Empirical analysis of query-based data aggregation within WSN through Monte Carlo simulation

Sami J. Habib and Paulvanna N. Marimuthu
Computer Engineering Department, Kuwait University, Safat, Kuwait

Abstract

Purpose – Energy constraint is always a serious issue in wireless sensor networks, as the energy possessed by the sensors is limited and non-renewable. Data aggregation at intermediate base stations increases the lifespan of the sensors, whereby the sensors' data are aggregated before being communicated to the central server. This paper proposes a query-based aggregation within Monte Carlo simulator to explore the best and worst possible query orders to aggregate the sensors' data at the base stations. The proposed query-based aggregation model can help the network administrator to envisage the best query orders in improving the performance of the base stations under uncertain query ordering. Furthermore, it aims to examine the feasibility of the proposed model to engage simultaneous transmissions at the base station and also to derive a best-fit mathematical model to study the behavior of data aggregation with uncertain querying order.

Design/methodology/approach – The paper considers small and medium-sized wireless sensor networks comprised of randomly deployed sensors in a square arena. It formulates the query-based data aggregation problem as an uncertain ordering problem within Monte Carlo simulator, generating several thousands of uncertain orders to schedule the responses of M sensors at the base station within the specified time interval. For each selected time interval, the model finds the best possible querying order to aggregate the data with reduced idle time and with improved throughput. Furthermore, it extends the model to include multiple sensing parameters and multiple aggregating channels, thereby enabling the administrator to plan the capacity of its WSN according to specific time intervals known in advance.

Findings – The experimental results within Monte Carlo simulator demonstrate that the query-based aggregation scheme show a better trade-off in maximizing the aggregating efficiency and also reducing the average idle-time experienced by the individual sensor. The query-based aggregation model was tested for a WSN containing 25 sensors with single sensing parameter, transmitting data to a base station; moreover, the simulation results show continuous improvement in best-case performances from 56 percent to 96 percent in the time interval of 80 to 200 time units. Moreover, the query aggregation is extended to analyze the behavior of WSN with 50 sensors, sensing two environmental parameters and base station equipped with multiple channels, whereby it demonstrates a shorter aggregation time interval against single channel. The analysis of average waiting time of individual sensors in the generated uncertain querying order shows that the best-case scenario within a specified time interval showed a gain of 10 percent to 20 percent over the worst-case scenario, which reduces the total transmission time by around 50 percent.

Practical implications – The proposed query-based data aggregation model can be utilized to predict the non-deterministic real-time behavior of the wireless sensor network in response to the flooded queries by the base station.

Originality/value – This paper employs a novel framework to analyze all possible ordering of sensor responses to be aggregated at the base station within the stipulated aggregating time interval.

Keywords Wireless sensor network, Aggregation, Query-based, Scheduling, Empirical modelling, Simulation, Monte Carlo simulation

Paper type Research paper

International Journal of Pervasive Computing and Communications
Vol. 8 No. 4, 2012
pp. 329-343

1742-7371
DOI 10.1108/17427371211283010

1. Introduction

Energy constraint is a major issue in wireless sensor networks (WSNs), as the tiny sensors are designed with limited lifespan. Data aggregation, which integrates data from the various sensors deployed within a WSN aims to reduce number of transmissions, thereby reduces energy consumption of the network. The various data aggregation techniques in practice are centralized aggregation, in-network data aggregation, cluster-based approach, tree-based approach and query-based aggregation. In query-based aggregation, the user is facilitated with a choice to retrieve information of their interest from the given WSN by placing queries through the base station, whereby the sensing may be event based or periodical sensing.

In query-based data aggregation, the sensors respond to the flooded query from the base station by sending their data back to the base station utilizing the same channel. The worst-case scenario occurs, when all the sensors try to respond at the same time to the base station, thereby losing all data. This necessitates a scheduling scheme to be proposed to enable loss less data transmission. However, a detailed evaluation is appreciated to select a best scheduling order to complete the aggregation of entire sensors within the stipulated query time interval. Out of various simulators available to examine the performance of the system, Monte Carlo simulation, based on modelling stationary probabilistic situations may be best suited to evaluate the performance of the system that initiates the queries/requests.

In this paper, we have extended the query-based data aggregation within Monte Carlo simulator (Habib and Marimuthu, 2011) to empirically analyze the performance of query ordering within the stipulated time interval for a WSN operating with multiple sensing parameters and with multiple aggregating channels at the base station. We have viewed aggregation as a two-way data flow, which constitutes flooding of queries from the base station to the sensors and the response from sensors back to the base station to reflect the status of the area, which is monitored by them periodically. The possible random querying orders generated by the Monte Carlo simulator in scheduling the sensor responses to the base station are analyzed to figure out the best- and worst-case scenarios, wherein the analysis accounts the number of completed transmissions within the stipulated time interval and the waiting time elapsed between the start of the aggregation process and the beginning of transmission of individual sensor. For the given small and medium size WSN of 25 and 50 nodes, the simulation results demonstrate an improvement in the throughput of data aggregation with increased channels at the base station to aggregate the sensors' data. Hereby, the total time to aggregate sensors data at the base station decreases by 47, 64 and 72 percent, respectively, with two, three and four aggregating channels operating in parallel against a single channel. It is also noticed that the average waiting time of individual sensors in best-case scenario shows a gain of 10-20 percent over the worst-case scenario within the selected time interval. Furthermore, a best-fit mathematical model is developed for the best-case performance to demonstrate the behavior of the proposed aggregation algorithm.

The rest of the paper is organized into six sections. Section 2 discusses the related work. We present the system model in Section 3. Section 4 describes the proposed query-based data aggregation model within Monte Carlo simulator. The simulation results are summarized and analyzed in Section 5. Section 6 concludes our work.

2. Related work

Most of the recent research works focus on energy efficient data aggregation schemes within WSNs, mainly to increase the lifespan of the deployed sensors. In a study to analyze the impact of data aggregation in energy savings, the researchers investigated the influence of source-sink placements and also the density of the network in data aggregation (Krishnamachari *et al.*, 2002). A new reactive and energy-efficient scheme for reporting events in WSNs was proposed by Mousannif *et al.* (2011). In this work, the nodes detecting a certain event were organized themselves into a cluster, and they elected a cluster-head to collect data from the cluster members to aggregate and to forward the data to the mobile sink. In another attempt to minimize the energy in data gathering applications, the authors (Hwang *et al.*, 2008) adopted hierarchical grid structure to minimize the total energy consumption and utilized tree architecture to minimize delay cost for data gathering application.

A data aggregation framework presented as a middleware on WSNs is utilized to compare the performance of tiny aggregation (TAG) method with and without data aggregation (Patil and Patil, 2010). The authors utilized a query-based data aggregation and the simulation results showed reduced energy consumption during data aggregation through multi-hop transmission. Buttyan and Holczer (2010) proposed a private data aggregation protocol that utilized a private communication scheme to broadcast the queries to the sensors arranged in the form of a ring. In response to the queries of the base station, all the sensors within the ring added some noise as they received the token except the queried sensor that added noise and the data. The aggregators collected the data from the sensors without disclosing their identity.

In a query-based data aggregation scheme to aggregate the sensors' data deployed in an application-specific environment, the authors utilized a generic querying algorithm to generate a variety of queries to the source, and they utilized the information gradients presented in the network to build a global tree structure (Prakash *et al.*, 2009). Ahvar (2010) developed an energy and distance-aware query-based data aggregation scheme, named "EDQD," in which the aggregator was selected from the set of neighbor sensors witnessing an event. The simulations were carried out using Glomosim simulator. In an inter-query-based data aggregation scheme in WSNs, the researchers redesigned the data structure of the data packets that had information matching multiple queries (Zhang *et al.*, 2005). Li *et al.* (2010) proposed a cluster-based energy efficient distributed scheduling algorithm and claimed that the algorithm ended with more minimal latency and energy consumption than the data aggregation scheduling algorithms (DAS) employing minimum connected dominating set theory.

Su and Cho (2010) proposed a data aggregation method employing honeycomb cell structure for a range query in WSNs. The authors showed that because of intrinsic characteristics, the honeycomb structure selected the shorter data transmission path length. In addition, they showed that the spreading events could be overseen through an aggregation tree. A parameter-based data aggregation for statistical information aggregation in a WSN was discussed by Jiang *et al.* (2010). The aggregation algorithm tried to extract the statistical information from the sensor data with reduced communication cost. In a tiny aggregation service for aggregation in low-power, distributed, wireless environments, the users were allowed to use simple and declarative queries to perform in-network aggregation (Madden *et al.*, 2002).

In addition to query-based data aggregation algorithms, a query-aggregation algorithm was proposed by a team of researchers to aggregate the queries to reduce the number of duplicating or overlapping queries (Yu *et al.*, 2004). The authors claimed that the proposed aggregation scheme reduced the overall energy consumption and increased the lifespan of the sensors. In a method to provide secured data transmission from the sensors, the authors (Wilke *et al.*, 2009) checked the authenticity of the queries and the aggregated data that might be hacked through any of the nodes involved in in-network data aggregation. They devised a protocol named "authenticated query flooding protocol" (AQF) and combined it with extended secure aggregation for WSNs (ESAWN) to enhance network security.

In general, the wireless systems performance is analyzed through analytical methods, simulations, test-bed measurements and physical measurements. Most of the above researches on query-based data aggregation focused on the development of analytical methods. Furthermore, simulation tools are affordable in comparison to the test-bed measurements, and these tools are widely utilized in developing and testing new protocols. On comparing the available simulators for wireless networks such as NS-2, Glomosim, J-Sim and, etc. Monte Carlo simulator provides features that are needed for the data aggregation such as studying the behavior of real time systems with uncertain timings in querying order. Additionally, our work analyzed the possibility to provide energy efficient data aggregation by scheduling of sensors data to the aggregator in a best possible way.

In this paper, we have extended a query-based data aggregation model (Habib and Marimuthu, 2011) designed for sensors deployed in an application-specific environment to derive an empirical model in scheduling simultaneous data transmissions and to analyse the aggregating efficiency of the network with multiple sensing parameters and with multiple aggregating channels at the base station. The query-based aggregation model within Monte Carlo simulator explores many scenarios, whereby the sensor data, the number of data aggregation channels and the data aggregation time span are varied and for each scenario, the simulator estimates the performance of the base station by accounting the number of completed transmissions and the idle time associated with the transmission of responses within the stipulated time duration.

3. System model

3.1 Aggregation model

The proposed query-based aggregation model is shown in Figure 1, in which the sensor nodes respond to the query flooded by the base station at user specified rates. The query is characterized by the following parameters: the aggregation time interval "Time", and the start time of response for each sensor s_t corresponding to the generated uncertain querying order. The simulator generates various time intervals and for each interval, it generates M! uncertain ordering of queries to be scheduled to the base station, whereby M represents the number of sensors deployed within the WSN.

The query-based aggregation model shows the simultaneous requests flooded from the base station to the given set of sensors $S = \{S_1, S_2, \ldots S_M\}$ at time instance $\boldsymbol{t_i}$ and the responses from the sensors at their prearranged request time $\boldsymbol{t_{re(Si)}}$. The timing diagram for aggregation at the base station employing K channels ($K \ll M$) for the given M sensors is shown in Figure 2. The attributes S_i and S_{i+1} in Figure 2 represent ith and $i+1$th sensor responses. The starting of data transmission Tx_{St} for each sensor is represented by equation (1) and the ending of each transmission Tx_{Fin} of each sensor is calculated using equation (2), wherein the variable i represents the sensor number and Tx represents the total estimated transmission time of the selected sensor. The idle time is defined as the time elapsed between the starting time of first sensor in the generated random querying order to the starting time of selected sensor i, which is presented in equation (3):

$$Tx_{St}(i) = Tx_{Fin}(i-1) \tag{1}$$

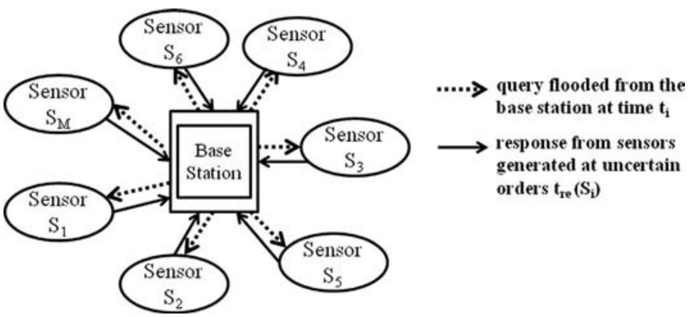
$$Tx_{Fin}(i) = Tx_{Fin}(i-1) + Tx(i) \tag{2}$$

$$Idle(i) = Tx_{St}(i) - Tx_{st}(1) \tag{3}$$

During the aggregation process, the total transmission time of the sensors with n sensing parameters is computed by summing up the individual transmission time taken by transmitting all the n sensing parameters. The multiple channels at the base station are simulated by enabling equal number of parallel transmissions.

3.2 Network model

We have considered a simple WSN comprises of a set of M sensors deployed randomly in the given area A and a base station located at the center of the area A, as shown in Figure 3. The sensors employed single hop routing model to directly send their data to the base station. The orders of tasks at each sensor are defined as sense, process and transmit and the base station queries all sensors at user specified rate to initiate



Note: M = number of sensors within WSN

Figure 1.
Aggregation model

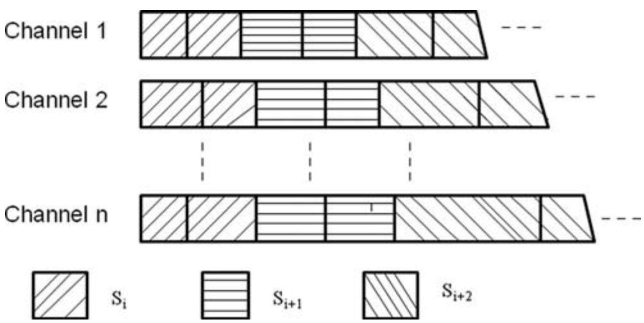


Figure 2.
Concurrent data aggregation at the base station

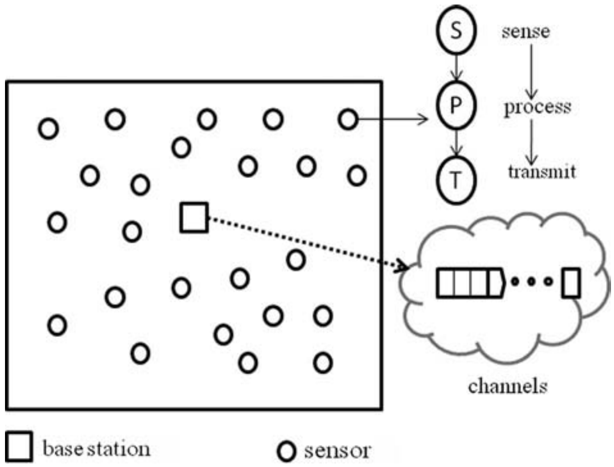


Figure 3.
WSN model

data transmission. It is assumed that the sensor nodes complete their sensing and processing tasks periodically and they are ready with their data for transmission. The periodical sensing nature forces the sensor to continuously sample/sense the data, thereby constrains related to the aggregator availability and storage space are added to avoid data loss during transmissions. The availability of the base station before starting the transmission should be verified, as the non-availability of the base station adds up additional data loss. Moreover, the non-availability of the base station for a long period forces the sensors to overwrite on previously sampled non-transmitted data stored in limited memory space.

The sensors are operating with a choice to sense single or double environmental parameters and the base station is devised with varying number of channels, which facilitates conflict-free reception of data from more than one sensor at a selected time instance. The increase in number of channels increases the number parallel transmissions within the stipulated time interval, thereby improves the performance and reduces the idle time (waiting time) of individual sensors.

4. Modeling query-based data aggregation within Monte Carlo simulator

We have extended the prior Monte Carlo simulator (Habib, 2006, 2008; Habib and Marimuthu, 2011) to explore the best possible way to schedule the sensors data with multiple sensing parameters and to offer conflict-free simultaneous transmissions within the WSN. In real environment, the unscheduled responses from the sensors to the query may lead to a worst-case scenario of losing all the data, whereby all M deployed sensors simultaneous flood the data to the base station causing the base station to receive none out of the M transmitted data.

We have considered periodical sensing and constrains are added in view of limited memory space and channel availability to prevent data loss during transmissions. Initially, the Monte Carlo simulation as shown in Figure 4, generates various uncertain ordering of queries to aggregate data at the base station operating with single channel. In this aggregation model, M is the number of sensors deployed within the WSN, and “Time” is an integer value specifying the duration of the allotted aggregation

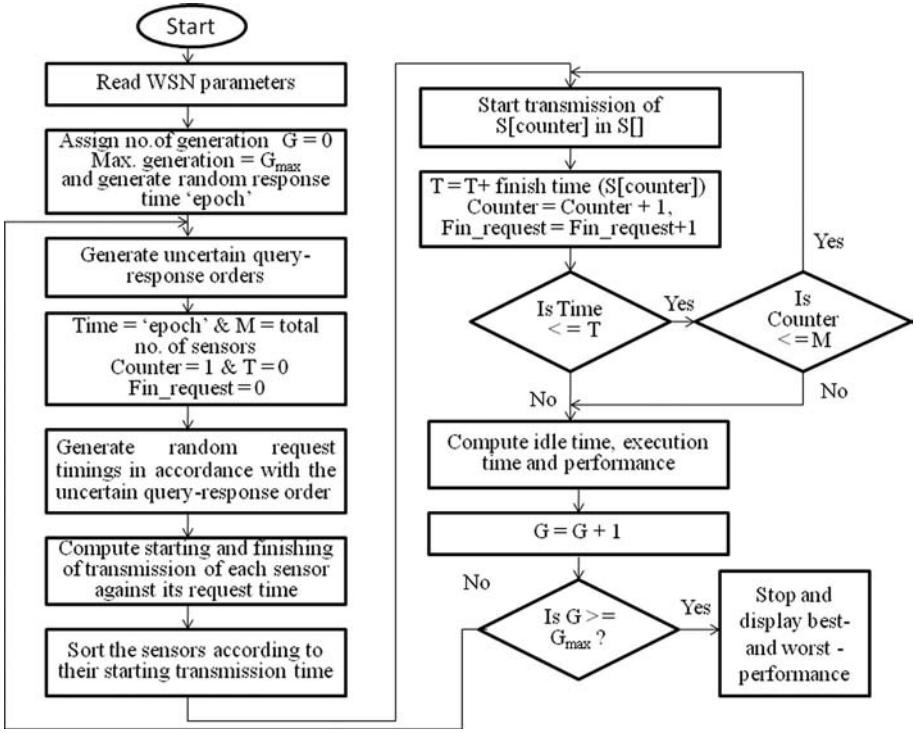


Figure 4.
Monte Carlo simulation
to model query-based
data aggregation

time interval. The parameter *S[ ]* specifies the transmission tasks timings of each of the sensors arranged in the order of uncertain request time generated by the Monte Carlo generator. The parameter *G* accounts the completed number of queries and G_{max} is an integer number representing the maximum number of generations for the given sensors. The parameter *Counter* is an integer number that is used as an index to the sensors in the sorted list. The Fin_request is an integer and it counts the number of finished transmissions within the stipulated time duration. The parameter *T* is a real number that keeps track of the total execution time of the sensors within the stipulated response period *Time*.

The inputs to the Monte Carlo simulator are the attributes of sensors (ID, coordinates, transmission task timings), the attributes of base station (ID, number of channels, location). The aggregation time interval "Time" and the number of uncertain order sets to be generated "G_{max}" are also keyed into the simulator. The simulation starts by initializing the parameters that are used within the experiment, such as the *T, Counter and Fin_request.*

The simulation generates a maximum of *M!* uncertain ordering of responses for the given *M* number of sensors and it assigns a random aggregation time interval "Time", which is a real number between zero and a maximum duration fed by the user. The successive request time of data transmission is computed in accordance with the finishing time of previous sensor in the sorted list S[] and a sort function is used to order the list of sensors in ascending order according to the request time.

The simulation algorithm tries to schedule all the sensors in the sorted list within the stipulated query time interval. The "Fin_request" updates the number of completed transmissions and the "Counter" updates the successive sensor from the sorted list to get ready for transmission. However, these parameters are incremented after the completion of each transmission within the stipulated time "Time". The performance of the base station, the aggregating efficiency is measured in terms of throughput of the base station, and is computed by accounting the number of completed transmissions within the given query time interval, which is presented in equation (4):

$$throughput = \frac{\text{number of sensors served by the base station}}{\text{Total number of sensors}} \tag{4}$$

Furthermore, for each generated querying order, the query-based aggregation algorithm computes the throughput and average idle time experienced by the individual sensors to list out the best- and worst-performances. The querying order with the least average idle time and maximum completed transmissions is selected as the best-case scenario; moreover, the querying order with maximum average idle time and minimum completed transmissions is selected as the worst-case scenario.

Additional parameters are introduced in Monte Carlo simulation in Figure 4 to track the number of completed transmissions with more aggregating channels at the base station. It is also ensured that the aggregation model takes care of the simultaneous scheduling with more than one channel, by allocating the next available sensor in the sorted list to the free channel.

5. Results and discussion

We have modelled the query-based aggregation within Monte Carlo simulator in view of handling uncertain transmissions in real environment and analyzing the possible querying orders for the given set of sensors to facilitate energy efficient transmissions. Our experimental setup is comprised of two WSN scenarios, whereby the number of sensors, the number of base station channels and the number of sensing parameters are varied. First set of experiments are conducted on a WSN having 25 sensors intended to sense single environmental parameter and aggregating the sensors' responses at the base station with single channel. The second set of experiments are carried out on a WSN with 50 sensors, sensing two different environmental parameters and aggregating data at a single base station possessing multiple channels. The query-based aggregation model embedded within a Monte Carlo simulator generates a maximum of *M!* (M factorial) uncertain querying orders for the given set of *M* sensors. However, we have restricted the number of generations to go up to few thousands to save run time in a standalone personal computer. For each uncertain querying order of sensors, our simulator computes the best-, worst-throughput of the base station and the average idle time experienced by the individual sensors during the data transmissions. The performance of the base station is estimated using equation (4), for each generated uncertain querying order with in the specified time interval "Time". The time interval "Time" is defined as the specific time allotted for the base station to aggregate the sensors data. We have carried out separate experiments in each of the proposed data aggregation schemes and the time interval is varied until the base station performance reaches the maximum.

The Monte Carlo simulator starts the aggregation process by generating an uncertain querying order and then, the sensors are sorted out based on their order of response to the base station to compute the starting time of transmission, whereby the starting time depends on the completion time of previous sensor.

5.1 Performance evaluation of 25 sensors

We have started with a small size sensor network comprised of 25 sensor nodes randomly deployed in a specified square area of $M \times M\,m^2$, projected to sense a single parameter and with the base station placed at the center of the selected square area to cover all the sensor nodes as shown in Figure 3. The time interval "Time" is varied from 80 time units to 200 time units to analyze the performance of the base station for various scheduling order generated by the simulator. The transmission time of each sensor is varied with their distance to base station. Our simulator lists the best-case and worse-case performance for each uncertain querying order set corresponding to the time interval "Time".

The best and worst-case performance for the given data set for varying time interval is shown in Figure 5, whereby the data aggregation follows a scheme of non-overlapping aggregation of sensors data. The response of the successive sensors is communicated after the completion of the previous sensor's transmission, thereby engaging the base station channel continuously. For the generated response-time "Time", the best-case performance varies from 56 to 96 percent and the worst-case performance at the base station varies from 36 to 96 percent.

5.2 Performance evaluation of 50 sensors

The second set of experiments are carried out on a medium size sensor network comprised of 50 sensor nodes deployed randomly in a specified square area of $M \times M\,m^2$, to sense two environmental parameters namely pressure and temperature. The lengths of two data packets resulting from two sensing parameters are varied to

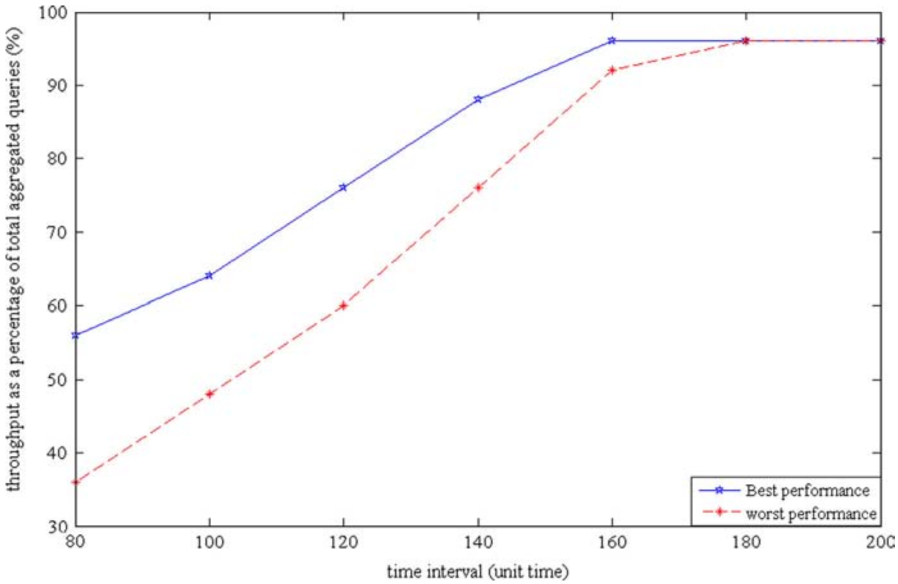


Figure 5.
Performance evaluation at the base station for 25 sensors

differentiate the nature of the data. The transmission time to base station from respective sensor depends on the Euclidian distance between the sensor and the base station and length of the data. We have carried out various experiments to explore the best- and worst-case scenarios in data aggregation at the base station, which queries the sensors to collect the data. For each uncertain ordering set generated within the stipulated time interval, our simulator computes the throughput as a percentage of total aggregated queries, average idle time experienced by the individual sensors and average execution time of the given system, as briefed in Section 4.

Our simulator varies the aggregation time interval from 100 time units to 900 units to allow the performance to reach around 100 percent. For each time interval “Time”, our simulator generates thousands of uncertain querying orders and computes the performance, idle time and execution time in scheduling the data to the base station. Finally, it compares the performance of the entire sensor sets, and selects the set with higher performance and lower idle time as the best order of aggregation. Figure 6 shows the best outcome of the entire uncertain sets generated within the Monte Carlo simulator.

The experiments are repeated by varying the number of aggregation channels at the base station by two, three and four. The performance of each case is recorded for the time interval starting from 100 time units up to the time during which the sensor’s performance is closer to 100 percent. Figures 7-9 show the best-case performance against various time intervals for the selected channels. It is observed from Figure 7 that both the plots generated for base station with single and two channels obey quadratic law. However, doubling of transmission channel facilitates maximum performance to be reached within 50 percent time period of single channel scenario. Being given the quadratic fitting equation and the coefficients, the performance as a percentage of total completed queries of the respective sensor network for any time interval t could be easily estimated by substituting the value of t.

The behavior of the performances follows a cubic fit with the selection of three and four aggregation channels and it shows still shorter time span to finish entire transmission as shown in Figures 8 and 9. The query-time decreases by 47, 64 and 72 percent, respectively, for base station with two, three and four channels against single channel.

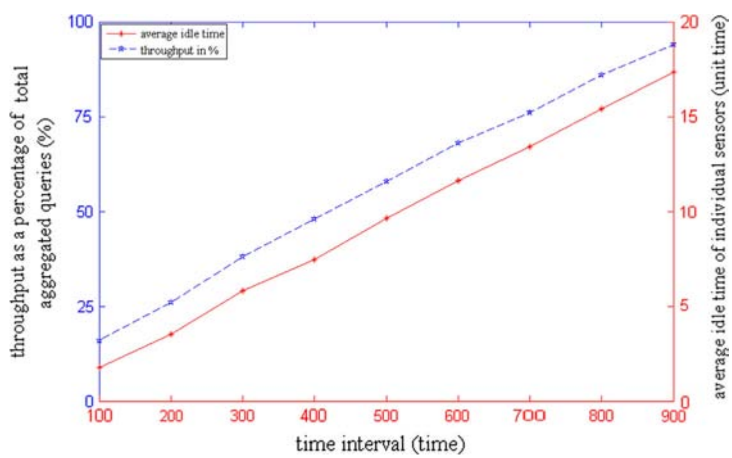


Figure 6.
Performance of the base station with single channel for a WSN of 50 sensors

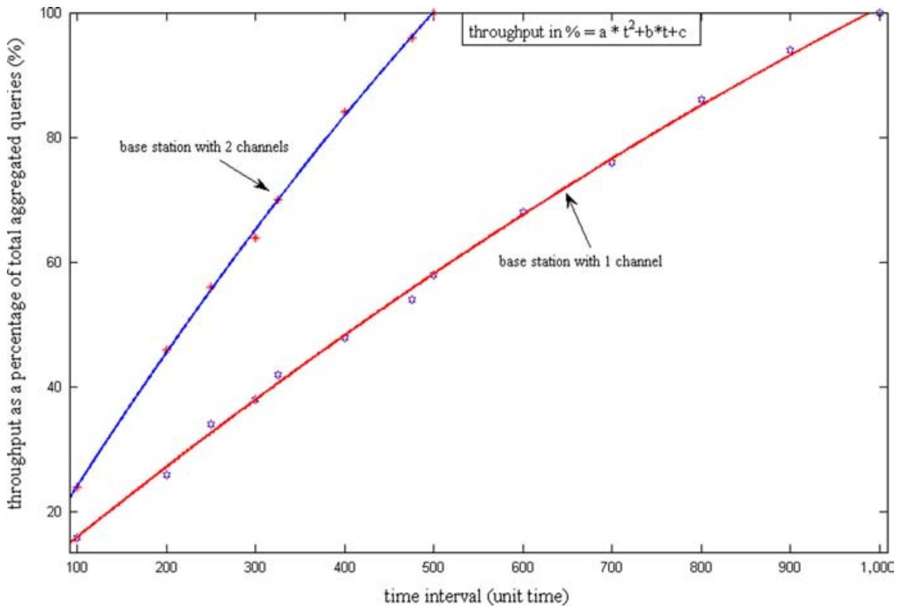


Figure 7.
Best-case performance for base station with single and two channels

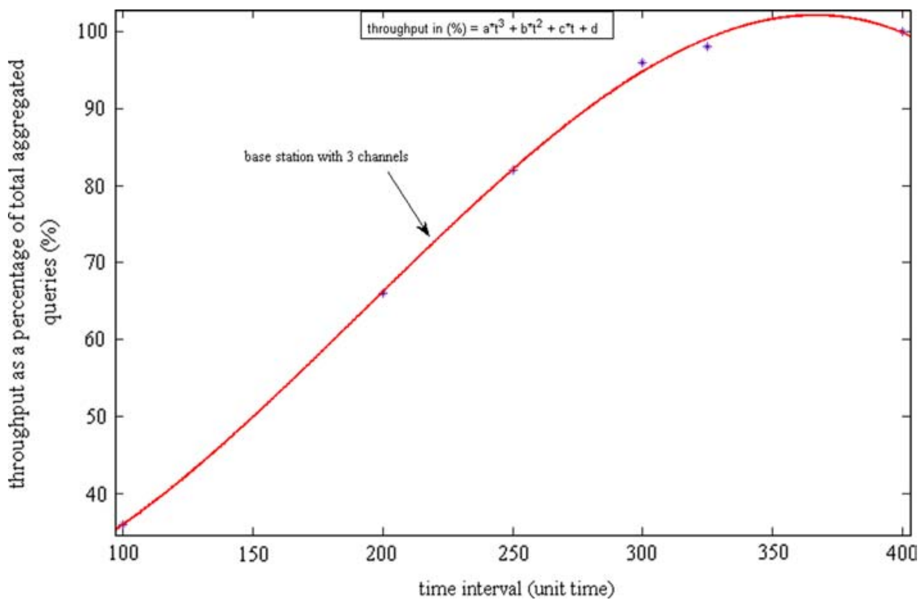


Figure 8.
Best-case performance for base station with three channels in parallel

The worst-case scenario corresponding to the stipulated time interval in all of the above experiments is shown in Figure 10, whereby a linear behavior is observed in all the scenarios. The difference in worst-case performance with the corresponding best-case performance is recorded as 8, 16 and 27 percent, respectively, for base station operating with two, three and four channels.

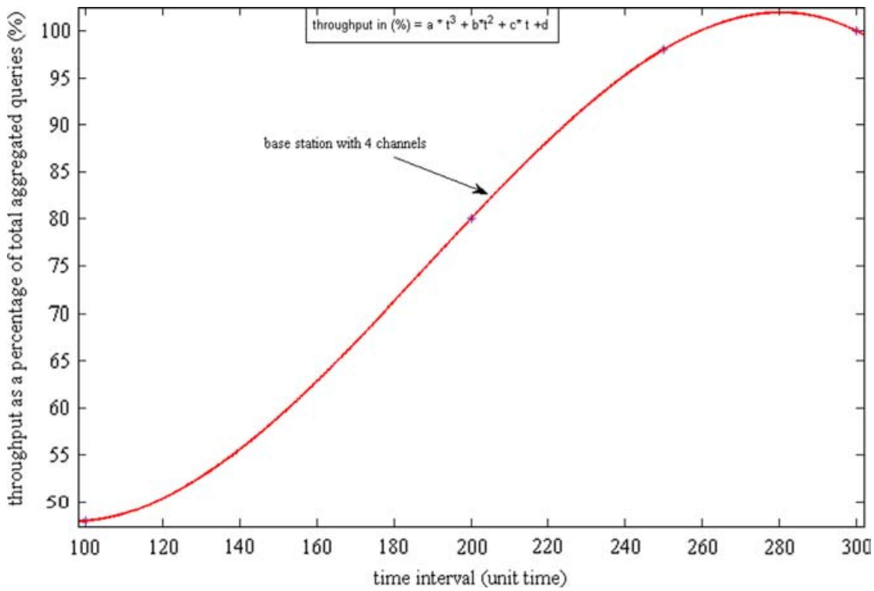


Figure 9.
Best-case performance for base station with four channels operating in parallel

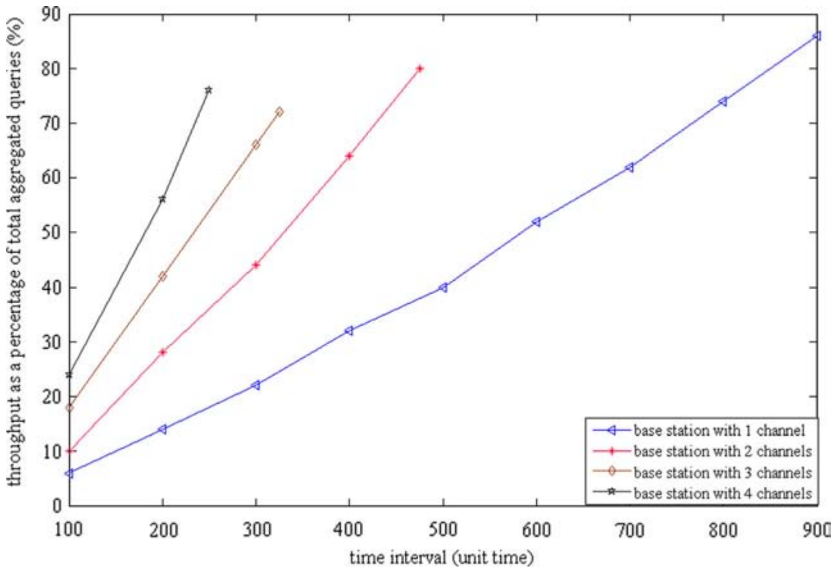


Figure 10.
Worst-case performance for base station with varying aggregating channels

The idle time is also found to be reduced with increasing channels, as it increases the number of simultaneous transmissions. The idle time as shown in Figure 11, demonstrate a decrease of 48, 66 and 75 percent, respectively, for base station operating with two, three and four channels compared to single channel. The maximum of average idle time plotted against the varying channels shows an exponential fitting as represented in equation (5), where *a, b* and *c* are the constants and *ch* represents the

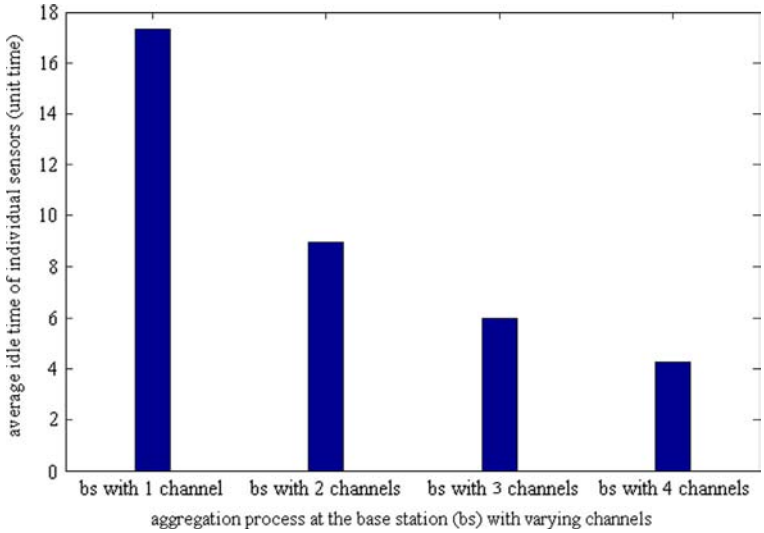


Figure 11.
Average idle time with varying channels at the base station

number of channels aggregating the sensors data in parallel. We have observed a 10-20 percent reduction in the idle time between the best-case and worst-case ordering within the generated uncertain order of sensors for the given time interval "Time":

$$idle\_time = a*ch^{b} + c \tag{5}$$

The above experimental results demonstrate that the query-based aggregation within Monte Carlo simulator can be utilized for choosing best scheduling querying order to provide energy efficient transmissions in real time environment, where the order of data aggregation queries is uncertain. The querying order selected as best-case showed an overall reduction of around 48-60 percent in transmission time than the worst-case performance. Moreover, for a given querying time interval, the aggregating efficiency of the base station could be improved by selecting optimum number of channels and best aggregating order of sensors. The analysis may be done offline to observe the behavior of the network, which may help the system administrator to decide the number of channels based on the quickness of the sensor response.

6. Conclusion and future work

The proposed query-based data aggregation within Monte Carlo simulator is employed to study the uncertain behavior of real time data transmission within WSNs. The aggregation selects the best scheduling order from thousands of randomly generated querying orders for a specified aggregation time interval and derives an empirical relationship from the behavior of the given sensors following a non-overlapping data aggregation scheme at the base station. The Monte Carlo simulator evaluates the performance of the base station by accounting the completed transmissions against the total deployed sensors and outputs the best- and worst-case querying order. The aggregation model tested for a WSN with 25 sensors, sensing single environmental parameter, showed a best-case scenario of performance ranging from 56 to 96 percent in the duration time of 80-200 time units, whereby the performance is the throughput presented as percentage of total completed tasks. The model analyzed the behavior of WSN with 50 sensors, sensing two environmental parameters and equipped with multiple channels at the base station to demonstrate the feasibility in aggregating simultaneous transmissions. In comparison with a single aggregating channel, the increase in number of channels by two, three and four decreases the total response-time by 47, 64 and 72 percent, respectively. Additionally, the average waiting time of individual sensors, in the generated uncertain querying order demonstrates a gain of 10-20 percent by the best-case scenario over the worst-case scenario within the specified time interval, which helps the system administrator to select the best aggregation ordering to produce better performance.

References

Ahvar, E. (2010), "EDQD: an energy-distance aware query-based data aggregation technique for wireless sensor networks", *Proceedings of International Workshop on Data Management for Wireless and Pervasive Communications, Perth, Australia, 13-20 April.*

Buttyan, L. and Holczer, T. (2010), "Perfectly anonymous data aggregation in wireless sensor networks", *Proceedings of IEEE 7th International Conference on Mobile Adhoc and Sensor Systems, San Francisco, CA, USA, 8-12 November*.

Habib, S.J. (2006), "A Monte Carlo simulator for evaluating server placement within network topology designs", *Proceedings of First International Conference on Performance Evaluation Methodologies and Tools, Pisa, Italy, 11-13 October*.

Habib, S.J. (2008), "Dynamic evaluation of server placement with a network design tool by using an embedded Monte Carlo simulator", *International Journal of Business Data Communications and Networking*, Vol. 4 No. 2, pp. 38-49.

Habib, S.J. and Marimuthu, P.N. (2011), "Query based data aggregation within WSN through Monte Carlo simulation", *Proceedings of the 9th International Conference on Advances in Mobile Computing and Multimedia, Ho Chi Minh City, Vietnam, 5-7 December*, pp. 162-8.

Hwang, S., Lu, K., Chang, T. and Dow, C. (2008), "Hierarchical data gathering schemes in wireless sensor networks", *International Journal of Pervasive Computing and Communications*, Vol. 4 No. 3, pp. 299-321.

Jiang, H., Jin, S. and Wang, C. (2010), "Parameter based data aggregation for statistical information extraction in wireless sensor networks", *IEEE Transactions on Vehicular Technology*, Vol. 59 No. 8, pp. 3992-4001.

Krishnamachari, B., Estrin, D. and Wicker, S. (2002), "The impact of data aggregation in wireless sensor networks", *Proceedings of International Workshop on Distributed Event Based Systems, Vienna, Austria, 2-3 July*.

Li, Y., Guo, L. and Prasad, S.K. (2010), "An energy-efficient distributed algorithm for minimum-latency aggregation scheduling in wireless sensor networks", *Proceedings of International Conference on Distributed Computing Systems, Genova, Italy, 21-25 June*.

Madden, S., Franklin, M.J., Hellerstein, J. and Hong, W. (2002), "TAG: a tiny aggregation service for ad-hoc sensor networks", *Proceedings of the Fifth Symposium on Operating Systems Design and Implementation, Boston, MA, USA, 9-11 December*.

Mousannif, H., Moatassime, H. and Rakrak, S. (2011), "An energy-efficient scheme for reporting events over WSNs", *International Journal of Pervasive Computing and Communications*, Vol. 7 No. 1, pp. 44-59.

Patil, N.S. and Patil, P.R. (2010), "Data aggregation in wireless sensor network", *Proceedings of IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India, 28-29 December*.

Prakash, R., Nourbakhsh, E. and Sahu, K. (2009), "Data aggregation in sensor networks: no more a slave to routing", *Proceedings of the Forty-Seventh Annual Allerton Conference on Communication, Control, and Computing, Allerton House, UIUC, Monticello, IL, USA, 30 September-2 October*.

Su, W. and Cho, G. (2010), "Honeycomb-based data aggregation for range query in WSN", *Proceedings of International Conference on Communications and Mobile Computing, Shenzhen, China, 12-14 April*, Vol. 3, pp. 177-81.

Wilke, J., Benenson, Z., Zitterbart, M. and Freiling, F.C. (2009), "Authentic query dissemination and data aggregation in WSN", *Proceedings of Sixth International Conference on Networked Sensing Systems, Pittsburgh, PA, USA, 17-19 June*.

Yu, W., Le, T., Xuan, D. and Zhao, W. (2004), "Query aggregation for providing efficient data services in sensor networks", *Proceedings of IEEE International Conference on Mobile Ad-hoc and Sensor Systems, Fort Lauderdale, Florida, USA, 25-27 October*, pp. 31-40.

Zhang, X., Yu, X. and Chen, X. (2005), "Inter-query data aggregation in wireless sensor networks", *Proceedings of International Conference on Wireless Communications, Networking and Mobile Computing, Wuhan, China*.

Further reading

Wang, P., Li, C. and Zheng, J. (2007), "Distributed data aggregation using clustered Slepian-Wolf coding in wireless sensor networks", *Proceedings of IEEE International Conference on Communications, Glasgow, Scotland, 24-28 June*, pp. 3616-22.

Corresponding author

Sami J. Habib can be contacted at: sami.habib@ku.edu.kw