and loss function minimization $J(\theta; x, \hat{x}) = \frac{1}{p}\sum_{i=1}^{p}\|x_i - \hat{x}_i\|^2$. The input space size should have more dimensionality than bottleneck space size. The input is directly copied as output when the input space is smaller than the hidden layer. Sparse bottleneck space is an alternative way instead of bottleneck neurons reduction. More hidden units are included in Sparse AE than inputs but at once hidden units are in small number. Regularizer is implements the sparsity constraints. After considering sparse space, loss function becomes:

$$J_{sparse}(\theta; x, \hat{x}) = J(\theta; x, \hat{x}) + \Omega(\mathcal{L}).$$

3.4 Random Forest

Huge collection of decision trees which are decorrelated are used the random forest algorithm for classification. The structure of decision tree is same as flowchart; a decision attribute is represented as internal node. As two branches every point is divided and decision result is represented with every branch, decision result class label is denoted with each leaf node. Branch split can exists in many positions in general. Gini Impurity is a measure for split quality, and defined as:

$$Gini(t) = 1 - p_{positive}(t)^2 - p_{negative}(t)^2$$

Where, positive probability is $p_{positive}(t)$, and negative probability is $p_{negative}(t)$ for the test. Separation effect is better when Gini impurity is small. Random forest takes the input as training data matrix S, feature number is denoted by n, data samples with p and for each data point class label with C . Matrix S is defined as:

$$S = \begin{pmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,n} & C_1 \\ f_{2,1} & f_{2,2} & \cdots & f_{2,n} & C_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ f_{p,1} & f_{p,2} & \cdots & f_{p,n} & C_p \end{pmatrix}$$

The row of the matrix S is shuffled for creating M subset matrices randomly with same size of input matrix S. Therefore these obtained subsets are named as bootstrapped datasets. Then create an each subset decision tree now. Random forest accuracy is measured with the difference between original set and each subset decision trees. This accuracy is used in parameters fine tuning.

4. RESULTS

In experiment results, proposed security framework for Iot performance evaluation is divided into two parts: OPTICS unsupervised device type identification and Random forest supervised dimension reduction for anomaly detection.

4.1 Performance of Device Type Identification

Suricata named open-source IDS and IPS tool is used for capturing the network traffic in represented embedded sensor. From 12 production lines network packets are collected which are belongs to target factory. Controller's events, robotic arm events and computer events are three classes which are from labeling the data manually with 21,447 total records

and it is used as device identification first dataset. Several types of traffic are represented with events in Suricata and explained with different fields or other protocols. 110-dimensional features were used. According to our experimental results with ordering points used in identification of clustering structure, the test accuracy is 98.6%. Therefore device identification uses the OPTICS for obtaining good efficiency. Then device dimensions are reduced by using feature selection methods which are used for device identification model performance is improved as 98.6%. Top 10 important features are obtained after applying AE for feature selection and device

identification confusion matrix is represented in below Table 1.

Table 1: The Confusion Matrix of Device Identification Results

Actual Vs. Predicted	Controller	Arm	PC
Controller	335	1	4
Arm	7	43	21
PC	6	16	1708

As shown in Figure 2, we can see an improved accuracy of 97.8% with the feature selection method, that is improved by 4.8% compared to without feature selection case and achieved more enhanced accuracy rate of 98.6% device type identification with the proposed OPTICS feature selection method. Therefore, feature selection methods effectiveness is used in classification performance improvement.

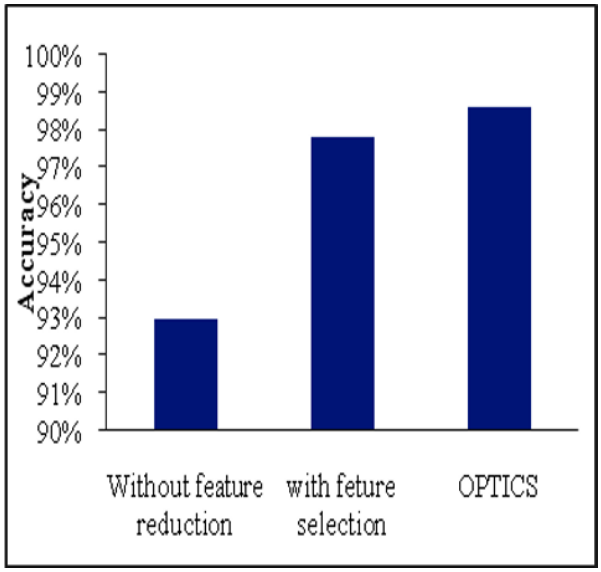

Fig. 2: Comparison Results of Device Type Detection Accuracy

4.2 Performance of Anomaly Detection method

There is no availability of real IoT attacking data so malicious network patterns are simulated first which are having different behaviors than normal device behavior. Normal packets and attacking packets are collected in two sites which are from devices. Anomaly detection using two datasets statistics are shown in below Table 2.

Table 2: The Confusion Matrix of Device Identification Results

Type	Packets	
	Site A	Site B
Controller	22389	21286
Arm	101	86
PC	22490	21372

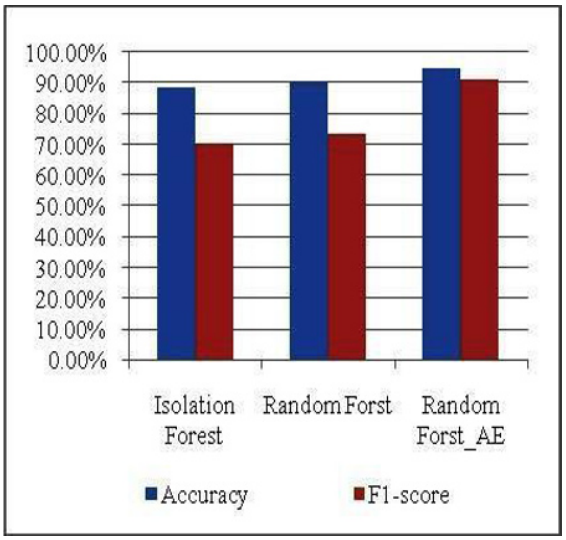

Fig. 3: PERFORMANCE COMPARISON OF ANOMALY DETECTION

The anomaly detection comparative performance is represented in Fig. 3 and Fig. 4 in terms of accuracy, F1-score and precision recall respectively. Best performance is observed when auto encoders are used for the reference dataset with achieved F1 score as 91.28% and 95.02% of accuracy. Learning normal behaviors effectiveness is observed clearly from above statements by using auto encoders. Therefore anomaly detection at any experiments are uses the auto encoders.

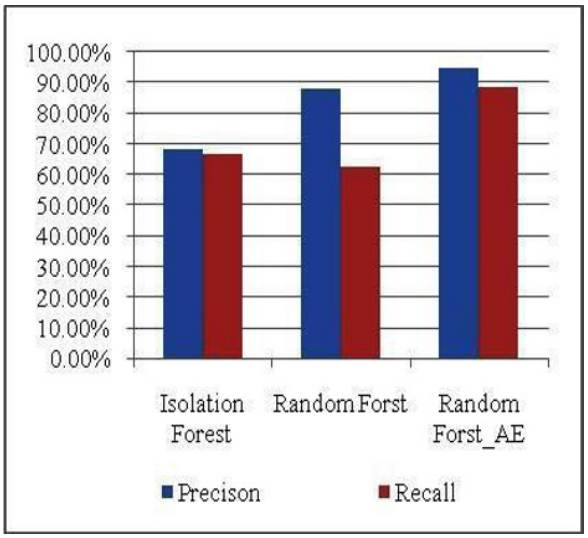

Fig. 4: Performance Comparison of Anomaly Detection

5. CONCLUSION

Security framework connection assistance for IOT device secured data communication was analyzed in this paper. The device performance is detected by the proposed architecture in first step and then it helps in controlling the IoT devices when there is a requirement. Unauthorized device accessing is eliminated by providing a secure IoT networks and any irregularities are detected by using network traffic in proposed hybrid learning framework. A multi-level security support architecture which combines clustering technique with deep neural networks for designing the resource oriented IoT devices with high security and these are enabling both the seen and unseen device classification. The datasets dimensions are reduced by considering the technique as auto encoder. In between accuracy and overhead classification good balancing is achieved.

References

1. Vikas Hassija, Vinay Chamola, Vikas Saxena, Divyansh Jain, Pranav Goyal, Biplab Sikdar, ":A Survey on IoT Security: Application Areas, Security Threats, and Solution Architectures", IEEE Access, Volume: **7**, (2020)
2. Fatima Hussain, Rasheed Hussain, Syed Ali Hassan, Ekram Hossain, "Machine Learning in IoT Security: Current Solutions and Future Challenges", IEEE Communications Surveys & Tutorials, Volume: **22**, Issue: 3, thirdquarter (2020)
3. Swaraja K , "Medical image region based watermarking for secured telemedicine", Multimedia Tools and Applications, 77 (21), (2018) ,pp. 28249-28280.
4. Shachar Siboni, Vinay Sachidananda, Yair Meidan, Michael Bohadana; Yael Mathov, "Security Testbed for Internet-of-Things Devices", IEEE Transactions on Reliability, Volume: **68**, Issue: 1, (March 2019)
5. S. Balaji, K. Nathani and R. Santhakumar, "IoT technology applications and challenges: A contemporary survey", *Wireless Pers. Commun.*, vol. **108**, pp. 363-388, (Apr. 2019).
6. Kumar, P., Singhal, A., Mehta, S., Mittal, A,"Real-time moving object detection algorithm on high-resolution videos using GPUs", Journal of Real-Time Image Processing, 11 (1), (2016) , pp. 93-109.
7. C. Kolias, G. Kambourakis, A. Stavrou and J. Voas, "DDoS in the IoT: Mirai and other botnets", Computer, vol. **50**, no. 7, pp. 80-84, (2017).
8. M. Hossain, R. Hasan and A. Skjellum, "Securing the Internet of Things: A meta-study of challenges approaches and open problems", *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. Workshop (ICDCSW)*, pp. 220-225, 2017
9. Raghunadha Reddy, T., Vishnu Vardhan, B., Vijayapal Reddy, P,"A survey on Authorship Profiling techniques", International Journal of Applied Engineering Research, 11 (5), (2016),pp. 3092-3102.
10. A. Mosenia and N. K. Jha, "A comprehensive study of security of Internet-of-Things", IEEE Trans. Emerg. Topics Comput., vol. 5, no. 4, pp. 586-602, (Dec. 2016).
11. S. Ray, Y. Jin and A. Raychowdhury, "The changing computing paradigm with Internet of Things: A tutorial introduction", *IEEE Design Test*, vol. **33**, no. 2, pp. 76-96, (Apr. 2016).
12. Mahalle, G., Salunke, O, Kotkunde, N., Gupta, A.K., Singh, S.K, t"Neural network modeling for anisotropic mechanical properties and work hardening behavior of Inconel 718 alloy at elevated temperatures", Journal of Materials Research and Technology, 8 (2), pp. 2130-2140.(2019)
13. K.-H. Yeh, "A secure IoT-based healthcare system with body sensor networks", IEEE Access, vol. **4**, pp. 10288-10299, (Dec. 2016).
14. Kumar, S.K., Reddy, P.D.K., Ramesh, G., Maddumala, V.R. "Image transformation technique using steganography methods using LWT technique", Traitement du Signal (2019)
15. X. Yao, X. Han, X. Du and X. Zhou, "A lightweight multicast authentication mechanism for small scale IoT applications", IEEE Sens. *J.*, vol. **13,** no. 10, pp. 3693-3701, (Oct. 2013).
16. Vikas Hassija, Vinay Chamola, Vikas Saxena, Divyansh Jain, Pranav Goyal, Biplab Sikdar, "A Survey on IoT Security: Application Areas, Security Threats, and Solution Architectures", IEEE Access, Volume: **7**, (2019)
17. Y. Liu, Y. Kuang, Y. Xiao and G. Xu, "SDN-based data transfer security for Internet of Things", IEEE Internet Things J., vol. **5**, no. 1, pp. 257-268, (Feb. 2018).
18. J. Wan et al., "Software-defined industrial Internet of Things in the context of industry 4.0", *IEEE Sensors J.*, vol. 16, no. 20, pp. 7373-7380, Oct. 2016.
19. O. Arias, J. Wurm, K. Hoang and Y. Jin, "Privacy and security in Internet of Things and wearable devices", *IEEE* Trans. Multi-Scale Comput. Syst., vol. **1**, no. 2, pp. 99-109, (Apr.–Jun. 2015).
20. P. C. S. Reddy, S. G. Rao, G. R. Sakthidharan and P. V. Rao, "Age Grouping with Central Local Binary Pattern based Structure Co-occurrence Features," *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, (2018), pp. 104-109, doi: 10.1109/ICSSIT.2018.8748473.
21. P.Chandra Sekhar Reddy , "Gender Classification using Central Fibonacci Weighted Neighborhood Pattern Flooding Binary Matrix (CFWNP_FBM) Shape Primitive Features", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-**8**, Issue-6, (August 2019).PP.5238-5244.
22. M. Z. Hasan, H. Al-Rizzo and F. Al-Turjman, "A survey on multipath routing protocols for QoS assurances in real-time wireless multimedia sensor networks", *IEEE* Commun. Surveys Tuts., vol. **19**, no. 3, pp. 1424-1456, 3rd Quart. (2017).