International Journal of
Engineering in Computer Science

E-ISSN: 2663-3590
P-ISSN: 2663-3582
www.computersciencejournals.com/ijecs
IJECS 2024; 6(2): 125-130
Received: 25-06-2024
Accepted: 13-08-2024

Machine learning approaches to detect dos and their effect on WSNS lifetime

Putta Srivani, G Bhoomika, J Srivigna and M Parimala Sai

DOI: https://doi.org/10.33545/26633582.2024.v6.i2b.135

Putta Srivani
Assistant Professor, Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad, Telangana, India

G Bhoomika
Student, Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad, Telangana, India

J Srivigna
Student, Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad, Telangana, India

M Parimala Sai
Student, Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad, Telangana, India

Corresponding Author:
Putta Srivani
Assistant Professor, Malla Reddy Engineering College for Women (Autonomous Institution), Hyderabad, Telangana, India

Abstract

Wireless sensor networks (WSNs) still face the dual problems of energy consumption and data security. Hence, one of the security responsibilities of WSN networks is to prevent them against Distributed Denial of Service (DDoS) and Denial of Service (DoS). Machine learning-based systems are now the only practical option for defending against these kinds of attacks, given traditional packet deep scan methods depend on open field inspection of transport layer security packets and open field encryption is becoming more popular. This research adds to the existing literature by evaluating the impact of machine learning algorithms on the lifetime of WSN networks and the traffic that flows through their nodes. We used a WSN-dataset of varying sizes to assess the performance metrics of various machine learning classification categories, including K-Nearest Neighbour (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Gboost, Decision Tree (DT), Naïve Bayes, LSTM, and Multi-Layer Perceptron (MLP). Results demonstrated that logical and statistical classification categories outperformed others on numerical statistical datasets, and that, across all performance criteria, the Gboost algorithm outperformed the competition. When doing these validations, the following performance measures were considered: accuracy, F1-score, FPR, FNR, and training execution time. In addition, the accuracy, F1-score, FPR, and FNR scores for the Gboost algorithm were 99.6%, 98.8%, 0.4%, and 0.13%, respectively, according to the test findings. In terms of training execution time, it averaged out all datasets at 1.41 seconds. Also, this article proved that numerical statistical data works best with datasets between 3,000 and 6,000 records in size, with at least 50% overlap between each category and all others. In addition, this article examined how Gboost affected the lifespan of WSNs, which was shown to be 32% shorter than other Gboost-free cases.

Keywords: Wireless sensor networks (WSNs), distributed denial of service (DDoS), machine learning

Introduction

Building the Internet of Things (IoT) relies heavily on wireless network technologies. For the simple reason that wireless networks are crucial for the exchange of interactive data between gadgets and people or between gadgets themselves [1]. There is no human interaction required for the data sharing capabilities of these devices, which are a component of automation and control systems, embedded systems, and Wireless Sensor Networks (WSN), among others. Typically, there are three levels to any application that makes use of these devices: perception, network, and application [2]. So that they can function for as long as feasible, particularly with restricted battery life systems, low-power devices execute the perception layer, whereas high-power devices often execute the application and network layers. With the ability to sense, record, calculate, and track, the perception level is comprised of many WSN nodes that interact with one another across different radio frequencies [3]. Because of their small size, limited memory space, slow processing speed, and short battery life, these WSN nodes are like miniature computers. Furthermore, Zigbee and 6LoWPAN are two protocols that are widely used in WSN networks between the physical and Media Access Control (MAC) layers of IEEE 802.15.4 [4]. Because WSN nodes aren't meant to be constantly watched, they may run in a variety of untrusted environments. Because of this, WSN nodes are susceptible to threats to their security, particularly those involving sensitive information [5]. In addition, supplying a charger for WSN nodes in these situations might be challenging when it comes to their limits, which include CPU power and energy [6]. Due to their reliance on public wireless networks and other WSN node disadvantages, WSN architecture has several issues. One of them is the perception layer's availability, privacy, and security. Data connection protection between WSN nodes from spoofing and eavesdropping via unauthorized WSN node modification or change is the primary concern in the realm of security [7, 8]. Someone may deactivate it by interfering with data packet transmission in WSN node availability via sinkhole, wormhole, Sybil, hello flood, or Denial of Service (DoS) assaults [3]. Threats of denial of service (DoS) may cause WSN nodes to drop data packets and use all their resources. As a result, this study will focus on reducing the impact of Distributed Denial of Service (DDoS) and Denial of Service (DoS) attacks in WSN networks while minimizing power consumption and accurately describing both types of assaults [9]. Additionally, the goal of a denial-of-service (DoS) or distributed denial-of-service (DDoS) attack is to exhaust the resources of a WSN node by flooding it with either legitimate or superfluous messages. Every tier of a WSN model system is vulnerable to these kinds of attacks, which cause WSN nodes to stop providing services to valid WSN nodes [10]. Since WSN networks are vulnerable to DDoS attacks, intrusion detection techniques are the greatest line of security [11]. There are two main types of intrusion detection systems: signature-based and anomaly-based. Regular network connection monitoring and comparison of continuing WSN network activities with current normal behavior traffics are necessary for the anomaly pattern approach [9, 11]. Consequently, in order to make active and passive DoS detection more effective, a number of strategies have been used. An approach to DDoS attack prediction and classification is supervised machine learning techniques. Typical techniques for this task include Logistic Regression (LR), decision Tree (DT), artificial neural networks, Support Vector Machine (SVM), deep learning, and K-Nearest Neighbor (KNN) [12]. While the authors in [1, 2, 13–17] did successfully improve detection accuracy, mean squared error, and sensitivity by utilizing various deep learning mechanisms, they failed to address how their proposal would impact WSN networks in terms of node power consumption and network lifetime. Datasets including WSN nodes were also not included in other articles; instead, normal networks' traffic was used [18]. In [19], the authors provide alternative methods for DDoS detection; their method involves separating WSN nodes into clusters and then connecting all of the nodes in each cluster using an authentication message to eliminate DoS assaults. Further, research in [20, 21] demonstrated that deep learning algorithms require enormous amounts of training data to produce accurate results in the classification process, making machine learning methods (LR, SVM, and DT) more suited for deployment in the real world. For that reason, it is pointless to apply these capabilities to training activities via the WSN network node [14]. With the use of machine learning and a clustering strategy for WSN nodes, an alternative method for detecting DDoS attacks has been suggested in [22, 23]. Data gathered from WSN nodes was also fed into a support vector machine (SVM) in [24] that takes into account spatial-temporal and attribute correlations. While this simulation and dataset have been considered in previous publications, no one has yet examined how the aforementioned techniques impact WSN networks. More importantly, anti-DDoS procedures have to be quick and easy since WSN nodes have limited resources. So, this study primarily adds to the existing literature by introducing novel WSNs settings, analyzing the performance of machine and deep learning algorithms on the WSN network dataset, and determining how these methods impact the longevity of WSN networks. Following is a synopsis of the work's main contributions: (1) In order to identify denial-of-service attacks and evaluate the influence of this detection on power consumption of WSN nodes, we presented a novel WSN network environment that integrated WSN nodes clustering technology, authentication, key management [25] and WSN-Dataset [23]. (2) Determining how different machine learning categorization methods fare when faced with varying dataset sizes. Several subsets were created from the original dataset, with the record count dropping with each subset. Thirdly, determine how DNS anomaly detection performance affects the longevity of WSN networks. Our document is structured as follows for the rest of it. In Section 2, we take a look at the relevant literature on machine learning, denial-of-service attacks, and WSN networks. Chapter 3 delves into the approach, construction of the environment, cluster management, machine learning test, and decision-making processes. Topics covered in Section 4 include gathering and organizing data. Machine learning methods and the lifespan of WSN networks are examined in Section 5 along with their application and assessment for complexity analysis. Section 6 wraps up the report with a conclusion and suggestions for further research.

Related Work

“A Survey of internet of things (IoT) in education: Opportunities and challenges”

The most difficult platform that will soon indicate the association of physical items is the Internet of Things (IoT). In order to study and summarize the Internet of Things (IoT) and its uses in many fields, a large number of review studies have been carried out. However, there hasn't been a thorough evaluation of the Internet of Things (IoT) in the classroom. Thus, this research primarily aims to showcase the current state of using IoT applications in education and to provide a range of possibilities and obstacles for future experiments. Specifically, this research study compiles the potential benefits and drawbacks of using the Internet of Things (IoT) in the following areas: medical, vocational, Green, and wearable technology education and training. It is concluded that developing nations are still in the early phases of adopting the Internet of Things and its applications, and that more study is strongly urged.

“A new digital watermarking method for data integrity protection in the perception layer of IoT”

Worldwide, governments and academic institutions have enthusiastically supported the Internet of Things (IoT) from its inception, and impressive results have been achieved. The perception layer of the Internet of Things (IoT) is crucial because it connects the IoT to the physical world; yet, security concerns have created a roadblock to the expansion of the IoT. In the perception layer, a system of wirelessly communicating sensor nodes with limited resources forms a self-organizing network. Consequently, the perception layer cannot be encrypted using the expensive approach. This work proposes a new, lightweight data integrity protection strategy that uses fragile watermarks to address the conflict between the perception layer's limited resources and its security. Our position random watermark (PRW) approach uses the temporal dynamics of sensing data to determine the embedding location, therefore improving security. Before being embedded to the dynamically determined place, the digital watermark is formed using the one-way hash algorithm SHA-1. This manner, we may remedy the security issues caused by the embedded location being fixed and ensure that the data remains undisturbed. Data integrity may be assured at a cheap cost using the suggested approach, according to security analysis and simulation findings.

"A novel block encryption algorithm based on chaotic s-box for wireless sensor network"

We suggest a block encryption technique called S-box that relies on chaotic substitution to guarantee the fundamental security of WSNs. In this research, we constructed a novel S-box by combining the linear congruence generator, the Baker map, the sinusoidal chaotic map, and the compound chaotic map. Also covered in this study are the limitations of WSN in terms of processing power and communication capabilities. The S-box is the foundation upon which the technique for creating round subkeys and the F function is built. The suggested encryption technique has minimal resource consumption and strong security, making it ideal for WSN, according to the rigorous performance and security testing.

"Congestion control in wireless sensor and 6LoWPAN networks: Toward the internet of things"

The next major obstacle for researchers will be the Internet of Things (IoT), a critical component of which is the IPv6 over low power wireless personal area network (6LoWPAN) protocol stack. To address the difficulties in connecting sensor nodes with limited memory, processing power, and availability to the Internet, new IP-based protocols for 6LoWPAN networks have been developed by the IETF ROLL and 6LoWPAN working groups. Network performance and quality of service metrics like throughput, latency, energy consumption, dependability, and packet delivery are adversely affected by congestion in 6LoWPAN networks caused by excessive network traffic. This article provides a high-level summary of the 6LoWPAN protocol stack, including all of its associated protocols and standards. We further categorize these algorithms into three types: hybrid, traffic, and resource control, according to the congestion control technique they use, and we evaluate and contrast several well-known congestion control methods in WSNs. We provide a comprehensive analysis of all current methods for 6LoWPAN congestion management. The study explains why WSN congestion management approaches work for 6LoWPAN networks and compares and contrasts the methods used to manage congestion in WSNs and 6LoWPAN networks. Last but not least, this study suggests some ways forward for further research into developing an innovative congestion control protocol that meets the needs of IoT applications.

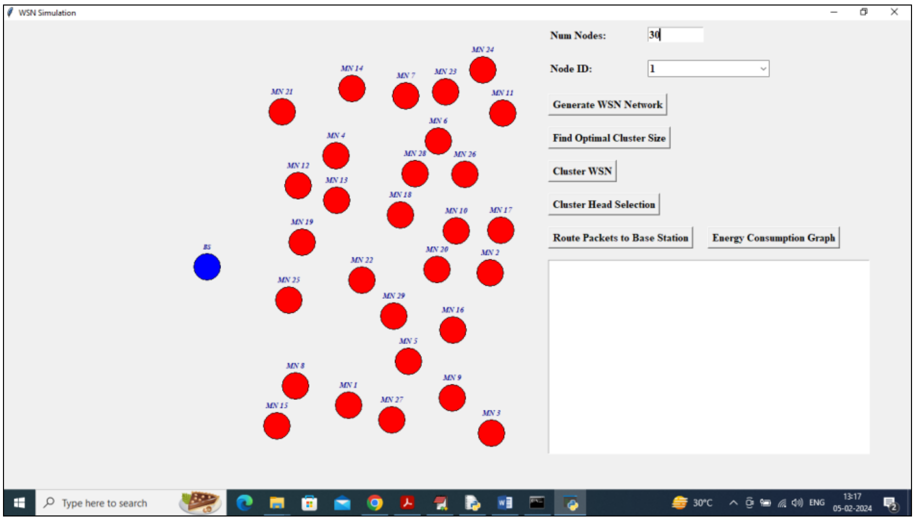
"Energy efficiency of encryption schemes applied to wireless sensor networks"

Secure communication in WSNs with respect to energy efficiency is the primary emphasis of this study. The cryptographic implementation algorithms and cipher modes of operation, including the formation of initialization vectors (IVs) and other details, are examined in our study on the link layer security of WSNs. Taking into account the algorithmic features and the impact of channel quality on cipher synchronization, we compare the computational energy efficiency of several symmetric key ciphers. Compared to stream ciphers, block ciphers use less computing resources during data encryption and transmission over a noisy channel. Various parameters impacting the communication energy cost of link layer cryptographic systems are examined in deeper detail. These elements include payload size, cipher mode of operation, IV distribution, and communication channel quality. An energy analysis model of secure data transmission at the link layer is developed in order to conduct a thorough performance comparison of various cryptographic algorithms. We build this model with a number of considerations for the communication cost and processing cost in mind, and we use simulation results to confirm that it is adequate. To avoid sacrificing security in WSNs for the sake of energy savings, we suggest encrypting data for WSN applications using a block cipher rather than a stream cipher, and then implementing a cipher feedback system for the cipher operation.

Methodology

1. **Upload WSN DOD Dataset:** Using this Module We can Upload Attack dataset to the application.
2. **Preprocess & Split Dataset:** Using this module we can apply processing techniques like shuffling, normalization and splitting into train and test.
3. **Run KNN, LR & SVM Algorithms:** SVM, KNN and Logistic regression training completed and can see accuracy and other metrics along with execution time as output.
4. **Run G Boost & Naive Bayes Algorithms:** GBOOST and Naïve Bayes training completed and can see accuracy and other metrics along with execution time as output.
5. **Run MLP & LSTM Algorithms:** MLP and LSTM training completed and can see accuracy and other metrics along with execution time as output.
6. **Comparison Graph:** In the graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars and in all algorithms GBOOST got high accuracy.
7. **Detect Attack from Test Data:** In this module selecting and uploading 'Test Data' file.

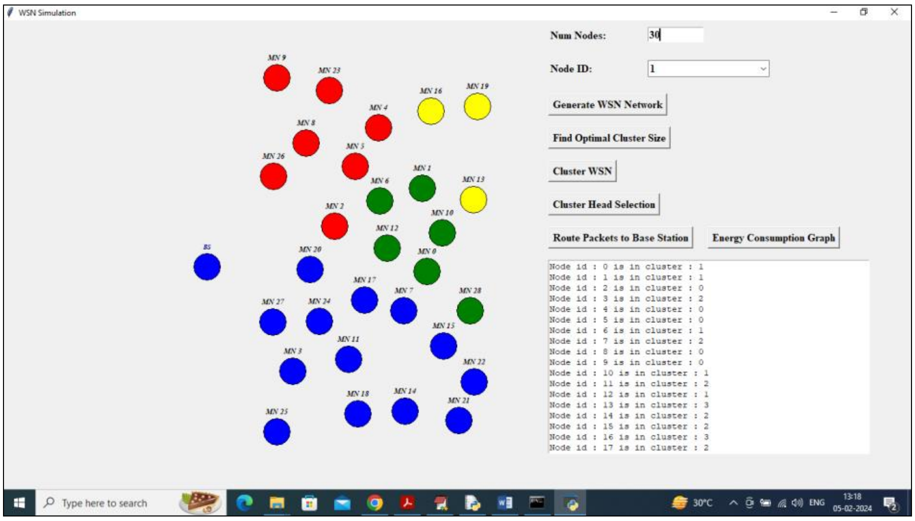
Results and Discussion



Num Nodes: 30

Node ID: 1

To cluster the nodes in the following screen, click on each button in turn; the red ones represent sensor nodes and the blue ones base stations.

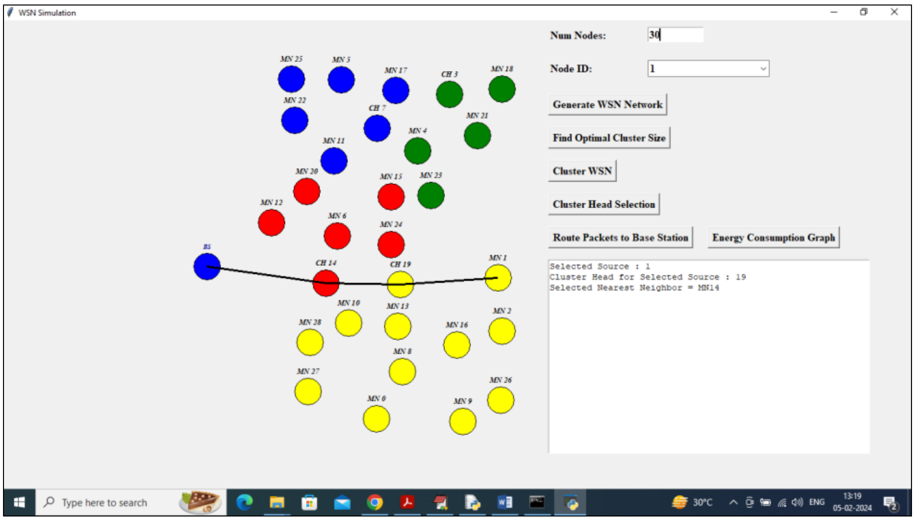


Num Nodes: 30

Node ID: 1

```
Node id : 0 is in cluster : 1
Node id : 1 is in cluster : 1
Node id : 2 is in cluster : 0
Node id : 3 is in cluster : 2
Node id : 4 is in cluster : 0
Node id : 5 is in cluster : 0
Node id : 6 is in cluster : 1
Node id : 7 is in cluster : 2
Node id : 8 is in cluster : 0
Node id : 9 is in cluster : 0
Node id : 10 is in cluster : 1
Node id : 11 is in cluster : 2
Node id : 12 is in cluster : 1
Node id : 13 is in cluster : 3
Node id : 14 is in cluster : 2
Node id : 15 is in cluster : 2
Node id : 16 is in cluster : 3
Node id : 17 is in cluster : 2
```

Nodes of varying colors represent various clusters; to transmit packets to the base station, choose a node and then click the "Route Packets to Base Station" button.

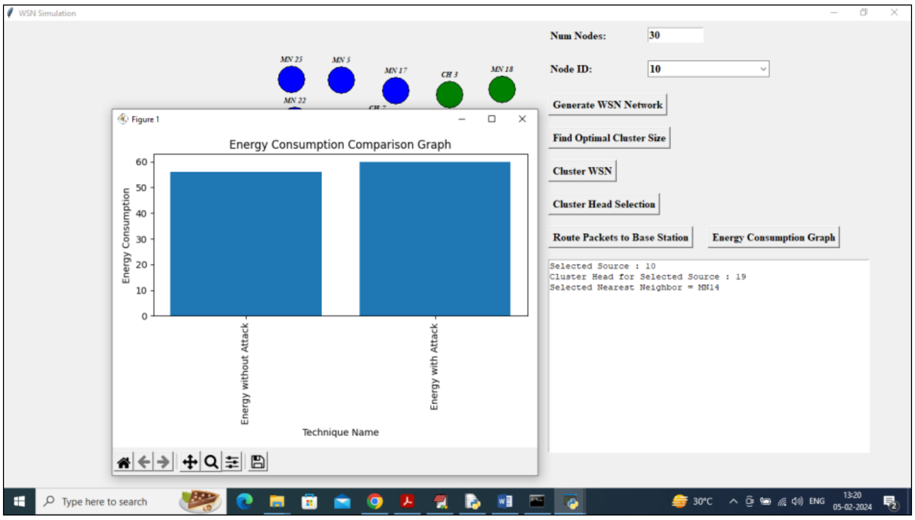


Num Nodes: 30

Node ID: 1

```
Selected Source : 1
Cluster Head for Selected Source : 19
Selected Nearest Neighbor = MN14
```

While the source is transmitting data to the base station, the receiving nodes will either disregard packets that are too big or duplicated, as seen in the previous screen.



Num Nodes: 30

Node ID: 10

```
Selected Source : 10
Cluster Head for Selected Source : 19
Selected Nearest Neighbor = MN14
```

In above graph by ignoring attack packets we can save some energy where x-axis represents packet type and y-axis represents energy consumption.

Conclusion

In this research, we have looked at how well several types of machine learning algorithms—Statistical-based, Logic-based, Instance-Based, and Deep Learning-based—perform on WSN-DS datasets when it comes to detecting denial-of-service attacks. Machine learning techniques such as KNN, LR, SVM, DT, Naïve Bayes, LSTM, and MLP were used. In order to evaluate the algorithms' efficacy on various dataset sizes, the WSN-SD was further partitioned. In addition, we chose one of these methods to study its effects on the WSN network's longevity. Each algorithm's accuracy, F1-score, FPR, FNR, and training execution time were obtained using Python 3.8 inside the Jumyter Network program. The WSN network lifespan may also be obtained using the Cooja simulator, with the help of a clustering management method that is supposedly utilized in contemporary references to control the simulation environment. The dataset obtained from WSN network traffic consists of numerical statistical values, according to the findings of the performance analysis of the machine learning algorithms in WSN-SD. As a result, the greatest performance measures are those of statistical and logical classification algorithms, with Gboost averaging out to the top across all WSN-DS datasets. The training method and high-performance prediction may be accomplished with a dataset of 3000 to 6000 records, provided that the data percentage between the labeled groups is enough to distinguish between them. When compared to KNN, LR, SVM, Naïve Bayes, LSTM, and MLP, Gboost enhanced the average accuracy across all WSN datasets by 0.29%, 2%, 26%, 5%, and 0.8%, respectively. It was almost identical to Gboost in terms of DT accuracy. Gboost outperformed KNN, LR, SVM, Naïve Bayes, LSTM, and MLP in terms of average F1-score by 2%, 5%, 41%, 12%, 58%, and 58%, respectively. Gboost and DT both achieved comparable results in F1-score. In addition, when placed against LR, SVM, DT, Naïve Bayes, LSTM, and MLP, Gboost decreased the average FPR in all WSD-DS datasets by 87%, 97%, 27%, 89%, 86%, and 72%, respectively. In comparison to Gboost, the KNN demonstrated a 36% improvement in reduction. The average FNR was lowered by 63%, 43%, 93%, 82%, and 41% by Gboost when compared to KNN, LR, SVM, LSTM, and MLP, in that order. Gboost had a FNR that was 35% higher than Naïve Bayes and 26% lower than DT. Last but not least, Gboost used 32% less time than DT, 927 percent less time than SVD, 999 percent less time than LSTM, and 913 percent less time than MLP during training. Also, as compared to Gboost, KNN, LR, and Naïve Bayes all used much less average training execution time—128%, 74%, and 808%, respectively. Gboost cut network lifespan by 32% in comparison to the baseline case. Our long-term goal is to expand our current WSN network dataset using the 6LoWPAN protocol and include additional metrics pertaining to packet loss rate, packet size, packet change ratio, and flow change ratio. Additionally, we may get a more robust conclusion about the node state by using the cumulative difference of properly categorized states calculated by the custom sniffer. In addition, we may use or assume a lightweight approach to identify the key aspects of datasets. In order to enhance the effectiveness of machine learning algorithms, the best features are chosen prior to the training step. Further investigation into how machine learning algorithms are affected by the proportion of each class in the dependent variables is required.

References

1. Al-Emran M, Malik SI, Al-Kabi MN. A survey of internet of things (IoT) in education: Opportunities and challenges. In: Studies in Computational Intelligence. Vol. 846. Springer International Publishing; c2020. p. 197-209.
2. Zhang G, Kou L, Zhang L, Liu C, Da Q, *et al*. A new digital watermarking method for data integrity protection in the perception layer of IoT. Security and Communication Networks. 2017;2017:1-12.
3. Yi L, Tong X, Wang Z, Zhang M, Zhu H, *et al*. A novel block encryption algorithm based on chaotic s-box for wireless sensor network. IEEE Access. 2019;7:53079-53090.
4. Al-Kashoash H, Kharrufa H, Al-Nidawi Y, Kemp AH. Congestion control in wireless sensor and 6LoWPAN networks: Toward the internet of things. Wireless Networks. 2019;25(8):4493-4522.
5. Zhang X, Heys HM, Li C. Energy efficiency of encryption schemes applied to wireless sensor networks. Security and Communication Networks. 2012;5(7):789-808.

6. Yang Y, Wu L, Yin G, Li L, Zhao H. A survey on security and privacy issues in internet-of-things. IEEE Internet of Things Journal. 2017;4(5):1250-1258.
7. Glissa G, Meddeb A. 6LoWPAN multi-layered security protocol based on IEEE 802.15.4 security features. In: 2017 13th International Wireless Communications and Mobile Computing Conference. Valencia, Spain; c2017. p. 264-269.
8. Lee C-C. Security and privacy in wireless sensor networks: Advances and challenges. Sensors. 2020;20(3):744.
9. Kaur T, Saluja KK, Sharma AK. DDoS attack in WSN: A survey. In: 2016 International Conference on Recent Advances and Innovations in Engineering. Jaipur, Rajasthan, India; c2016. p. 23-27.
10. Islam MNU, Fahmin A, Hossain MS, Atiquzzaman M. Denial-of-service attacks on wireless sensor network and defense techniques. Wireless Personal Communications. 2020;116:1993-2021.
11. Khan K, Mehmood A, Khan S, Khan MA, Iqbal Z, *et al*. A survey on intrusion detection and prevention in wireless ad-hoc networks. Journal of Systems Architecture. 2020;105:101701.
12. Praveen Kumar D, Amgoth T, Annavarapu CSR. Machine learning algorithms for wireless sensor networks: A survey. Information Fusion. 2019;49:1-25.
13. Cauteruccio F, Cinelli L, Corradini E, Terracina G, Ursino D, *et al*. A framework for anomaly detection and classification in multiple IoT scenarios. Future Generation Computer Systems. 2021;114:322-335.
14. Cheng J, Zhou J, Liu Q, Tang X, Guo Y. A DDoS detection method for socially aware networking based on forecasting fusion feature sequence. Computer Journal. 2018;61(7):959-970.