A Cross-Layer Framework for Sensor Data Aggregation for IoT Applications in Smart Cities

Abrar Alkhamisi
Faculty of Computing and Information Technology
King Abdulaziz University
Kingdom of Saudi Arabia
e-mail: abrar200828@yahoo.com

Mohamed Saleem Haja Nazmudeen
Universiti Teknologi Brunei
Brunei Darussalam
e-mail: mohamed.saleem@utb.edu.bn

Seyed M. Buhari
Faculty of Computing and Information Technology
King Abdulaziz University
Kingdom of Saudi Arabia
email: mesbukary@kau.edu.sa

Abstract— **Many Internet of Things (IoT) applications like smart parking, waste management, and traffic congestion management, are being developed for smart cities. These applications make use of billions of sensors which in turn generates a huge amount of data that comes under the category of Big Data. For IoT/Smart-city applications to make use of these data efficiently there needs to be a proper framework through which the required sensor could be easily searched and made use of. The existing Extract-Transformation-Loading (ETL) tools and other search mechanisms for sensors assume there exist registries where the sensors can be searched for the desired criteria through ontologies or other suitable techniques. However, there has not been enough contribution to efficiently retrieve the sensor data and to make it available in the required format for the registries to search for. In this paper, we analyze a distributed cross-layer commit protocol (CLCP) for data aggregations and its support for query based search for IoT application.**

Keywords- Distributed, cross-layer, CLCP, wireless sensor, aggregation, IoT, Smart city

I. INTRODUCTION

The concept of IoT started with the advent of RFID technology and has grown rapidly with the support of other technologies such as wireless sensor networks (WSN), actuators, smartphones, social network and others [1]. The use of IoT has grown tremendously and is still growing in various sectors such as food processing, agriculture, smart parking, waste management and others [2], [3]. WSN network plays an important role in the growth of IoT as the hardware becomes cheaper, more powerful and has enhanced the battery life [4], [5], [6], [7]. These sensors are being widely deployed for various applications and they generate huge volume of data otherwise called as big data. The applications which utilize these big data need to analyze it properly to make use of the data efficiently. These applications require a middleware to process their requests with the SOA approach. In order to facilitate this, IoT architecture has been designed to have four layers as shown in Figure 1. Application layer is where all the IoT applications are running which can interact with the service layers otherwise called as the middleware. The applications can query the sensing layer through the middleware and obtain the data required for their applications. The service layer provides both device and platform independent interface for the applications to query the sensor data. The network layer deals with the interconnection of sensors and deals with efficient and suitable means of transferring the data from sensors in the sensing layer to data store in the service layer.

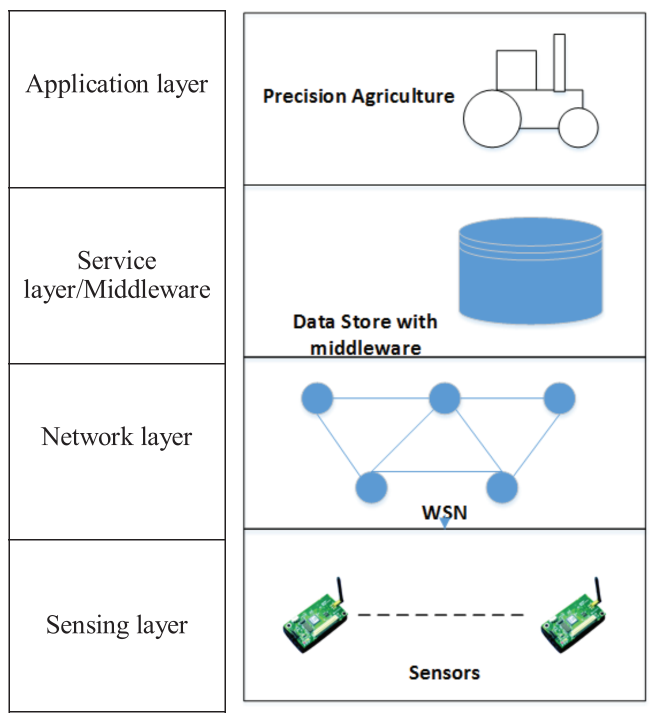


Fig. 1. IoT Architecture

For IoT/Smart-city applications to make use of these data efficiently there needs to be a proper framework through which the required sensor could be easily searched and made use of. The system level application that runs in the service layer to extract data from the sensors and store in the repositories is termed as Extract-Transformation-Loading

978-1-5090-1846-8/16/$31.00 ©2016 IEEE

(ETL) tool [8]. This and other similar tools assume there exist a standard port through which the sensors can be searched for the desired criteria through ontologies or other suitable techniques. Most of the research works in developing these tools focus on the application and the service layer only. However, there has not been enough contribution which deals with efficient retrieval of sensor data and to make it available to the ETL tools in the service layer for further operations. Moreover, sensor layer in IoT applications which makes use of WSN technology are constrained with energy supply due to battery capacity and communication bandwidth. There has been many research work done earlier to minimize the energy requirement and improving the throughput in WSN. Among other things data aggregation is one of the key approach utilized by many routing protocols to minimize the data transmission. Moreover, the query based data retrieval employed by the service layer requires in-network data aggregation to be performed for efficient operation. For efficient data aggregation according to the query generated by the service layer there needs to be cross-layer communication between the application layer and the network layer of the sensor nodes. The existing distributed cross-layer commit protocol (CLCP) [9] only makes use of the cross-layer communication to handle the network failure and not for query based data aggregation. In this paper, we would like to explore the suitability of CLCP in WSN for efficient query-based data aggregation that minimizes sensor energy consumption.

The rest of the paper is organized as follows, Section 2 discuss the related work, Section 3 discuss the design of our proposed framework, Section 4 elaborates on the simulation setup, Section 5 discuss the simulation results and we conclude the paper in Section 6.

II. Background and Related Work

In this section we briefly explain the existing approaches in data aggregation and cross-layer communication approaches employed so far in WSN. The data aggregation approach in sensor networks has been extensively studied before by many researchers [4], [10], [11]. Earlier studies show that energy required for transmitting a bit is as much as processing few thousands of instructions [12]. Therefore, data aggregation plays an important role in reducing the number of transmission for the WSN nodes. Most of the data aggregation in WSN is being performed in the application layer of the node. Moreover, the approach proposed in [11] is not efficient for query-based data retrieval methods that are used by most of the ETL systems for IoT applications. In this paper we propose to merge the data aggregation into the network layer and investigate the possibility of further reducing the energy consumption.

Among the cross-layer communications employed in energy conservation in sensor network we have come across research efforts on communication and coordination between Media Access Control (MAC) layer and the physical layer. However, between the application and the network layer there is only one attempt so far by authors in [13]. There they use filters to intercept the network packets for data aggregation. In this paper, our work focuses on a method that is based on a modified CLCP protocol. Our approach automatically identifies optimal WSN nodes in which data aggregation would take place for energy conservation.

Our earlier work on cross-layer communication [14],[15] focused on a similar approach in distributed service discovery in Peer-to-Peer networks, however such an approach was not studied in WSN.

III. Proposed Cross-Layer Approach for Data Aggregation

CLCP protocol which is being studied in this paper addresses the following problems in WSN data retrieval [9].

- Handle mobile ad-hoc environments without assuming any fixed infrastructure.
- Ensures failure tolerance.
- Operates both on application and network layer.

However, it does not deal with data aggregation problem. The query request is initiated in the application layer and operates mostly on the network layer. In our design we have adapted the CLCP protocol to have the aggregation feature with the cooperation of application and network layers throughout the protocol operation. In CLCP, by default cluster head selection process is carried out based on the CL_factor that considers two parameters which are residual energy and average distance of cluster members. The cluster member with the highest CL_factor is chosen as the cluster head of corresponding cluster. In our approach the cluster members are chosen based on the query reply which identifies the target sources node for a certain query generated by ETL registry. Figure 2 shows the layer in which our adapted CLCP operates. Our proposed approach is elaborated in the following steps which has two stages.

Stage 1: Aggregating node (AN) identification

Requirements:

1. Sensor nodes are capable of taking routing decisions based on application level messages.
2. Any query generated from ETL registry needs to be forwarded to root node otherwise called as sink node in WSN.

Steps:

1. The query message, constructed by encapsulating the query and its query id(automatic), is forwarded to the WSN network by the root node.
2. In the network, clusters are formed based on whichever nodes send the acknowledgement for the query and are marked as the member of the cluster for the specific query id.
3. Then the CL_factor is used within the cluster to select the AN node.

Stage 2: Data collection

1. The root node now sends the data collection signal to the target nodes only.
2. The target nodes start sending the data for a specific period as mentioned in the query.
3. As the data passes through the AN node, data aggregation takes place with message level intelligence at the network layer. In our approach the aggregation takes place at the network layer itself and there is no need to send the packet to the application layer. This process is shown in Figure 3.
4. Once the query period expires then the AN previously identified are removed and fresh cluster along with ANs will be formed for any new queries.

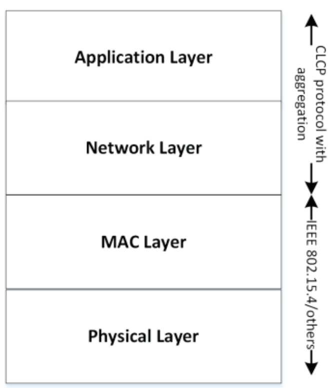
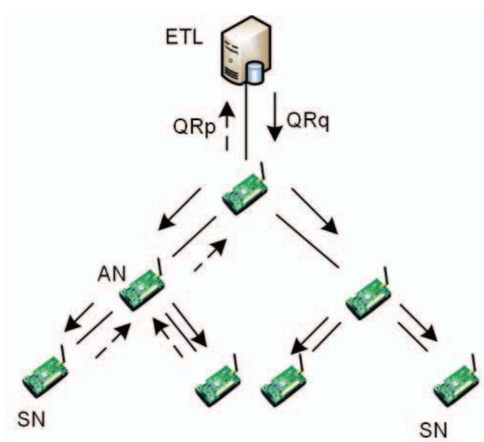


Fig. 2. Layered Architecture



QRq	Query request
QRp	Query reply
AN	Aggregating sensor node
SN	Non aggregating sensor node

Fig. 3. Network Architecture

Consider a scenario in which the ETL issues a query "SELECT temperature FROM sensor WHERE floor = 3 FOR 10 seconds" [13]. In this case our approach forwards the query and identifies the target source nodes. Then the cluster and AN is formed as explained in Stage 1 above. This AN node is specific for this query only. For a different query there could be different source nodes and therefore different ANs could be identified. The network layer identifies the query with the corresponding query id in cooperation with the application layer, thus using different ANs for different queries. The current approach can be improved by having a caching mechanism and we leave that to the future work as it is not included in the scope of this paper.

IV. Simulation

A. Cluster Cell formation in Network:

The network area is divided into cells based on network size in terms of height and width and communication range. Each cell is assigned with an ID and formed as a cluster. Each cluster have some cluster members to form the clustering. Then cluster member sends data to cluster head. After aggregating the received data, cluster head sends data to sink. Proposed Scheme cluster head selection method differs from existing scheme.

B. EECP (Energy Efficient Clustering Protocol) based cluster head selection and data aggregation:

Cluster head selection process is carried out based on Energy_factor that considers ratio of residual energy and initial energy of cluster members. The cluster member with the highest Energy_factor is chosen as the cluster Head of corresponding cluster. After cluster head selection, cluster members send all their data to their respective cluster head. Then, cluster head performs data aggregation. Further on, the aggregated data is sent from cluster head to sink via routers.

C. The proposed CLCP-based Cluster head selection and data aggregation:

Cluster head selection process is carried out based on the CL_factor that considers two parameters; residual energy and average distance of cluster members. The cluster member with the highest CL_factor is chosen as the cluster Head of corresponding cluster.

With Data Aggregation: After cluster head selection, all cluster members send data to the cluster head, which performs data aggregation. Then the aggregated data is sent from cluster head to sink via routers.

Without Data Aggregation: After cluster head selection, the cluster head forwards the data that it receives from cluster members, without data aggregation. Then the data is sent from cluster head to sink via routers.

NS2 simulator is used to simulate the environment and analyse the impact of using CLCP method. The nodes are distributed in an area of 600x600m. Different nodes are used to send data from a source to the sink node, numbered as 0. Same simulation scenario is used with CLCP-with-aggregation, CLCP-without-aggregation and EECP methods for comparison, varying number of rounds. Totally 5 simulation runs were carried out for the 5 scenarios of varying number of rounds (15 seconds each), named as 1, 2, 3, 4, and 5. Using cluster head, the nodes are grouped together to form cluster, as shown in Figure 4.

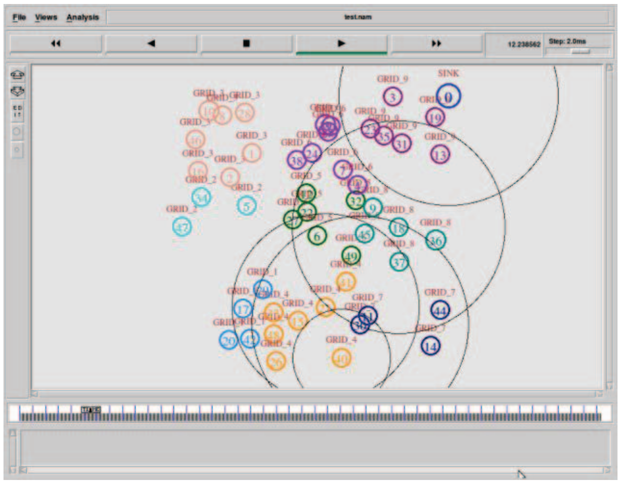


Fig. 4. Simulation Environment with cluster formation

TABLE I. SIMULATION SETUP DETAILS

SIMULATOR	Network Simulator 2.35
NUMBER OF NODES	50
AREA	600m x 600m
COMMUNICATION RANGE	200m
INTERFACE TYPE	Phy/WirelessPhy
MAC TYPE	IEEE 802.11
QUEUE TYPE	Droptail/Priority Queue
QUEUE LENGTH	50 Packets
ANTENNA TYPE	Omni Antenna
PROPAGATION TYPE	TwoRayGround
ROUTING AGENT	CLCP,EECP
TRANSPORT AGENT	UDP
APPLICATION AGENT	CBR
INITIAL ENERGY	50 Joules
SIMULATION TIME	100seconds

D. Actual Residual Energy

Figure 5 shows the impact of residual energy due to CLCP with and without aggregation and EECP methods. Varying the number of rounds, the residual energy drops by about 0.05J. This reduction in energy is due to usage of energy for data transmission in each round.

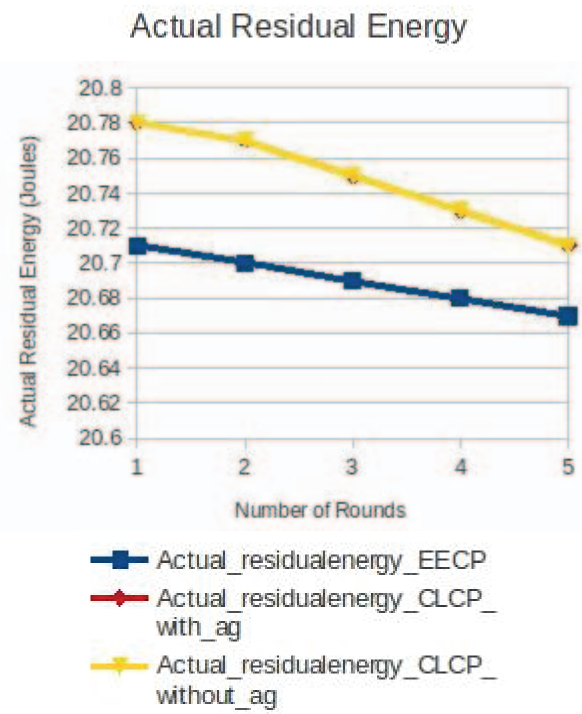


Fig. 5. Residual Energy

CLCP, with or without aggregation, achieves increased actual residual energy when compared to EECP.

E. Throughput

Figure 6 shows the impact of throughput due to different methods.

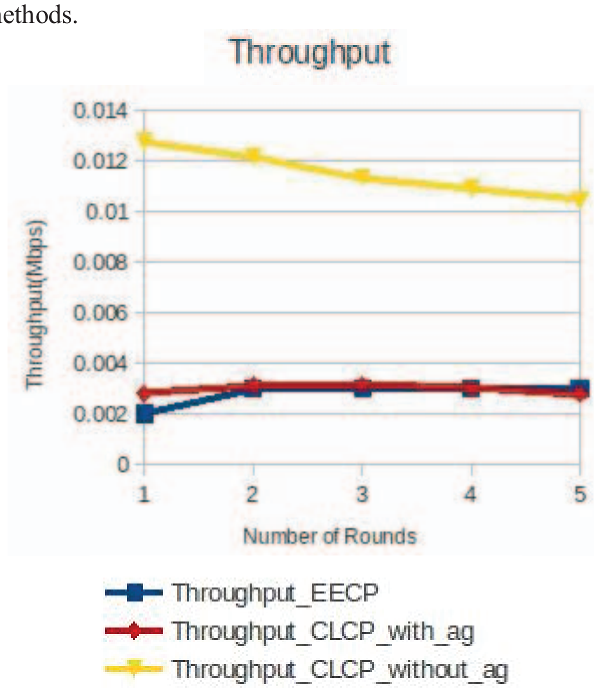


Fig. 6. Throughput

When number of rounds are increased throughput is reduced. CLCP without aggregation achieves better throughput when compared to CLCP with aggregation as well as the EECP methods.

F. *Delay*

Figure 7 shows the impact of delay due to CLCP with and without aggregation and EECP methods.

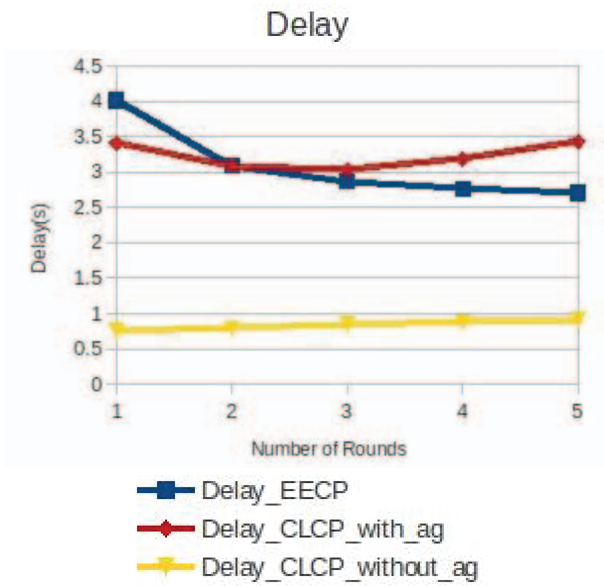
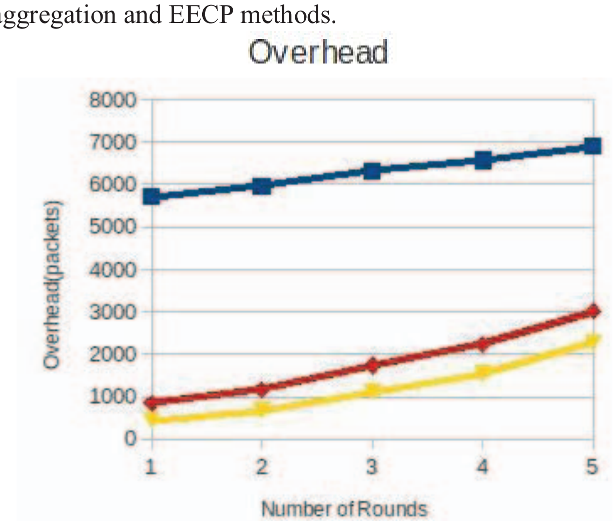


Fig. 7. Delay

When number of rounds are increased delay is increased and CLCP without aggregation achieves reduced delay when compared to CLCP with aggregation.

G. *Overhead*

Figure 8 shows the due to CLCP with and without aggregation and EECP methods.



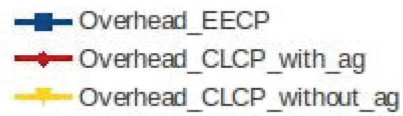


Fig. 8. Overhead

When number of rounds are increased, overhead is increased and CLCP achieves reduced overhead when compared to EECP.

V. Conclusion and Future work

Based on the given testing scenario, the CLCP method without aggregation outperforms other approaches in reducing overhead and increasing the throughput. But, the usage of residual energy is similar for both the CLCP approaches. Further study is to be performed with some realistically available smart meter dataset to test the outcome of this approach, along with multiple cluster groupings. Also we are planning to study the effect of multiple ANs within a cluster in the WSN.

VI. References

[1] S. Chen, H. Xu, D. Liu, B. Hu, and H. Wang, “A Vision of IoT: Applications, Challenges, and Opportunities With China Perspective,” *IEEE Internet Things J.*, vol. 1, no. 4, pp. 349–359, Aug. 2014.

[2] “50 Sensor Applications for a Smarter World,” 09-May-2016. [Online]. Available: http://www.libelium.com/top_50_iot_sensor_applications_ranking/.

[3] L. D. Xu, W. He, and S. Li, “Internet of Things in Industries: A Survey,” *IEEE Trans. Ind. Inform.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.

[4] J. Yick, B. Mukherjee, and D. Ghosal, “Wireless sensor network survey,” *Comput. Netw.*, vol. 52, no. 12, pp. 2292–2330, Aug. 2008.

[5] C. Perera, A. Zaslavsky, P. Christen, M. Compton, and D. Georgakopoulos, “Context-Aware Sensor Search, Selection and Ranking Model for Internet of Things Middleware,” in *2013 IEEE 14th International Conference on Mobile Data Management*, 2013, vol. 1, pp. 314–322.

[6] F. Viani, A. Polo, M. Donelli, and E. Giarola, “A Relocable and Resilient Distributed Measurement System for Electromagnetic Exposure Assessment,” *IEEE Sens. J.*, vol. 16, no. 11, pp. 4595–4604, Jun. 2016.

[7] M. Benedetti, L. Ioriatti, M. Martinelli, and F. Viani, “Wireless Sensor Network: A Pervasive Technology for Earth Observation,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 3, no. 4, pp. 488–496, Dec. 2010.

[8] M. Mesiti, S. Valtolina, L. Ferrari, M. S. Dao, and K. Zettsu, “An editable live ETL system for Ambient Intelligence environments,” in *Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on*, 2015, pp. 393–394.

[9] S. Obermeier, S. Böttcher, and D. Kleine, “CLCP #150; A Distributed Cross-Layer Commit Protocol for Mobile Ad Hoc Networks,” in *2008 IEEE International Symposium on Parallel and Distributed Processing with Applications*, 2008, pp. 361–370.

[10] S. Sahana and R. Amutha, “Energy Efficiency in Wireless Sensor Networks using Data Aggregation,” in *International Journal of Engineering Development and Research*, 2014, vol. 2.

[11] L. Krishnamachari, D. Estrin, and S. Wicker, “The impact of data aggregation in wireless sensor networks,” in *Distributed Computing*

Systems Workshops, 2002. Proceedings. 22nd International Conference on, 2002, pp. 575–578.

[12] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, “Energy conservation in wireless sensor networks: A survey,” *Ad Hoc Netw.*, vol. 7, no. 3, pp. 537–568, May 2009.

[13] Johannes Gehrke, “Query Processing in Sensor Networks,” *IEEE Pervasive Computing*, vol. 3, no. 1, pp. 46–55, 01-Jan-2004.

[14] A. B. Mohideen, M. Buhari, and H. M. Saleem, “Cost Efficient Proxy-Based Real-Time Streaming System,” 2012.

[15] M. Saleem, M. F. Hassan, and V. S. Asirvadam, “Modelling and simulation of underlay aware distributed service discovery,” in *Communications (APCC), 2011 17th Asia-Pacific Conference on*, 2011, pp. 417–421.