THINK INDIA JOURNAL

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

An Efficient and Secure Data Collection Mechanism By Using Internet of Things

Sofia, Dr. Arun Malik, Dr. Isha, VikasVerma

School of Computer Science and Engineering

Lovely Professional University, Punjab

Abstract

With high demand of efficient data collection from Internet of Things devices, researchers' interest in providing new data collection methods is getting growing every day. Many data collection methods have been proposed, but still more efficiency is required. In this paper, a novel and secure data collection method from Internet of Things devices has been proposed. In this paper, creation of aggregator node and data collection from nodes within a cluster mechanisms have been proposed. Integrity of data will be maintained using hashing mechanisms. In earlier scenarios, there was an issue regarding fault tolerance in a network. To resolve this issue, aggregator node will be updated after each successful transmission of data to the fog server. As the huge amount of data transmission can create various challenges in network, so this proposed mechanism will provide the better results in terms of energy efficiency, storage requirement and latency as well.

Introduction

As now-a-days, IoT smart devices are expanding in whole world and due to which a biggest change in world is coming. Today everything we are using have embedded sensors. Sensor enables various physical devices to monitor the things and provide the results to the users. IoT has widely used in every field like smart home systems[1], smart cities[2], [3], healthcare[4], [5], Unnamed Aerial Vehicles[6], [7], monitoring conditions of crops in agriculture, fire alarms, environmental monitoringetc.

Internet of Things process mainly comprises with the two phases- data collection and data mining. Data collection in IoT refers with the receiving or extracting from various IoT devices. Data mining is extracting the useful patterns from the collected data. Data collection is one of the most critical phase in the whole process. This whole process resides on the communication within devices and server or repositories. Communication can be perform using various protocols. There are various protocols that can be used for communication inIoTapplications. Protocol like XMPP (Extensible Message Persistent Protocol), CoAP (Constrained Application Protocol), MQTT (Message Queuing Telemetry Transport) are mainly used protocols[8]. XMPP protocol uses the XML technology for the communication and file transmission within nodes in distributed network. XMPP protocol can be used in audio, video data file transfer based on TCP, instant messaging, chats etc. CoAP protocol mainly based on UDP which can be customized for constrained node and networks. CoAP uses four various message types for requesting resouces- PUT, GET,DELETE and POST[8]. MQTT protocol is lightweight protocol which provides M2M communication. In MQTT, publish-subscribe architecture has been used. Node who needs data send the subscribe message to a particular node and that node publish the data to subscribed node. These protocols has various advantages and disadvantages[8].

Now-a-days, there are numerous attacks which are breaking the confidentiality, integrity of the data. In IoT applications, security of data is the biggest challenge. There can be many attacks that are breaching the security of data[9]. In the regard of IoT data security, IoT developers and researchers should take some initiatives. However, everyday new security methods are developing. But still, there is high need to secure the data more.

Literature Review

There are numerous data collection methods already exists in order to improve the network lifetime, energy efficiency, storage requirements etc. In order to aggregate the data more securely and efficiently, [5] propose two different algorithms – message receiving and message extracting. Message receiving algorithm is to aggregate the data at aggregator node whereas message extracting algorithm is proposed to use at fog server[5]. After the proper aggregation of data, it was stored using local repositories and later uploaded to cloud repositories. This method shown less energy consumption, storage requirements, good transmission ratios etc. [5]. A hybrid scheme including HBH ARQ (Hop-by-Hop Automatic Repeat Request) and PR (Packet Reproduction) was proposed [10] to make data transmission more energy efficient and reliable. The results were evaluated using extensive simulators. Compression sensing theory had been proposed to transmit the compressed data as well as to increase the network lifetime[11]. This proposed schemewasdividedintotwopartitions-clusteringofdatabasedontheirspatial

Correlation and other is reconstruction of data at edge of the network [11]. This reconstruction of data was done by ADMM algorithm. A data collection method or algorithm was proposed for tracking the wearable and mobile devices within IoT systems[12]. This was proposed to collect the complex data even at very minimum energy consumption. The algorithm was partitioned into three phases- Initial phase, IoT edge phase and Fusion Center phase[12]. Results were evaluated using real datasets. As data collection process is for concurrent users, so to eliminate the single user problem, concurrent data aggregation tree mechanism were proposed[13]. It compares the existing single user structure with proposed structure. Effective durations were achieved using thismechanisms[13].

Sometimes, biggest challenge becomes to integrate the data because data is usually heterogeneous in nature. In order to remove this problem, a model based on Hidden Markov Process for integrating heterogeneous data were proposed[14]. Co-operative event detection case study also used in this mechanism. Later, good performance in terms of power consumption and accuracy were achieved[14]. Data collection form the sensors or IoT devices is easy but to preserve the privacy of collected data APPA(anonymous privacy preserving scheme with authentication) mechanism were proposed[15]. Paillier algorithm were used to securely aggregate the data from various sensory devices. To provide the real time services for device registration and device update, fog-enhanced systems were used[15]. Later, good performance in terms of computational complexity and communicational overhead were achieved. HealthIIoT- enabled monitoring mechanism were proposed to collect the healthcare data from various mobile devices and IoT systems[16]. Water marketing, signal enhancements and various analytical techniques were used to secure the data from data thefts and other attacks[16]. Results were evaluated using both real world data and simulations. Energy consumption is the big challenge while transmitting the data from one end to other. If data is being sent in compressed form then energy consumption can be reduced. So, a model low density parity check code (LPDC) was designed to compress the data size[17]. And to resolve the same problem, one another method using Lyapunov optimization theory were proposed[18]. Encrypting data can be compressed using this LDPC code. With the help of using this same model, privacy and confidentiality of data can be maintained. The basic idea of this research was to transfer the data accurately,topreservetheconfidentiality,integrityofthedataandtoutilizetheefficient amount of the energy[17]. A new multi representative re-fusion method were proposed for the data collection from multi sensors with high performance in terms of energy consumption as well as network lifetime[19].While sending data to third party, a trusted authority always a requirement. But a novel privacy preserving raw data collection method was proposed to communicate with the third party without any trusted authority[20]. In this method, data remains in the raw format to increase the value for data consumer. So, that no outsider can know the source of data himself. Another data collection method SW- SS were proposed for the security and confidentiality of data. In this, PrivacyProtector mechanism were used to protect the patients' whole data from the attacks that can breach the security of the data[21]. When data is aggregated from the various IoT nodes and later updated in cloud repositories or in local repositories, then there can be delays in this process. So, to minimize the delays a probabilistic model for many- to-one communication process has been proposed[22]. A delay and energy-efficient data collection method based on matrix filling theory has been proposed to collect the randomly generated data[23]. In this TDMA theory has been used for efficient transmission. Moreover, DEEDC method used clustering data aggregation method. A smart gateway data collection method using MDA plug-in mechanism over traditional gateway data collection method has been proposed to provide the better quality of performance[1]. Moreover, in this method cloud controller has been used to deal with the controlpolicies.

As large amount of data can be transmit from multiple sensing nodes which can create congestion and delay related challenge, so to tackle with this problem CDCAPC mechanism were proposed[24]. A new ring based data collection scheme were proposed, which aggregate the data ring by ring from outside to inside in wireless sensor networks[25]. All the packets were unicast to the destination and later, these were adjusted using fuzzy logic concept. The proposed scheme ensures the reliability as well as the energy efficiency[25]. The biggest challenge to store the data in cloud is security of data because cloud is not much secure. To address the problem related with the security of data, proxy re-encryption method was proposed[26]. Another framework known as SecureData, including FPGA hardware based cipher and sharing cipher secret algorithm has been proposed for the security of data collection[27]. A model for Android platform known as OppNet- Opportunistic Networking Connectivity Servicewere Proposed[28]. This model includes the context awareness routing and deployment of citizen centric data collection method. Battery level and backpressure were two parameters of the model[28]. Later, a novel architecture including two sub architectures- MF-R architecture (Apache Pig and Apache HBase technologies were used for the data collection and storage) and GC architecture (securing data using fog computing) were proposed[29]. Later, DDSV method for optimizing delay and delivery ratio for data collection in multimedia was proposed [30]. Unnamed Aerial Vehicles were utilized by using compressive data gathering solution methodology which helps to transfer the data from cluster node to sink node[7].

System Model and Problem Identification

In this section, we present the system model for data collection from IoT nodes. Figure 1, shows that how IoT nodes sending data to fog server. Earlier many methods had proposed by different researchers, but they all are having some problems. The major problem was with aggregated node. Some methods don’tmention any alternative way of collecting data, if the most important node i.e. aggregated node fails. As, if aggregated node will fail whole transmission or model will fail. To overcome this problem, a new model has been proposed. It will also help to avoid congestion in the network. The whole mechanism is updating the aggregating node after each transmission according to energy. Fog server is playing the vital role in this wholeprocess.

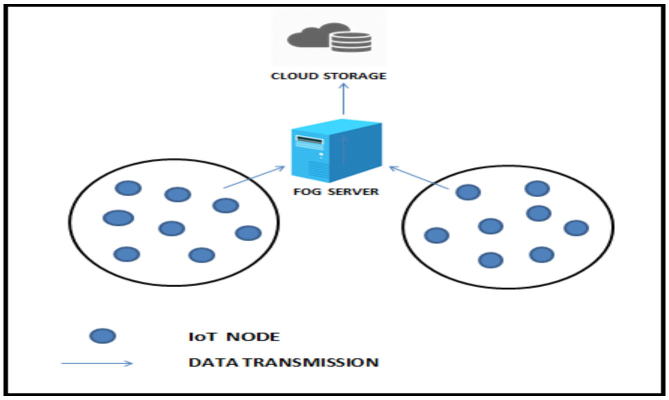

Figure 1: System model for collecting data from IoT nodes

In this scenario, there are various clusters of IoT nodes based on their regions. The first task is to find the distance of IoT nodes from the fog server. There can be various ways to find the distance. But we opt Dijkstra algorithm to find the distance of nodes from the fog server. After finding the distance of each node,the second task is to analyze the shortest distance. The node having the shortest distance from the fog server will be consider as the first cluster head or aggregated node of the cluster. The cluster head or aggregated node broadcast the message about its cluster head information to all the nodes within the cluster and fog server. For the communication between nodes and fog server, MQTT (Message Queuing Telemetry Transport) protocol has been used. As user needs to store or collect the data at fog server,so fog server first initiates the communication by sending the subscribe message to cluster head. Then, later cluster head broadcast the subscribe message to each of the IoT node within its cluster. All cluster members send publish message with its data to cluster head. But before publishing the message, symmetric key mechanism will be done. When cluster head will broadcast the cluster head information to its cluster members and fog server, then at that time cluster head concatenates the secret key information in the data. Then cluster members performs the compression of data by using this key. They will perform the XOR operation between the data and the key. Then, the resultant data will be known as compressed data. Later, cluster members will be finding the hash of the data. At last, they will send the publish message to the cluster head by concatenating the compressed data, device ID, key and hash of the data. The whole data coming from various nodes will be collected at cluster head or aggregating node. The cluster head or aggregating node will concatenate the whole data as one unit and it publish this data to the fog server. Now, whole function will be perform at the fog server. At fog server, the hash of the data after uncompressing the data will be perform. If hash matches will the sent hash by the source, then data will accepted and store at the fog server otherwise it will be rejected. Now, next task is to creating the next cluster head. When all nodes send the data to cluster head(aggregated node) or fog server, then at that time they will concatenate its energy residual information. Energy Residual is the energy left in the node after performing any function. The IoT node having the maximum energy residual will be selected as the next cluster head or aggregated node. The selection of cluster head will be done after each transmission. In this way, whole process will goingon and large amount of data from the IoT node can be collected and stored at the server.

In IoT data collection models, storage requirements, energy efficiency and latency is one of the most challenging parameters. This whole scenario will be evaluated based on the energy efficiency, storage requirement and latency. Later, whole analysis will be done to prove that this model is using efficient amount of resources. Updating the cluster head is quite challenging, but this model is performing very accurately without disturbing any of otherfunction.

Proposed Algorithm

This section presents the whole working of the model. In this section, two algorithms have been proposed. Algorithm 1. (Creation of Aggregator Node) further subdivides into two parts- Creation of Initial Aggregator Node and Updated Aggregator node. Creation of initial aggregator node will represent that how the data will be collected in first cycle. The whole process is given in Algorithm 1 (I). Updated Aggregator node represents that how an aggregated node will change after each transmission. The main idea to update the node is to increase the network lifetime. If aggregator node will be updated, then the data transmission load to the fog server will be dividing to different nodes and the chances of node failure will decrease. Whereas, Algorithm 2. represents the data collection from each node and sending of the same data at the fog server. The whole process of securing data, collecting data, accepting data, rejecting data will be given in Algorithm 2. The idea process of transmitting data have taken from MQTT protocol.

Algorithm 1. Creation of Aggregator Node

I. CreationofInitialAggregatorNode

1. Deploythenodesandmaketheclustersofthatnodes.
2. Establishthepathsfromfogservertoeachnodewithinacluster.
3. Findthedistancefromthefogservertoeachnodepresentinacluster.
4. Aftermeasuringthedistancebetweeneachnodeandfogserver,nextpartistoanalyzethe resultsthathavecomeaftercalculatingthedistance.
5. Choosethesmallestdistancefromallcalculateddistances.
6. Makethatnodeasaggregatornodewhichishavingsmallestdistanceforfogserver.

II. Updated AggregatorNode

1. AggregatornodewillbeupdatedaftereachtransmissionusingAlgorithm2.
2. Aggregatornodewillbeupdatedusingenergyresidualstatusinresponsepacketasmentioned in Algorithm2.
3. Afteranalyzingtheenergyresidualstatuses,choosethemaximumenergyresidualofanode.
4. Makethatnodeasupdatedornewaggregatornode.
5. Thiscyclewillberepeatingtillwholedatacollection.

Algorithm 2. Data Collection from nodes within a cluster

1. Fog server initiates the communication for collecting data from the nodes by sending the requesttoaggregatornodeusingsubscribemessagepacket.
2. WhenAggregatornodereceivesthemessagefromfogserver,thenaggregatornodewillstart theconversationwiththenodeswithinthecluster.
3. Aggregatornodesendthesubscribemessagetoallofthenodeswithinthecluster.
4. When nodes collects the subscribe message, then all of the nodes will start preparing for responses.
5. Each Node will compress the data using symmetric key encryption and send the publish message to aggregatornode.
6. Publishmessagewillcontainthecompresseddataofthenodes,hashofthedata,energy residualstatusandnodeid.
7. Whenaggregatornodereceivesthepublishmessage,thenitwillanalyzethe dataand concatenateitsowndataandsendtothefogserverasasingleunit.
8. Aftercollectingresponsefromtheaggregatornode,fogserverwillcheckthehashesofthe data.
9. Fogserverwillcalculatethehashofdataofeachnode.
10. Comparisonwillbedonebetweenreceivedhashesandcalculatedhashesbyfogserver.
11. Ifthecalculatedhashmatcheswiththereceivedhashofanode,thenthatnode'sdatawillbe acceptedotherwisedatawillbediscardedfromthesamenode.
12. Thiscyclewillberepeatingineachtransmission.

Basic Terminologies used in Algorithm 2:

Subscribe message: Packet which initiates the communication and carries the secret key.

Energy Residual: Remaining energy of a node after performing a function.

Symmetric key mechanism: Performing the XOR operation between original data of a node and shared key.

Compressed Data: Data after performing symmetric key mechanism.

Node Id: Node Id helps to create unique nodes in a cluster. It will be assigned number to each node like 1,2,3...It will help to provide information regarding data of each node.

Publish message: Packet which consists of compressed data of the nodes, hash of the data, energy residual status and node id.

Flow of data in proposed model

In this section, the whole data exchange process between nodes and fog server via aggregator node is explained. After doing whole pre-processing mechanisms, fog server will initiate the data exchange process. Fog server will send the subscribe message to the initial aggregator node. Later, aggregator node will send the subscribe message to all of the nodes within the cluster. Then, nodes will start preparing the data packet and sends the publish message to the aggregator node. Aggregator node will send the final publish message after concatenating the whole data to the fog server. Figure 2. representing the same mechanism and performs the efficient data collection from the various unique node and stores at the fog server.

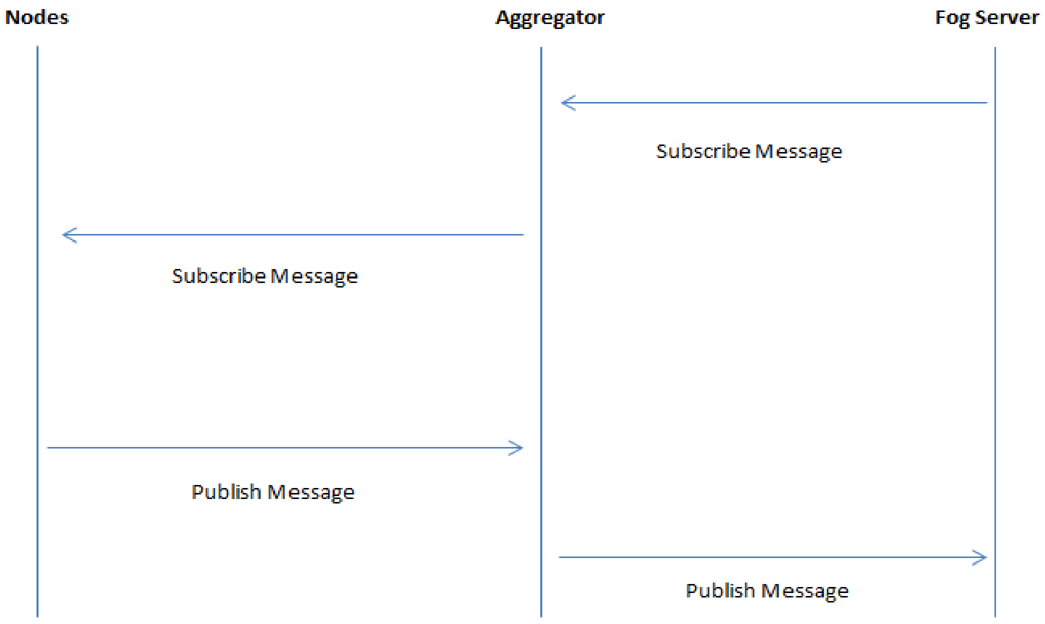

Figure 2 Data flow mechanism

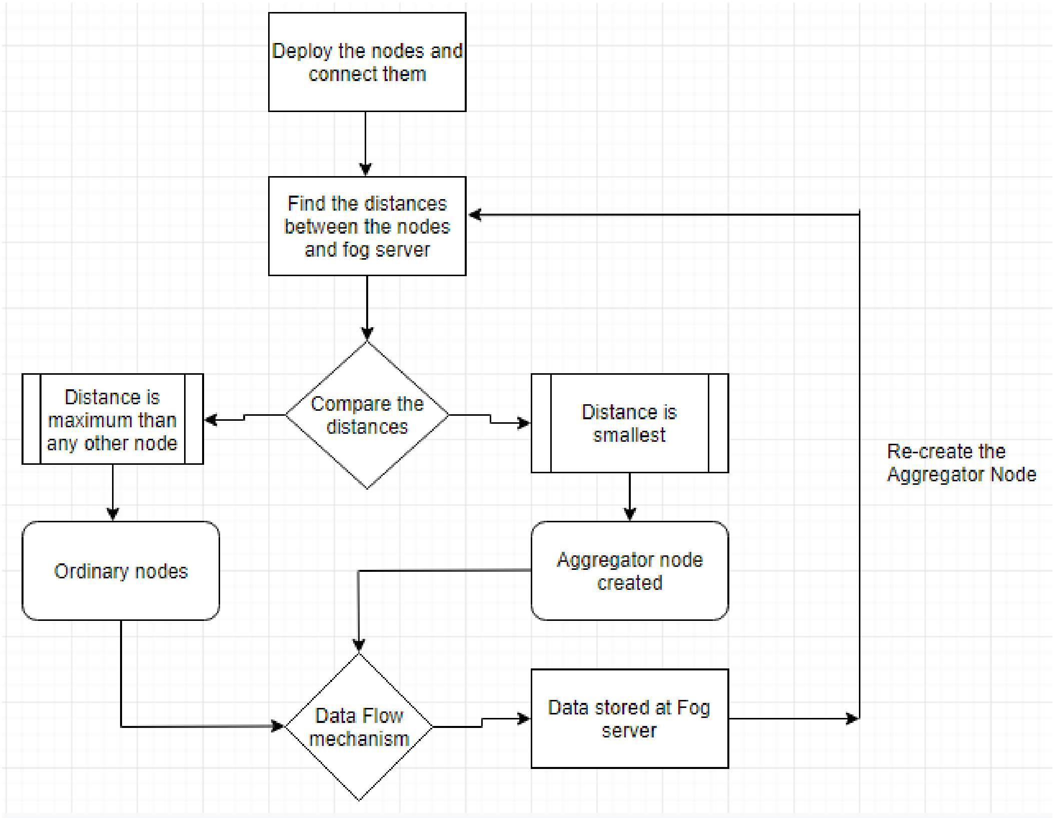
Deploy the nodes and connect them
Find the distances between the nodes and fog server
Compare the distances
Distance is maximum than any other node
Distance is smallest
Re-create the Aggregator Node
Ordinary nodes
Aggregator node created
Data Flow mechanism
Data stored at Fog server

Figure 3 Flow chart of proposed mechanism

Conclusion

In this paper, an efficient and secure data collection algorithm has been proposed. This proposed work comprises of creation of aggregator nodes and collection of data from the nodes within the cluster. The performance of proposed algorithm will be better than existing methods because after each transmission of data, the aggregator node keeps on updating. This is totally resolving the problem of failure of nodes due to heavy data exchange with the fog servers or base station. This proposed algorithm will helps in improving the energy efficiency, storage requirement and latency.

References

[1] P. Wang, F. Ye, and X. Chen, “A Smart Home Gateway Platform for Data Collection and Awareness,” *IEEE Commun. Mag.* , vol. 56, no. 9, pp. 87–93,2018.

[2] A. Orsino, G. Araniti, L. Militano, J. Alonso-Zarate, A. Molinaro, andA. Iera, “Energy efficient IoT data collection in smart cities exploiting D2D communications,” *Sensors (Switzerland),* vol. 16, no. 6, pp. 1–19, 2016.

[3] A. Brékine*et al.* , “Internet of Things Security and Data Protection,” pp. 81–92,2019.

[4] S. Balakrishna, M. Thirumaran, and V. K. Solanki, *A Handbook of Internet of Things in Biomedical and Cyber Physical System,* vol. 165. Springer International Publishing,2020.

[5] A. Ullah, G. Said, M. Sher, and H. Ning, “Fog-assisted secure healthcare data aggregation scheme in IoT-enabled WSN,” *Peer-to-Peer Netw. Appl.* , 2019.

[6] V. P. Alanis *et al.* ,*Smart Technology*, vol. 213. Springer International Publishing,2018.

[7] D. Ebrahimi, S. Sharafeddine, P. H. Ho, and C. Assi, “UAV-Aided projection-based compressive data gathering in wireless sensornetworks,” *IEEE Internet Things J.* , vol. 6, no. 2, pp. 1893–1905,2019.

[8] B. H. Çorak, F. Y. Okay, M. Güzel, Ş. Murt, and S.Ozdemir, “Comparative Analysis of IoT Communication Protocols,” *2018 Int. Symp. Networks, Comput. Commun. ISNCC 2018*, 2018.

[9] R. Gurunath, M. Agarwal, A. Nandi, and D. Samanta, “An overview: Security issue in IoT network,” *Proc. Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud), I-SMAC 2018*, pp. 104–107,2019.

[10] J. Zhang, P. Hu, and J. Long, “A hybrid transmission based data collection scheme with delay and reliability guaranteed for lossy WSNs,” *IEEE Access*, vol. 7, no. c, pp. 70474–70485,2019.

[11] G. Li *et al.* , “Energy efficient data collection in large-scale internet of things via computation offloading,” *IEEE Internet Things J.* , vol. 6, no. 3, pp. 4176–4187,2019.

[12] N. A. M. Alduais, I. Abdullah, and A. Jamil, “An EfficientData Collection Algorithm for Wearable / Mobile Tracking System in IoT /WSN,” *2018 Electr. Power, Electron. Commun. Control. Informatics Semin. EECCIS 2018*, pp. 250–254, 2019.

[13] C. T. Cheng, N. Ganganath, and K. Y. Fok, “Concurrent data collection trees for IoT applications,” *IEEE Trans. Ind. Informatics*, vol. 13, no. 2, pp. 793–799,2017.

[14] S. Cheng, Y. Li, Z. Tian, W. Cheng, and X. Cheng, “A model for integrating heterogeneous sensory data in IoT systems,” *Comput. Networks*, vol. 150, pp. 1–14,2019.

[15] Z. Guan *et al.* , “APPA: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT,” *J. Netw. Comput. Appl.* , vol. 125, pp. 82–92,2019.

[16] M. S. Hossain and G. Muhammad, “Cloud-assisted Industrial Internet of Things (IIoT) - Enabled framework for health monitoring,” *Comput. Networks*, vol. 101, pp. 192–202,2016.

[17] J. Jang, I. Y. Jung, and J. H. Park, “An effective handling of secure data stream in IoT,” *Appl. Soft Comput. J.* , vol. 68, pp. 811–820,2018.

[18] H. Huang, S. Guo, W. Liang, K. Wang, and A. Y. Zomaya, “Green Data-Collection from Geo-distributed IoT Networks through Low-Earth-Orbit Satellites,” *IEEE Trans. Green Commun. Netw.* , vol. PP, no. c, pp. 1–1, 2019.

[19] A. Liu, X. Liu, T. Wei, L. T. Yang, S. C. Rho, and A. Paul, “Distributed multi-representative re-fusion approach for heterogeneous sensingdata collection,” *ACM Trans. EmneddedComput. Syst.* , vol. 16, no. 3, p. 73, 2017.

[20] Y. N. Liu, Y. P. Wang, X. F. Wang, Z. Xia, and J. F. Xu, “Privacy- preserving raw data collection without a trusted authority for IoT,” *Comput. Networks*, vol. 148, pp. 340–348,2019.

[21] E. Luo, M. Z. A. Bhuiyan, G. Wang, M. A. Rahman, J. Wu, and M. Atiquzzaman, “PrivacyProtector: Privacy-Protected Patient DataCollection in IoT-Based Healthcare Systems,” *IEEE Commun. Mag.* , vol. 56, no. 2, pp. 163–168,2018.

[22] Z. Qin, D. Wu, Z. Xiao, B. Fu, and Z. Qin, “Modeling and Analysisof

Data Aggregation from Convergecast in Mobile Sensor Networks for Industrial IoT," *IEEE Trans. Ind. Informatics*, vol. 14, no. 10, pp. 4457– 4467, 2018.

[23] X. Xiang *et al.* , "Delay and energy-efficient data collection scheme-based matrix filling theory for dynamic traffic IoT," *Eurasip J. Wirel. Commun. Netw.* , vol. 2019, no. 1,2019.

[24] T. Rahman, X. Yao, and G. Tao, "Consistent Data Collection and Assortment in the Progression of Continuous Objects in IoT," *IEEE Access*, vol. 6, pp. 51875–51885,2018.

[25] J. Zhang, P. Hu, F. Xie, J. Long, and A. He, "An Energy Efficient and Reliable In-Network Data Aggregation Scheme for WSN," *IEEE Access*, vol. 6, no. c, pp. 71857–71870,2018.

[26] W. Wang, P. Xu, and L. T. Yang, "Secure data collection, storage, and access in cloud-assisted Iot," *IEEE Cloud Comput.* , vol. 5, no. 4, pp. 77– 88,2018.

[27] T. Hayajneh, A. N. Abdalla, M. Z. A. Bhuiyan, J. M. Zain, M. M. Hassan, and H. Tao, "Secured Data Collection with Hardware-based Ciphers for IoT-based Healthcare," *IEEE Internet Things J.* , vol. PP, no. X, pp. 1– 1,2018.

[28] F. Shi,U. Adeel,E. Theodoridis,M. Haghighi,andJ. Mccann,"OppNet: Enabling Citizen-Centric Urban IoT Data Collection Through Opportunistic Connectivity Service," pp. 723–728,2016.

[29] G. Manogaran, R. Varatharajan, D. Lopez, P. M. Kumar, R. Sundarasekar, and C. Thota, "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system," *Futur. Gener. Comput. Syst.* , vol. 82, pp. 375–387,2018.

[30] T. Li, S. Tian, A. Liu, H. Liu, and T. Pei, "DDSV: Optimizing Delay and Delivery Ratio for Multimedia Big Data Collection in Mobile Sensing Vehicles," *IEEE Internet Things J.* , vol. 5, no. 5, pp. 3474–3486,2018.