Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

Deep Learning-Based Double Authentication Protection for MQTT Transactions in Iiot Environment Against Detected Dos Variants

Dr.S.Thavamani M.Sc.,Mphil.,PhD[1], Ms.U.Sinthuja M.Sc.,Mphil[2,3]

[1]Associate Professor, Department of Computer Application, Sri Ramakrishna College of Arts and Science [Formerly SNR SONS College], Coimbatore-6, Tamilnadu.

[2]PhD Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and Science [Formerly SNR SONS College],

[3]Assistant Professor, Department of Information Technology, Hindusthan College of Arts & Science, Coimbatore-6, Tamilnadu.

ORCID : 0000-0001-5873-3459

Abstract: Recently, there has been a great deal of applicability for fast message transmission protection for Industrial Internet of Things (IIoT) applications employing the double authentication paradigm. Of course, malicious people are naturally drawn to any new technology that is widely used by industry and would stop at nothing to take full advantage of it by employing cutting-edge technologies such as Denial of Service (DoS). MQTT, or message queuing telemetry transport, is said to be the easiest protocol for Industrial IoT devices to use. MQTT Protocol-based message transmission protection using double or triple authentication method to secure the MQTT messages from detected DoS attack Variants in IoT. Preprocessing, feature extraction, and hybrid network are the three components of the LHMDA-DL approach that enable secure MQTT message transfer from identified DoS variants. When compared to current approaches, the results show a significant improvement in processing time, traffic overhead, message transmission accuracy, and message transmission error rate.

Keywords: *Industrial IoT, LSTM, Message Queuing Telemetry Transport, Hellinger Neighbor Embedding, Lebesgue Hash Message Double Authentication.*

1. Introduction

The volume of data transferred over global networks increases yearly due to the numerous gadgets linked to computers and information technology networks[1]. The term "Industrial Internet of Things" (IIoT) describes the application of IoT concepts and technology in an industrial environment. It entails tying industrial machinery, systems, and devices to the internet to collect and share data, improving industrial process monitoring, control, and communication. Advances in communication protocols in recent times have made it possible to connect a wide variety of things that are part of the Internet of Things (IoT). Even in a setting with limited bandwidth, technologies such as Message Queue Telemetry Transport (MQTT) have a reasonable chance of eliminating network management.

The amount of IIoT data is increasing as the Industrial Internet of Things (IIoT) develops. Data analytics services are advanced by machine learning. Data analytics service transactions and data packet transactions must be taken into consideration at the same time to streamline data flow and investigate the economic potential of IIoT data. To facilitate the trading of data commodities, centralized data trading platforms are developed. Centralized platforms, however, are unreliable and weak. It is a difficult problem to implement Data Packets in a decentralized manner[2]. Detection systems, which rely on state-of-the-art algorithms, are essential to the security domain. They can secure the foundational system by utilizing methods like machine learning to detect or anticipate security breaches. The best method, named Long Short Term Memory (LSTM), with 87% accuracy, is what has shown as one of the Deep Learning-based strategies for developing the Artificial Intelligence-based Interpolation Technique for IoT Environments[3]. By considering the discussed previous environment the following considerations have been taken to proceed to the next setup.

1.1 Contribution

To facilitate the flow of data and explore the potential value of data in IIoT. A transaction solution containing CMR and CAMR is proposed based on the Lebesgue Hash Message Double Authentication in this paper.

- The MQTTset data training, feature selection, and hybrid networks are taken into consideration when proposing an LHMDA-DL deep classifier to improve message transmission accuracy with the lowest error.
- An Embedding-based Hellinger Distributed Stochastic Neighbour To generate dimensionality-reduced network samples, preprocessing is used. The highly associative Distributed Stochastic Neighbour Embedding function is the basis for this. Followed by Gaussian Mixture-based feature selection applied.
- Hash Message Double Authentication Rules and Long Short-Term Memory are used to protect messages during transmission. These rules are also used to identify the messages of DoS attackers and to secure messages while they are being transmitted.
- The effectiveness of the suggested approach, LHMDA-DL, is quantitatively verified and contrasted with that of other traditional approaches.

2. Related Works

A few relevant papers are examined and discussed in this part. Data analytics service transactions were examined in [4-7]. Data analytics services are regarded as digital goods in [7]. A Bayesian digital commodity auction-based model for optimal pricing and profit maximization was presented. A profit-driven data collection framework for the crowdsourcing-aware data trading market was offered in [5], and a smart data pricing approach for IoT providers was proposed in [4], which determined the purchase price of data packets for the data owner and the subscription cost for the service user. A service platform business concept that links users with wireless sensor networks was proposed[6]. Users can obtain data analytics services from the service platform using data purchased from the wireless sensor network. A proposed business strategy for the service platform connecting users and wireless sensor networks was published in [7]. Users can obtain data analytics services from the service platform using data that has been purchased from the wireless sensor network.

In [8], a trustworthy access control strategy was established via a suggested electronic medical data sharing framework built on the interplanetary file system (IPFS), smart contracts, and blockchain technology. It is safe for patients and healthcare professionals to exchange electronic medical data.Based on blockchain technology, attribute-based encryption (ABE) technology, and IPFS distributed storage, a distributed data storage and sharing system was put out [9]. The problem that the cloud server in the conventional cloud storage system could not return all search results or could return inaccurate results was resolved by the suggested framework, which carried out the function of the ciphertext keyword search. In [10], a method for exchanging IoT data without the need for third-party verification was put forth. The study examined the authentication and interaction between the organizations engaged in data transfers. In [11], a secure data storage and sharing plan based on smart contracts and consortium blockchains for car edge networks was presented. This plan has the potential to successfully stop unwanted data sharing.

A distributed digital asset delivery proof was presented in [12], taking into account the lack of trust, security, and transparency in the current digital asset transfer systems. Transaction logs that are immutable and tamper-proof are provided by smart contracts and blockchain technology. In [13], a distributed data vending framework based on similarity learning and data embedding is described. A real-world application for exchanging electronic medical information was used to examine the concept. Studies on data packet transactions were initiated by a few open-source programs.By utilizing blockchain technology, AAAChain hopes to create an autonomous, decentralized data open platform made up of numerous vertical sector applications [14].

3. Materials Methods

In order to secure MQTT messages from identified DoS attack variants in IIoT environment, a hybrid deep learning technique known as Lebesgue Hash Message Double Authentication fused Deep Learning (LHMDA-DL) based message transmission security is proposed in this section. To get rid of the duplicate instances, the LHMDA-DL first preprocesses. The preprocessed network samples are then subjected to a Lebesgue measure Gaussian Mixture in order to extract features. Finally, a hybrid network using long short-term memory and hash message double authentication rules is used to confirm the goal.

3.1 Problem Statement

The main emphasis of the research is to focus on most important attacks like Denial of Service(DoS) in IoT mainly, by focusing on a widely used IoT protocol called MQTT.

- Because of the loss of safety mechanisms in IoT devices, many IoT devices come to be soft targets by the attacker, sometimes with data loss or traffic and so on.
- Attack Detection at the time of Publish and Subscribe is needed.
- Attack Detection in Application Layer without secure data communication between devices leads to increased false positive rates.
- Trusted Devices with multiple Attack vectors need to be concentrated.

3.2 Dataset Description

By guaranteeing safe data transmission between devices, the suggested approach seeks to provide a dataset related to the Internet of Things context, or more specifically, the MQTT communication protocol, to provide the industrial community with a first dataset for use. The MQTT dataset [15] consists of Internet of Things (IoT) sensors built on top of MQTT, which defines every element of a real network. Additionally, Eclipse Mosquitto is used to instantiate the MQTT broker, and the total network comprises eight sensors related to the industrial sector.

3.3 Hellinger Distributed Stochastic Neighbor Embedding (HDSNE) based Preprocessing

Consistent data in various patterns are presented by the MQTTset dataset. Processing is therefore required prior to enabling the security of MQTT communications from detected DoS attack variants, as using the raw data obtained from the MQTTset dataset to do so cannot preserve significant results. The MQTT protocol was utilized to investigate the abstract behavior of 25,000 samples. Additionally, redundant instances from the MQTTset dataset are removed using HDSNE to minimize the dataset's dimension. The HDSNE used in our work embeds high dimensional data in low dimensional space in a way that, with high probability, models dissimilar objects (i.e., sensors or devices) by remote locations and models similar objects (i.e., sensors or devices) by adjacent locations.

Input: Dataset 'DS', Features '$F = \{F_1, F_2, ..., F_m\}$', Data '$D = \{D_1, D_2, ..., D_n\}$', Sensor '$S = \{S_1, S_2, ..., S_s\}$'
Output: dimensionality reduced network samples '$x \in DRNS$'
1: **Initialize** 'm', 'n', Network Samples 'NS', 's'
2: **Begin**
3: **For** each Dataset 'DS' with Features 'F', Data 'D', Sensor 'S' and Network Samples 'NS'
4: Formulate sample vector matrix as given in (1)
5: Measure probability of object similarity positioned in a room as given in (2)
6: Evaluate similarity of data point or sensors as given in (3)
7: Measure Hellinger distance with which similarity between two probability distributions is quantified as given in (4)
8: **Return** dimensionality reduced network samples '$DRNS$'
9: **End for**
10: **End**

Algorithm 1Hellinger Distributed Stochastic Neighbor Embedding-based Preprocessing

3.4 Lebesgue measure Gaussian Mixture(LMGM) based Feature Extraction

Using MQTT user-chosen network samples as a baseline, we first estimate the local neighborhood histogram of each selected goal feature. A neighborhood's quantitative features may be obtained and good perceptions of large features can be obtained by the MQTT user with the use of the local neighborhood histogram, which is essentially a histogram measured according to the neighborhood principle. After that, we used Gaussian Mixture Models to

accurately describe these target feature local neighborhood histograms, resulting in a large number of potential GMMs (i.e., features extracted).

Here is the definition of the Gaussian mixture model used to extract the features with reduced dimensionality based on the probability density of a random feature. On the basis of the Lebesgue correlative measure (CM2), "Likelihood F" is determined in (1-3) equations

$$Prob(F) = \sum_{i=1}^{m} w_i * N\left(F|\mu_i, \sigma_i\right) \quad \text{(1)}$$

$$DRNS = H^2\left(S_i, S_j\right) = \frac{1}{2}\int \left(\sqrt{S_i(F)} - \sqrt{S_j(F)}\right)^2 \alpha\,(dF) \quad \text{(2)}$$

$$Likelihood\ (F) = \exp exp\left(\frac{\Delta^2}{D}\right) \quad \text{(3)}$$

For training the proposed technique, the proposed work considers the network traffic features Connect and ConnectAck of CONNECT and CONNACK messages, respectively, in addition to the extracted features. Getting the two crucial variables, Connection Message Ratio (CMR) and Connection Acknowledgement Message Ratio (CAMR), is the next stage. Subsequently, the variable CMR represents a percentage of connection requests made by the publisher, expressed mathematically as follows in equations (4&5).

$$CMR = \frac{N_{connect}}{N\ Total} \quad \text{(4)}$$

$$CAMR = \frac{NConnAck}{N\ Total} \quad \text{(5)}$$

The Hash Message Double Authentication Rules Will be carried out by the MQTT broker if any anomalies are found. In the absence of such, the subscriber receives the standard data packets. The HMAC inference engine creates the authentication rules as an IF-THEN statement based on the input metrics CMR and CAMR.

Input: Dataset 'DS', Features 'F=F1, F2, ...,Fm', Data 'D=D1, D2, ...,Dn', Sensor 'S=S1, S2, ...,Ss'
Output: Computationally efficient secure message transmission with detected DoS variants
Step 1: **Initialize** dimensionality reduced network samples 'DRNS', 'm', 'n', 's', count of the connect feature 'Nconnect', total number of MQTT messages 'N', connection acknowledgement messages from the MQTT broker **Step 2**: **Begin** **Step 3**: **For** each Dataset 'DS' with Features 'F', Data 'D', Sensor 'S' and dimensionality reduced network samples 'DRNS' **//Feature extraction** **Step 4**: Formulate Gaussian mixture model to extract the dimensionality reduced features as given in equation (5) **Step 5**: Obtain the fittest GMM using the Lebesgue measure as given in equations (6) and (7) **Step 6**: **If** 'Likelihood F>0.5' **Step 7**: **Return** features extracted 'FE' **Step 8**: **End if** **Step 9**: **If** 'Likelihood F≤0.5' **Step 10**: Non trivial features **Step 11**: **Go to** step 27 **Step 12**: **End if** **//Feature selection** **Step 13**: Evaluate Connection Message Ratio as given in equation (8) **Step 14**: Evaluate Connection Acknowledgement Message Ratio as given in equation (9) **Step 15**: If 'CMR=LOW and CAMR=LOW' **Step 16**: **Go to step** 18 **Step 17**: **End if** **Step 18**: Formulate Hash Message Authentication Code Double Authentication rules as given in equations (10) and (11) **Step 19**: **If** 'CMR=LOW and CAMR=MEDIUM' **Step 20**: **Go to** step 18 **Step 21**: **End if** **Step 22**: **If** 'CMR=HIGH and CAMR=MEDIUM' **Step 23**: **Go to step** 18 **Step 24**: **End if** **//Secure message transmission** **Step 25**: Formulate forgetting gate as given in equation (12) **Step 26**: Update the information according to the resultant forgetting gate activation vector as given in equation (13) **Step 27**: Obtain output gate's activation vector and cell input gate's activation vector as given in equations (14) and (15) **Step 28**: Update cell state as given in equations (16) and (17) **Step 29**: **End for** **Step 30**: **End**

Algorithm 2: HMDAR & LSTMQ

First, relevant features are retrieved using the Gaussian mixture model, as per the above algorithm, with the dimensionality reduced network samples and the associated MQTT client or sensor provided data packets. This is followed by the extraction of high likelihood features using the fittest GMM and the measurement of Lebesgue. Second, using the retrieved features, the Connection Message Ratio and Connection Acknowledgment Message Ratio are used to determine which features are the most selective. Consequently, MQTT messages identified as DoS variations are subject to the application of Double Authentication rules using the hash message authentication code. Lastly, computationally effective message transmission protection is guaranteed by utilizing the hyperbolic tangent function.

4. Results & Discussion

In this section, we developed Lebesgue Hash Message Double Authentication fused Deep Learning (LHMDA-DL) based message transmission protection to secure MQTT messages from detected DoS attack Variants in [16]. Using the MQTTset Dataset, the effectiveness of the LHMDA-DL approach has been examined and verified in terms of processing time, traffic overhead, message transmission error rate, and correctness.

4.1 Processing Time

The following is a mathematical statement of the processing time. From the equation-6 formulated with network samples which are denoted as NS by adding the published and subscribed time of MQTT Table-1 gives the comparison between Secure Reliable Message Communication (SEC-RMC), Deep Learning and IoT-Based Monitoring System as well as LHMDA-DL. The proposed system is identified with minimal processing time by comparatively in figure-1.

$$PT = \sum_{i=1}^{N} NS_i * Time\,[Pub + Sub] \quad (6)$$

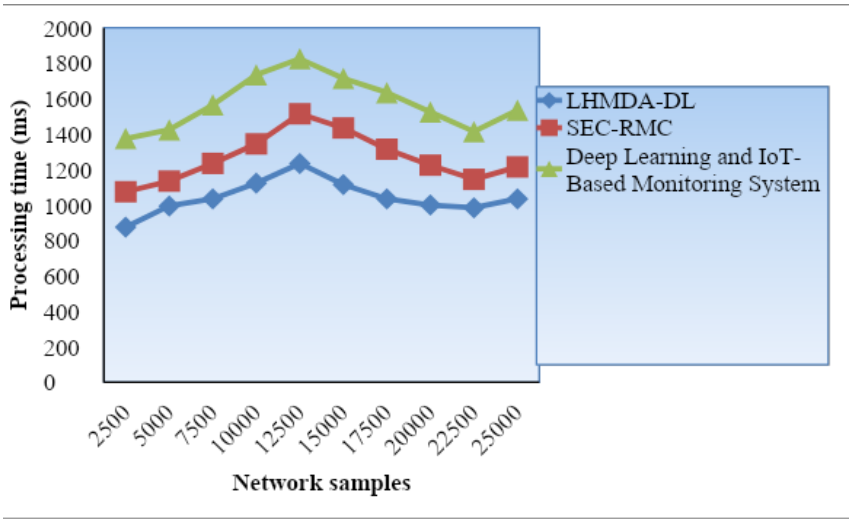

Figure-1 : Processing Time Comparison of Proposed Scenario

Table -1: Processing Time Comparison of Proposed Scenario

Network samples	Processing time (ms)		
	LHMDA-DL	SEC-RMC	Deep Learning and IoT-Based Monitoring System
2500	875	1075	1375
5000	995	1135	1425
7500	1035	1235	1565
10000	1123	1345	1735
12500	1235	1515	1825
15000	1115	1435	1715
17500	1035	1315	1635
20000	1000	1225	1525
22500	985	1145	1415
25000	1035	1215	1535

4.2 Trafic Overhead

This section discusses the traffic overhead about the recommended techniques for enhancing MQTT interaction security. The following is a mathematical formulation of the traffic overhead.

$$TO = \sum_{i=1}^{N} NS_i * Mem\,[Pub + Sub] \quad (7)$$

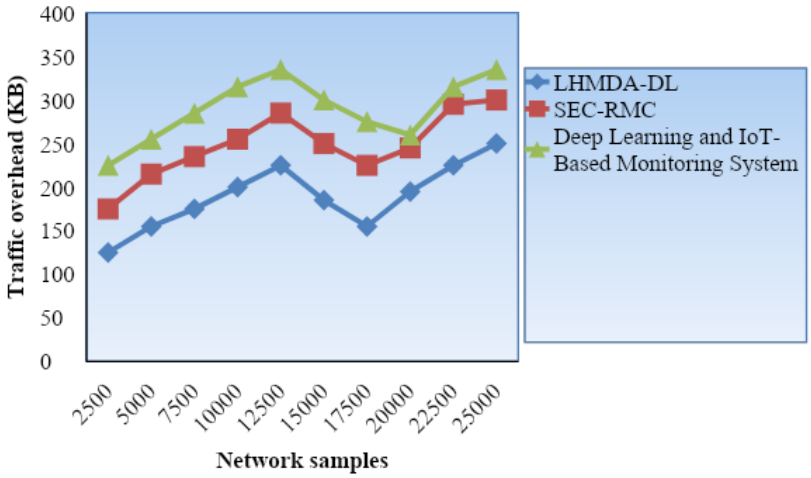

Figure -2: Traffic Overhead

Table -2: Traffic Overhead Comparison of Proposed Scenario

Network samples	Traffic overhead (KB)		
	LHMDA-DL	SEC-RMC	Deep Learning and IoT-Based Monitoring System
2500	125	175	225
5000	155	215	255
7500	175	235	285
10000	200	255	315
12500	225	285	335
15000	185	250	300
17500	155	225	275
20000	195	245	260
22500	225	295	315
25000	250	300	335

4.3 Message Transmission Accuracy

This section presents the message transmission accuracy analysis of the method's efficacy and efficiency. The evaluation of the message transmission accuracy is provided below.

$$MTA = \sum \frac{NS_{recvd}}{NS_{sent}} * 100 \quad (8)$$

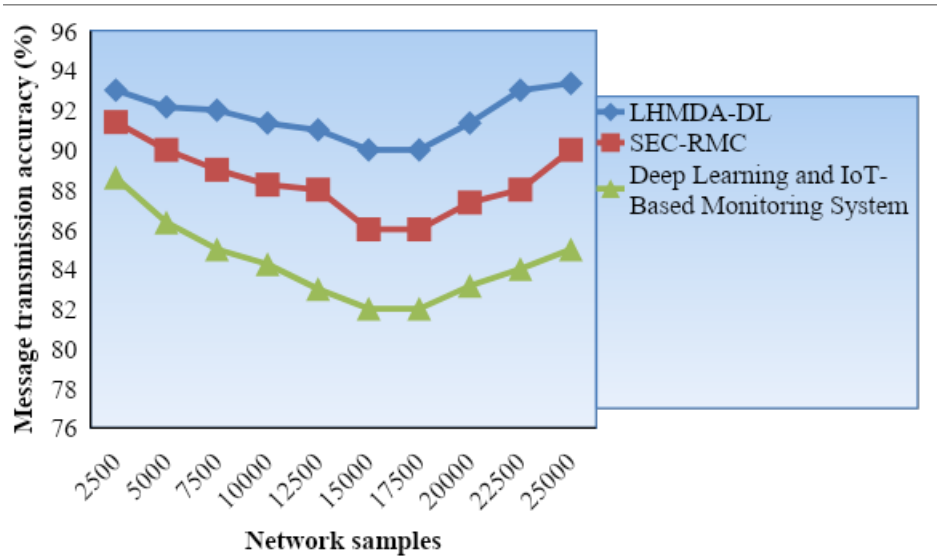

Figure-3 Message Transmission Accuracy

Table -3: Message Transmission Accuracy Comparison of Proposed Scenario

Network samples	Message transmission accuracy (%)		
	LHMDA-DL	SEC-RMC	Deep Learning and IoT-Based Monitoring System
2500	93	91.4	88.6
5000	92.15	90	86.35
7500	92	89	85
10000	91.35	88.25	84.25
12500	91	88	83
15000	90	86	82
17500	90	86	82
20000	91.35	87.35	83.15
22500	93	88	84
25000	93.35	90	85

4.4 Message Transmission Error Rate

This stage finally yields the message transport error rate. This part assesses the message transmission error rate in order to verify the method's effectiveness.

_____(9)

$$MTE = \sum \frac{NS_{mixed}}{NS_{sent}} * 100$$

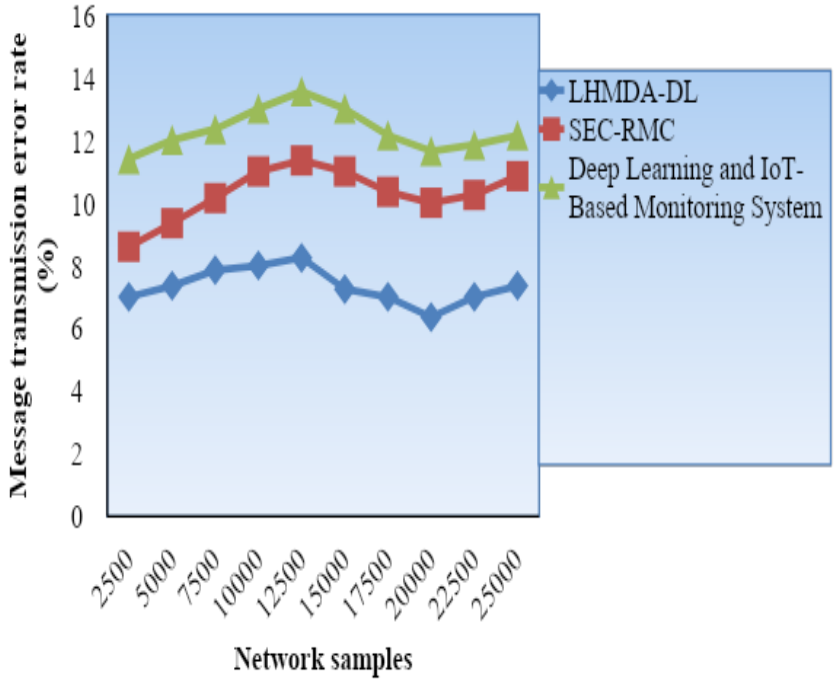

Figure-4: Message Transmission Error Rate

Table-4 : Message Transmission Error Rate Comparison of Proposed Scenario

Network samples	Message transmission error rate (%)		
	LHMDA-DL	SEC-RMC	Deep Learning and IoT-Based Monitoring System
2500	7	8.6	11.4
5000	7.35	9.35	12
7500	7.85	10.15	12.35
10000	8	11	13
12500	8.25	11.35	13.55
15000	7.25	11	13
17500	7	10.35	12.15
20000	6.35	10	11.65
22500	7	10.25	11.85
25000	7.35	10.85	12.15

5. Conclusion

According to industry trends, IIoT devices are employed in fields where collaboration, message transmission, and automation of industrial processes are important. In order to secure MQTT messages from identified DoS attack variants, a Lebesgue Hash Message Double Authentication fused Deep Learning (LHMDA-DL) based message transmission security is presented in this paper. Through the use of the Hellinger Distributed Stochastic Neighbour Embedding-based Preprocessing technique, first dimensionality reduced network samples are created. Next, a Lebesgue measure Gaussian Mixture based Feature Extraction technique is used to extract relevant features from network data that have been decreased in dimensionality. Finally, the hybrid deep learning model receives as input the dimensionality-reduced network samples and relevant extracted features.

6. References

[1] Li, Y.; Chi, Z.; Liu, X.; Zhu, T. Chiron: Concurrent high throughput communication for IoT devices. In Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services, Munich, Germany, 11–15 June 2018, pp. 204–216.

[2] Y. Jiang, Y. Zhong, and X. Ge, "Smart Contract-Based Data Commodity Transactions for Industrial Internet of Things," in IEEE Access, vol. 7, pp. 180856-180866, 2019, doi: 10.1109/ACCESS.2019.2959771.

[3] S. Thavamani and U. Sinthuja, "LSTM based Deep Learning Technique to Forecast Internet of Things Attacks in MQTT Protocol," 2022 IEEE Fourth International Conference on Advances in Electronics, Computers and Communications (ICAECC), Bengaluru, India, 2022, pp. 1-4, doi: 10.1109/ICAECC54045.2022.9716585.

[4] D. Niyato, D. T. Hoang, N. C. Luong, P. Wang, D. I. Kim, and Z. Han, ‘‘Smart data pricing models for the Internet of Things: A bundling strategy approach,’’ IEEE Netw., vol. 30, no. 2, pp. 18–25, Feb. 2016.

[5] Z. Zheng, Y. Peng, F. Wu, S. Tang, and G. Chen, ‘‘Trading data in the crowd: Profit-driven data acquisition for mobile crowdsensing,’’ IEEE J. Sel. Areas Commun., vol. 35, no. 2, pp. 486–501, Feb. 2017.

[6] L. Guijarro, V. Pla, J. R. Vidal, and M. Naldi, ‘‘Maximum-profit two-sided pricing in service platforms based on wireless sensor networks,’’ IEEE Wireless Commun. Lett., vol. 5, no. 1, pp. 8–11, Feb. 2016.

[7] Y. Jiao, P. Wang, S. Feng, and D. Niyato, ‘‘Profit maximization mechanism and data management for data analytics services,’’ IEEE Internet Things J., vol. 5, no. 3, pp. 2001–2014, Jun. 2018.

[8] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, ‘‘Blockchain for secure EHRs sharing of mobile cloud based e-Health systems,’’ IEEE Access, vol. 7, pp. 66792–66806, 2019.

[9] S. Wang, Y. Zhang, and Y. Zhang, ‘‘A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems,’’ IEEE Access, vol. 6, pp. 38437–38450, Jun. 2018.

[10] A. Suliman, Z. Husain, M. Abououf, M. Alblooshi, and K. Salah, ‘‘Monetization of IoT data using smart contracts,’’ IET Netw., vol. 8, no. 1, pp. 32–37, Jan. 2019.

[11] J. Kang, R. Yu, X. Huang, M. Wu, S. Maharjan, S. Xie, and Y. Zhang, ‘‘Blockchain for secure and efficient data sharing in vehicular edge computing and networks,’’ IEEE Internet Things J., vol. 6, no. 3, pp. 4660–4670, Jun. 2019.

[12] H. Hasan and K. Salah, ‘‘Proof of delivery of digital assets using blockchain and smart contracts,’’ IEEE Access, vol. 6, pp. 65439–65448, 2018.

[13] J. Zhou, F. Tang, H. Zhu, N. Nan, and Z. Zhou, ‘‘Distributed data vending on blockchain,’’ 2018, arXiv:1803.05871. [Online]. Available: https://arxiv.org/abs/1803.058

[14] AAAChain White Paper. Accessed: Dec. 16, 2019. [Online]. Available: https://aaachain.net/#whitepaper

[15] Vaccari I, Chiola G, Aiello M, Mongelli M, Cambiaso E. MQTTset, a New Dataset for Machine Learning Techniques on MQTT. Sensors. 2020; 20(22):6578. https://doi.org/10.3390/s20226578

[16] S. Thavamani and U. Sinthuja, "LSTM based Deep Learning Technique to Forecast Internet of Things Attacks in MQTT Protocol," 2022 IEEE Fourth International Conference on Advances in Electronics, Computers and Communications (ICAECC), Bengaluru, India, 2022, pp. 1-4, doi: 10.1109/ICAECC54045.2022.9716585