Model of Data Collection Control in the Internet of Things on the Basis of Social Network Technologies

N A Zhukova¹, A M Thaw² and E L Evnevich¹

¹St. Petersburg Research Center of the Russian Academy of Sciences, 14-th Linia, VI, No. 39, St. Petersburg, 199178, Russia
²Faculty of Software Engineering and Computer Systems, ITMO University, Kronverkskiy prospekt, 49, St. Petersburg, 197101, Russia

nazhukova@mail.ru

Abstract. The paper deals with the issues of data collection in the dynamic Internet of Things network. An approach proposed is based on the models and methods of social networks analysis along with clustering methods commonly used in this case. The developed hybrid model tracks contacts between nodes of the network thus creating “social relationships” for each node and estimates the number and the frequency of “social contacts” characterizing the degree of “friendships”. After fixing the contacts between nodes the model takes into consideration only the nodes connected by “friendship” relation. Data collection based on “social structure” provides higher security level and shows the enhancement as regards energy saving and delays of data transfer.

1. Introduction

Data collection in IoT (Internet of Things) systems is becoming increasingly problematic due to the transition from static nodes to multiple mobile nodes. Dynamic changes in the network topology result in the complexity of data collection and in the lack of reliability of connections between nodes.

Data collection technologies used in IoT systems could be given a brief review as: cloud computing technologies, fog computing technologies, middleware platforms, sensor network technologies, data encryption, social network analysis techniques (SNA) and data routing techniques. The cloud and Fog computing are focused on decentralized processing of raw data collected from the devices [1]. Various data analysis methods can be used for data processing, including machine learning methods [2]. Fog computing is carried out in close proximity to data sources of end devices. System performance enhancement is achieved by reducing the amount of data being transferred to the cloud for processing, analysis and long-term storage.

Sensor networks are essential components of smart environment. Basically, IoT sensors have low cost and energy. These devices are characterized by the such parameters: battery capacity and ease of deployment [3, 4]. Data being transferred from devices are subject to attacks. In order to ensure data security data encryption techniques are used to prevent unauthorized monitoring of data and access to them, of data modification during collection, processing and storage. Traditional cryptographic algorithms cannot be used to secure the collected and transferred data from devices with limited resources [5]. SNA technique assumes creation of new groups of devices, establishment of “social relations” between devices and control of the established relationships [6].
The problem of data security assurance in the mobile networks can only be treated and solved in close connection with the two other ones – high energy consumption and network latency. That is the solution of security problem shouldn’t result in the growth of amount of energy consumed and in the speed decrease of data transfer from devices to the server.

There are a number of models based on these technologies which address the challenges of data collection in the context of dynamic changes in the network. The study [7, 8] proposes a data collection model using “BeSmart” platform which provides the possibility to control the collection and processing of heterogeneous data coming from the nodes with limited resources. The model optimizes the balance between data accuracy and power consumption, but doesn’t take into consideration network latency.

The work [9] presents a linear model based on Fog computing. Fog nodes are distributed over the network. User devices are placed at the end of fog nodes sequence. The model can efficiently reduce power consumption but is not able to meet the requirements to security level and deal with data transfer delay challenges.

Data collection model based on node clustering is described in [10]. Data are transferred by the nodes to the cluster head (CH), by CHs to the fog nodes and finally to the cloud. Cloud is responsible for processing and storing enormous amount of data using high-end servers and data centers. According to the estimates of the model it provides effective solution of the latency and power consumption problem however it suffers from low security of the data collection and transfer as the model is used in a large number of real-time applications.

The above approaches to data collection in the dynamic networks of IoT are mainly used to meet certain separate requirements: network load, data routing, transfer delays, etc. Security issues are not paid sufficient attention to. The purpose of the present study is to identify and to apply appropriate social networking methods for data collection security assurance, reduction of energy consumption and data transfer delays in dynamic IoT networks. The research was conducted in the frames of further development of Social Internet of Things (SIoT) paradigm [11]: use of social networking methods in IoT resulting in enhanced capabilities of search, selection and interaction of objects and services.

The method proposed in the paper takes advantages of the SIoT network analysis methods, clustering methods and fog computing which combined ensure the trust and security of data collection between IoT devices.

2. Data Collection Security Model
Determining trust between devices is an effective way to improve the security of data collection in the IoT network. Using the “social contacts” of devices, their “friendship” and “similarity of interests”, the system establishes “social relationships” between devices and measures “social trust” using indicators of connectivity, honesty, privacy, etc.

2.1 Nodes degrees distribution
Networks arising in a natural way (social, communication, biological, etc.) could be adequately modeled by scale-free graphs. Cluster networks having a topology like a natural network-can therefore be represented by scale-free graphs as well. In a network with cluster structure the central elements – cluster heads – are assigned to serve as the intermediate for data transfer from the nodes to the server. So, devices establish connection with the corresponding cluster head for data transfer via it.

Each node $i$ in the SIoT is linked by relations of “friendship” with other nodes. To establish relations of “friendship” between nodes (devices) it is necessary to perform the following actions:
- determine the frequency of “interaction” (contacts) between nodes (devices),
- establish the maximum degree of “friendship” for each node (the degree of “friendship” is a number of connections that the node can establish with other nodes),
- collect data from all “friendship” nodes by using clustering methods.
Degree of node (i.e., number of links between a node and other nodes) could be considered as a number of packages received by that node. Let us consider one of the vertices of the graph – some node “i”. For this node “i” the input semi-degree $d-1$ is the number of edges of the graph entering the vertex (node “i”) and the output semi-degree $d-1$ is the number of edges of the graph issuing from the node “i”. The output semi-degree shows to how many nodes the node “i” can transmit data, and the input semi-degree determines the number of nodes from which the node “i” can receive data.

Maximum number of connections with other nodes should be determined in order to avoid defining malicious ones as the cluster heads. The probability of establishment of a new connection of some node with the cluster head “c” can be expressed by the formula.

$$P(i) = \left(1 - \frac{k_c}{k_{max}}\right) \frac{E_c k_c}{\sum_{c=1}^{N(t)} E_c k_c},$$

where $k_c$ – is the number of connections (links) of the cluster head; $k_{max}$ – maximum number of connections to any node in the cluster; $E_c$ – energy of the cluster head; $N(t)$ – number of cluster heads at time $t$; $E_i$ – energy of the node “i”; $k_i$ – number of connections of the node “i”. Larger value of $E_c k_c$ indicates the possibility of connecting a new node to the cluster since the cluster head “c” has an energy reserve. It follows from formula (1) that if the degree of the vertex – cluster head – reaches the maximum limit for the number of connections $k_{max}$, the probability that the new node can establish a connection with the cluster head becomes equal to zero. It could be seen that the nodes may have various probabilities of inclusion in the cluster depending on energy, number of connections with other nodes. It allows increases the trust and security of data collection with consideration of consumed energy.

2.2 Establishing network “social structure”

The use of social networking methods in IoT allows the objects to establish autonomously the so-called “social relations” [6]. For such networks the following relationship types could be defined:

- Ownership Object Relationship (OOR) is a type of relationship between objects that are owned by the same user,
- Social Object Relationship (SOR) – those are relationships that are established between objects which belong to friends and are in contact from time to time,
- Parental Object Relationship (POR) – set up between nodes that belong to homogeneous objects created by the same manufacturer (for example, products of the same company),
- Co-location Object Relationship (C-LOR) is a relationship between objects that are in the same location, etc.

By limiting “friendship” with other nodes up to a number of $k_{max}$, it becomes possible to control “social relations”. In accordance with formulas (1) all devices can establish "friendship" with the maximum number of nodes equal to $k_{max}$. The parameter $k_{max}$ also makes possible to limit the energy consumption of the network and enhance security by avoiding data collection from malicious nodes. The main idea of this method is to improve the process of collecting data from the devices due to supporting efficient interaction between the devices taking into consideration the scalability of the network. Thus, data collection in cluster networks can be performed only from a limited number (not exceeding $k_{max}$) of “friendships” nodes, which meet the security requirements to data collection in the IoT networks (figure1).

Each node controls its “friendship” requests taking into account one the following schemes:

- Scheme 1. Each node must reject the “add to friends list” request after reaching the number of links equal to $k_{max}$;
- Scheme 2. Each node must sort its “friends list” by decrease of degree of “friendship”. To maximize the number of “friends” the node must accept the first $k_{max}$ “friends”;
• Scheme 3. Each node must sort its “list of friends” by the degree of “friendship”. To minimize the number of friends the node must reject the first $k_{max}$ “friends”;
• Scheme 4. In order to maximize its own local cluster coefficient, the node sorts its friends list in decreasing order in terms of their common friends and rejects the nodes with the lowest number of common friends;
• Scheme 5. In order to minimize its own local cluster coefficient, the node sorts its friends in increasing order of the common friends. Rejects the nodes with the highest number of common friends.

The schemes help mobile nodes to control the “friendship list” thus forming the best set of $k_{max}$ “friends” and, hence, to choose the best way to collect data.

3. Simulations of data collection control
Simulation is carried out according to the above mentioned steps which are determining the frequency of “interaction” between nodes, establishing the maximum degree of “friendship” for each node and collecting data from all “friendship” nodes by using clustering methods.

3.1 Data collection control based on “friendship” relationships
“Friendship” connections between nodes are established at fog layer. SOR and OOR relations are determined for the data set of a smart city [12]. The total number of objects is 16 216, of which 14 600 – private devices; 1616 – public devices. Private devices are divided into two categories: mobile (smartphones, cars, etc.) and static (personal computers, printers, smart home sensors). Public devices are also divided into mobile (cars, taxis) and static (streetlights, sensors for parking and traffic monitoring). SORs are being established during "random interaction" between nodes and OOR relationships are established for the nodes (devices) belonging to the same user or organization. Therefore, OOR relationships are established for a short time, whereas the number of SOR relationships increases over time (figure 2).
It is assumed that algorithms based on fog computing collect the required data from nodes between which SOR are established. After creating friendships between devices, when the Fog service requests the data from a node, the required data is transmitted over the nodes which have established friendship relations with the requested node. Each node interacts with other nodes on two layers - the social layer and the data collection layer. The Fog service is an intermediate node in establishing “friendship” between devices. The “friendship” relationships help to establish trust relations between a limited number of IoT devices and to avoid collecting data from malicious nodes.

3.2 Simulation of “friendship” data collection based on clustering techniques

The simulation is carried out for mobile devices that suffer from power consumption problem. 200 nodes with the same initial energy 2J are located in the modeling area 300 m × 300 m. It is assumed that all nodes have equal computing resource and the maximum number of “friendships” $k_{\text{max}} = 50$. An identifier is assigned to each of 50 nodes of the same “social group” (node ID, current location). In one group maximum number of member nodes is 50. The list of the simulation parameters is given in Table 1.

Table 1. Simulation parameters of friendship data collection

| Parameters                          | Value        |
|-------------------------------------|--------------|
| 1 Number of mobile nodes            | 200          |
| 2 Simulation area                   | 300*300m²    |
| 3 Initial energy                    | 2J           |
| 4 Amplification energy for path-loss| 10 pJ / bit/ m² |
| 5 Amplification energy for multipath-loss | 0.0013 pJ / bit/ m³ |
| 6 Threshold distance                | 87.7 m       |
| 7 Energy consumption for data aggregation | 5 pJ / bit/ signal |
| 8 Data size                         | 500 bit      |
| 9 Selected friendship nodes         | 50           |
Figure 3 shows the node IDs in different groups. Due to the limitation of the maximum number of friendship nodes in each “social group” impossible to add more than $k_{\text{max}}$ nodes to each of them. The groups with a certain number of nodes are protected from addition of inaccessible or malicious node due to collecting data only from trusted nodes. It should be noted that the number of “friendships” remains the same at each round of data collection. Until a new “friendship” request is received each device can handle the behavior of other devices in its “friendship” group. Figure 4 shows a cluster head that collects data from 50 “friendship” nodes. At each round of data collection, the cluster head is reselected from “friendship” nodes. Figure 5 shows the total energy consumption in data collection rounds. The model is compared with the previously developed by the authors of LEACH-M algorithm [13] and demonstrates significant economy in energy consumption. Figure 6 shows the overall data transfer delay. The model is compared again with the modified LEACH-M algorithm and the model described in [14] and shows its advantage as regards to data transfer delays. Thus, simulation results confirm that the proposed data collection control model allows obtain significant gains in terms of consumed energy resources, enhances network response time and ensures higher level of security in comparison with the available models used in mobile IoT.

Fig. 3. Identiﬁers of devices from a certain “social group”.

Fig. 4. Paths of data collection between cluster head and selected “friendly” nodes.

Fig. 5. Total energy consumption.

Fig. 6. Total data transfer delay.
4. Conclusion

Novel model of data collection in IoT mobile network developed as a combination of clustering technologies, fog computing and a model of “social relationships” used in the analysis of social networks. By grouping nodes with as many “friendships” connections as possible the model helps to avoid the use of malicious nodes thus augmenting security and trust of data collection. Simulation results in comparison with those of traditional models demonstrated less power consumption and data transfer delays. Further development of the proposed model assumes using social network analysis methods to enhance the performance of data collection in IoT networks.

5. References

[1] Mostafa N 2020 A Dynamic approach for consistency service in cloud and fog environment Proc. 5th Int. Conf. on Fog and Mobile Edge Computing (FMEC) (Paris: Curran associates) 28–33
[2] Akinyelu A A 2019 Hybrid machine learning-based intelligent technique for improved big data analytics Proc. 6th Int. Conf. on Soft Computing & Machine Intelligence (ISCMII) (Johannesburg: IEEE) 7–11
[3] Da Cruz M A A, Rodrigues JJ P C, Al-Muhtadi J, Korotaev VV and de Albuquerque V HC 2018 A Reference model for Internet of Things middleware IEEE Internet of Things J. 5 (2) 871–883
[4] Fialho V and Fortes F 2019 Low power IoT network sensors optimization for smart cities applications Proc. Int. Conf. on Smart Energy Systems and Technologies (SEST) (Porto: IEEE) 1–6
[5] Raghav V and Raheja S 2018 Analysis of cryptographic algorithms for IoT security Proc. Int. Conf. on Advances in Computing, Communication Control and Networking (ICACCCN) (Greater Noida (UP): Curran Associates, Inc.) 16–20
[6] Tyagi V and Kumar A 2017 Internet of Things and social networks: A survey Proc. Int. Conf. on Computing, Communication and Automation (ICCNA) (Greater Noida: Curran Associates, Inc.) 1268–70
[7] Dong L and Li R 2019 RPL based named data routing protocol for low power and lossy wide area networks Proc. IEEE 5th World Forum on Internet of Things (WF-IoT) (Limerick: IEEE) 442–447
[8] Fotiou N, Siris V, Mertzianis A and Polyzos G 2018 Smart IoT data collection Proc. Symp. on a World of Wireless, Mobile and Multimedia Networks (WoWMoM) (Chania) 588–599
[9] Oma Ryuji et al. 2018 An energy-efficient model of fog and device nodes in IoT Proc. 32nd Int. Conf. on Advanced Information Networking and Applications Workshop (WAINA)(Cracow) 301–306
[10] Fattoum Marwa et al. 2019 A two-level clustering based on position, data correlation and residual energy in WSN. Proc. 15th Int. Wireless Communications & MobileComputing Conf. (IWCMC) (Tanger: Curran Associates, Inc.) 961–966
[11] Atzori Luigi et al. 2011 SIoT: Giving a social structure to the Internet of Things IEEE Communications Letters 15 (11) 1193–95
[12] Marche Claudio et al. 2018 A dataset for performance analysis of the Social Internet of Things Proc. IEEE 29th Ann. Int. Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC) (Istanbul: IEEE) 1–5
[13] Vodyaho A I, Ahmed A S, Zhukova N A and Thaw A M 2020 Cluster-oriented model for data collection in mobile IoT networks Proc. 9th Mediterranean Conf. on Embedded Computing (MECO: IEEE) (Budva) 1–6
[14] J. Voas, R. K. Kuhn, P. Laplante, and S. Applebaum 2018 Internet of Things (IoT) Trust Concerns Proc. NIST Cybersecurity White Paper.