The Effect of Data Aggregation on the Performance of a Wireless Sensor Network Employing a Polling Based Data Collecting Technique

Mohammad M. Abdellatif #1, José Manuel Oliveira *2, Manuel Ricardo #3

 # INESC TEC, Faculdade de Engenharia, Universidade do Porto
* INESC TEC, Faculdade de Economia, Universidade do Porto
Rua Dr. Roberto Frias, 378, 4200-465, Porto, Portugal
{1mma,2jmo,3mricardo}@inescporto.pt

Abstract—**Wireless Sensor Networks (WSNs) consist of small devices capable of sensing various variables in the environment, process and communicate them through the network. These devices interact together to carry out monitoring tasks. A photo-voltaic (PV) power plant is an example of such network, where each solar panel has a sensor connected to it. The number of interconnected solar panels can become very large in order to cover a large area. Each sensor senses the output of the panel and sends this value to a central node for processing.**

In this paper we study and compare the performance of a multi-hop wireless sensor network, employing a polling based data collecting technique with data aggregation, against the performance of a one hop network employing two different data collecting techniques. The study considers a wireless network with fixed number of nodes using different values of the offered load, estimating the network throughput for each technique and offered load. The use of a multi-hop setup was chosen in order to reduce transmission power and interference among nodes. Results show that the multi-hop network, using the polling based data collecting technique with data aggregation, performs close to the one hop network using the other two techniques. The study involves both simulation and testbed experimentation.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are made of small devices capable of sensing various variables in the environment, process these variables and communicate them through the network [1]. WSN applications are various including smart-cities, environmental monitoring, distributed sensing in industrial plants, and health care [2].

Usually in WSNs, low data rate is employed. However, other challenging issues related to the reliability of the communication links and to the efficient use of batteries also appear in the network [3]. Also, because of the many-to-one feature of the data collection applications in WSNs, wireless interferences and collisions, and the huge sizes of these networks, the scheduling of data transmissions becomes a challenging problem which needs to be carefully addressed. Another aspect that needs consideration is the data collection technique, i.e., the way sensors send their data to the sink. Each node can send its data to the sink as soon as it is ready, or nodes can wait for a request from the sink before sending their data. As the size of the network grows, the selected data collecting technique will have a huge impact on the performance of the network as a whole.

This work studies how the data collection technique used to gather information in a well defined scenario affects the throughput of a WSN. The communications scenario is provided by the SELF-PVP project [4]. This project aims to increase the efficiency of a photo voltaic (PV) power plant. It assumes that 200,000 solar panels are deployed in a matrix like deployment spread over a large area (250 hectares). The objective is to put the solar plant operating at an optimum power delivery point. Each solar panel will have a sensor/actuator that will sense local variables, communicate these values with other sensors and compute local errors with the goal of optimizing the overall performance of the panels' array.

A WSN in this scenario can be viewed as a grid of sensors where each sensor is connected to a solar panel. Panels in the same column will have the same value of the current passing through them. In order to reach the optimum power delivery point, most of the communications will be among nodes in the same column. Based on this, we divide the network into small networks of columns. Throughout this paper, we focus on the communication within a column.

This paper is a continuity of the work done in [5] where the performance of an IEEE 802.15.4 wireless sensor network employing three different data collecting techniques within a column was evaluated. In Technique 1, client nodes send their data as soon as it is ready. In Technique 2, the sink sends a broadcast poll message to its neighbors, which is propagated through the network triggering each client node to send its data to the sink. Technique 3 is similar to Technique 2 except that the poll message is sent to half of the network first and then the sink waits to receive the data from this half before sending the poll message to the other half.

In this paper, we evaluate the performance of Technique 3 with data aggregation in a multi-hop network. Since the information coming from the nodes is similar and headed to the same destination, it is reasonable to assume that data aggregation will improve the performance of the network. The performance is then compared against the performance of a one hop network with no data aggregation. Also, the functionality of a sink command message was enabled. This gives the sink node the ability to communicate with the client nodes after each

cycle in order to send different commands that might be used by the clients. The study considers a fixed size network, with different values of the offered load, and estimates the network throughput for each value of the offered load. The evaluation of the performance of the three data collection techniques was carried out using the Contiki COOJA simulator [6] which can be downloaded from [7].

The main contribution of this paper is the evaluation of the effect of data aggregation on the performance of Technique 3 with and without the sink command message. This evaluation enabled us to conclude that Technique 3 with data aggregation over a multi-hop network performs close to the other two techniques in a one hop network. This means that we can use Technique 3 with lower transmission power which leads to less interference among the nodes to achieve the same throughput as the one hop network.

The rest of the paper is organized as follows. Section II gives a background on the work related to this research work. Section III presents the topology and the communications architecture proposed for the WSN, as well as the different techniques for data collecting and communication through the WSN. Section IV describes the simulation and testbed environment used to obtain the results in this paper which are evaluated in Section V. Finally, paper conclusions and ideas for future work are listed in Section VI.

II. RELATED WORK

In [8], the authors proposed a scheme for Adaptive Packet Concatenation (APC). In this scheme, several data link layer packets destined to the same node are concatenated in a super packet that is sent once. This reduces protocol overhead coming from sending multiple packets to the same destination. They have proved analytically that this scheme can improve the throughput of the network 4-16 times. The idea is similar to the work done in this paper. However, their work was done in an IEEE 802.11 network performing the aggregation at the data link layer, while we are using IEEE 802.15.4 MAC protocol and we are aggregating data on the application layer.

In [9] Patel et al. have studied query processing as well as data aggregation in wireless sensor networks. They have showed that optimizing data aggregation in a WSN can help reduce energy consumption in the network and in turn increase the life span of the sensors' battery as well as removing any redundancy in data transmission. They have used TinyOs and its simulation tool TOSSIM to simulate a multi-hop sensor network. They have used a multi-hop network in order to reduce both transmission and receiving powers reducing the overall power consumption as well as interference between nodes. Our work here follows a similar path. However, we focused more on data throughput and less on energy consumption.

In [10], the authors have presented a comprehensive survey on data-aggregation algorithms in WSNs. They showed that techniques for data aggregation aim to improve aspects such as network lifetime, latency, and accuracy of the sensed data. They presented an overview of many data aggregation algorithms found in literature showing their advantages and disadvantages. They showed that the performance of any data aggregation algorithm is connected to the infrastructure of the network. And so, not any protocol can perform well in every type of network. We have used a basic method for data aggregation here in a simple multi-hop network in order to reduce interference between the nodes.
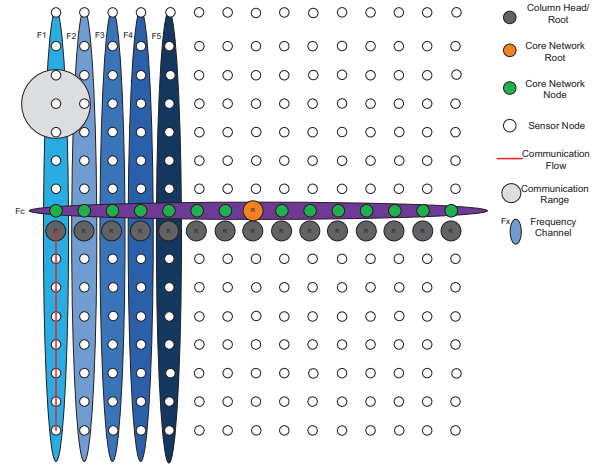
III. PROPOSED ARCHITECTURE

Fig. 1. Network topology and communications architecture.

As the network we are using for this study is a photo voltaic power plant where panels are spread over a huge area, we decided to simulate the network as a grid system of sensors where each sensor is connected to a solar panel and has constant distances to each of its neighbors. Additionally, since most of the communications are done within each column, we decided to divide the whole network into small networks of columns. Each sensor reads the output voltage of its respective panel and sends it to a central node, the sink. And since the changes in the light conditions is slow by nature, it is expected that the required traffic load is going to be low.

The proposed WSN nodes topology and communications architecture is shown in Figure 1. Such setup is referred to as a strip-based deployment [11]. It allows for full coverage while each node having a maximum of two neighbors. This allows nodes to have small transmission range and use small transmission power which reduces interference among nodes. In a real life scenario, nodes within a column are placed side by side with about a distance of 1 meter between each consecutive sensor. While columns have a more wider separation in order to allow for maintenance access.

Additionally, in order to reduce the interference between columns, each column is to operate in a different frequency channel than the other columns. To enable inter-column communication, a core network will be created operating in a frequency different from all the columns, and when the need for inter-column communication arises, the respective column sinks will have to switch to this core frequency and communicate through the core network. However, in this paper only intra-column communication is considered. A group of columns with their access network can be considered as a
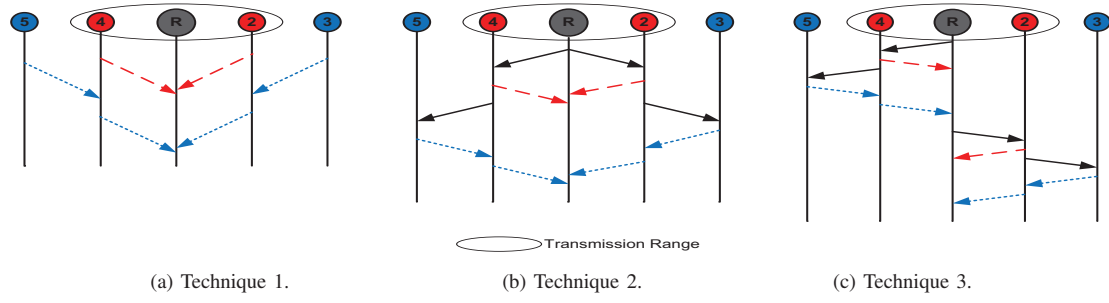
(a) Technique 1. (b) Technique 2. (c) Technique 3.

Fig. 2. Initially proposed data collection and communication techniques.

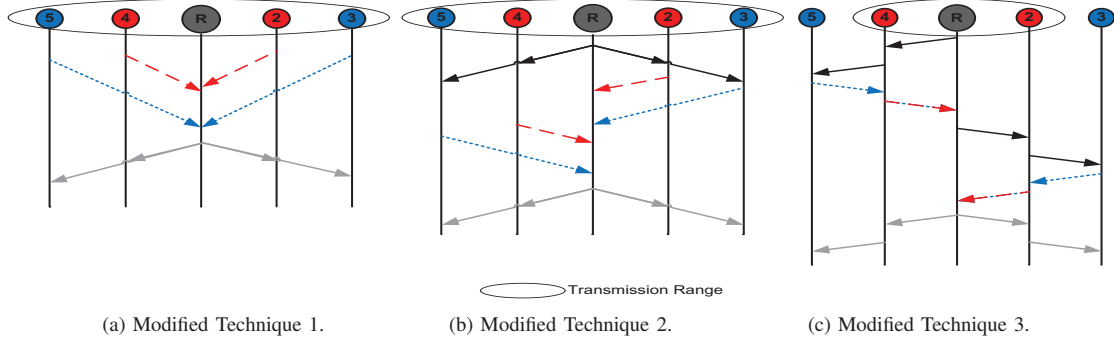(a) Modified Technique 1. (b) Modified Technique 2. (c) Modified Technique 3.

Fig. 3. Modified data collection and communication techniques.

building block which can be replicated as the network grows. In [5], three techniques were analyzed for data collection and communication between the nodes. Figure 2 shows an example of how these techniques work with 4 client nodes in a multi-hop setup. In Technique 1, client nodes generate and send data to the sink which is the basic way of communication in WSN. In Technique 2, the sink sends a broadcast poll message to its neighboring nodes and each node replies with its data as soon as it receives this poll before forwarding it to its other neighbors, thus propagating the poll message across the column. This is done in order to reduce collisions between messages. Finally, in Technique 3, the procedure is similar to Technique 2, except that the sink first sends the poll to one of its neighbors and waits until it receives the information from all the nodes on this side of the network before sending the poll to the other neighbor in order to receive data from the rest of the nodes on the other side of the network. The messages shown in this figure are only the application layer packets; we rely on the MAC layer to perform the acknowledgments and retransmissions of messages when collisions occur.

In this paper, we test the effect of data aggregation on Technique 3, and compare it with the performance of a one hop network employing Techniques 1 or 2 while also enabling the sink command message. This command message enables the sink to communicate with the nodes after each cycle to send a command that can be used in the next cycle. The modified techniques are presented in Figure 3.

In modified Technique 1, the nodes are direct neighbors of the sink, and so, they send their data directly to it. After the data is collected from all the nodes, the sink sends a broadcast command message to all the nodes. In modified Technique 2 the broadcast poll is sent to all the neighbors, the nodes then

take turn sending their data to the sink. And finally, the sink sends the broadcast command message to all of the nodes. Technique 3 has been modified to include data aggregation. When a node receives the poll message, it forwards it to its neighbor and waits until it receives data from this neighbor. After that, it adds its data to the received message in order to create an aggregated message which is then forwarded in the direction of the sink. The sink then repeats this procedure to the other side of the network, and then sends the broadcast command message to its neighbors which is then forwarded to the rest of the nodes.

Modified Techniques 1 and 2 were tested in a one hop setup as this is the setup that achieves the highest throughput of data since every message can reach the sink in one hop. The motivation behind this is to prove that modified Technique 3 while operating in a multi-hop setup with low transmission power and so low interference, can perform close to the other two modified techniques.

IV. SIMULATION AND TESTBED ENVIRONMENTS

Simulations and testbed experiments were done using Contiki2.5 [7] and Sky motes [12]. The three modified techniques were simulated in Contiki's built-in COOJA simulator using a network with a sink and 8 clients. We have selected 8 clients for the network to be symmetric around the sink and with a manageable network size. This setup can be replicated as the network grows. Simulations were performed for different values of the offered load, where the offered load is the data rate that each client is offering to the network. These values were chosen in order to test the limit of the system. Simulations were repeated 10 times for each case with each run working until the sink receives a total of 10000 packets in the case of the simulations and 4000

packets in the case of the testbed experiments.

Additionally, Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) was used in the simulations as the medium access protocol with acknowledgment messages. IEEE 802.15.4 PHY 2.4 GHz [13] was used as the physical layer technology. The channel model used in the simulations is the Unit Disk Graph Medium (UDGM) provided by COOJA with free space distance loss and no fading. We have chosen this model because in a real life scenario, both the nodes as well as the environment are static. And so, there will be low or no shadowing/fading to affect the data transmission, which is a reasonable assumption. In modified Technique 3, nodes were placed apart so that each node has always two neighbors falling within its transmission and interference ranges. This was done easily in the simulator by controlling the transmission and interference ranges of the nodes in simulation environment. On the other hand, in the testbed, small modification had to be made in the radio driver in order to reduce the receiver sensitivity as well as the transmission powers in order to assure that each node has only two neighbors falling within its transmission range.

V. Results and Analysis

In this work, we analyzed the throughput and the packet loss of the network. Throughput is calculated per node as the number of received data packets at the sink from such node divided by the simulation time. Then this value is averaged over the total number of simulated nodes in order to get the average throughput of the network. Packet loss per node is calculated as the number of transmitted packets from this node minus the number of received packets at the sink from the same node divided by the number of transmitted packets from the same node. To be clear, we are counting only the IPv6 packets that arrive correctly at the destination, the retransmitted layer two frames are not counted when calculating the packet loss. Additionally, the load is defined here as the data rate that each node is offering to the network.

A. Link analysis

Figure 4 shows the throughput of a network with one client for different values of offered load. The client sends data to the sink with a constant rate. This rate is changed gradually in order to test the capacity of the system. From the figure we can see that the throughput increases linearly with the offered load until a point with 47.76 pac/s at a value of offered load of 50 pac/s after which it saturates. This can give us an idea about the limitation of the system which can help us when doing the rest of the simulations by establishing a base line of the maximum throughput a single node can achieve in this scenario. These results coincide with the results we got in [5] in which the one hop delay was 20ms per packet per node, which means that a node can send up to 50 pac/s. Which is close to the link analysis performed.

We are assuming that hop delay comes mainly from the packet transmission delay, medium access delay, and processing delay. The propagation delay for such small distances between the nodes is very low and so can be neglected.
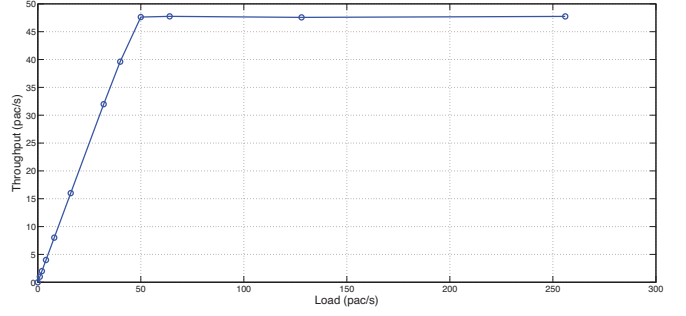

Fig. 4.   Throughput vs. offered load for a system with one client node.

B. Performance of the three modified techniques with no sink command messages

Figure 5 shows the throughput of a network with 8 clients for the three modified techniques for different values of the offered load with no sink command messages. Results show that the throughput of a one-hop network using modified Techniques 1 and 2 saturates at 5.3 and 5.5 pac/s/node, respectively, for a value of offered load of 6 pac/s/node. A multi-hop network using modified Technique 3 with data aggregation, operates close to the other two modified techniques until a value of offered load of 5 pac/s/node after which the performance degrades. This is due to the fact that as the value of the offered load increases, the time between the polls decreases, and so there will come a point when the next poll is sent before all the nodes had time to send their data to the sink. This will lead to congestion which increases the packet loss and causes the throughput to decrease. Using the
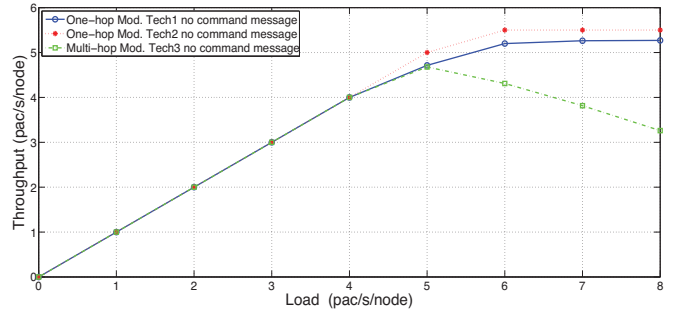

Fig. 5.   Throughput vs. offered load for an 8 clients network for the modified techniques with no sink command messages.

results of the link analysis found in the previous subsection, an approximation of the maximum average throughput of a network with 8 nodes can be roughly estimated by dividing the one node network maximum throughput by 8, which gives us 6.25 pac/s/node. This is close to the results found from the simulations.

C. Performance of the three modified techniques with sink command messages enabled

Figure 6 shows the throughput of a network with 8 clients for the three modified techniques for different values of the offered load with the sink command messages enabled. We can see that the overall throughput decreases a little when

the sink command messages are enabled which is due to the increased number of messages to be sent in each cycle, which in turns leads to more congestions and packet losses. However, the difference is not that big and we can see that the trend of the results is the same, with modified Techniques 1 and 2 saturating at a value of 6 pac/s/node offered load and with 5.2 and 5.3 pac/s/node, respectively. And with modified Technique 3 performing close until 5 pac/s/node of offered load.
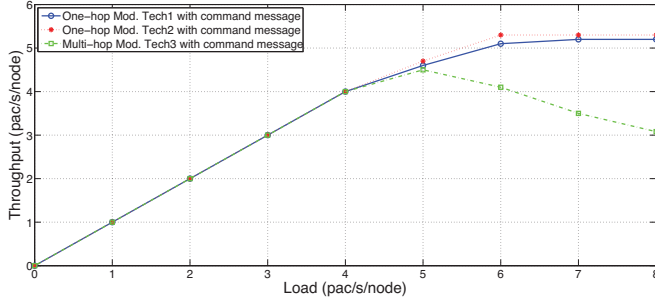

Fig. 6. Throughput vs. offered load for an 8 clients network for the modified techniques with sink command messages enabled.

Figure 7 shows the sink command message throughput of a network with 8 clients for the three modified techniques for different values of the offered load with the sink command messages enabled. The sink command messages throughput is calculated as the mean number of sink command messages received by the nodes divided by the simulation time. The sink command message contains information from all nodes in the network which means that its size is equal to the size of the full aggregated packet in the case of modified Technique 3. We can see from the figure that the trend is the same as the previous figures.
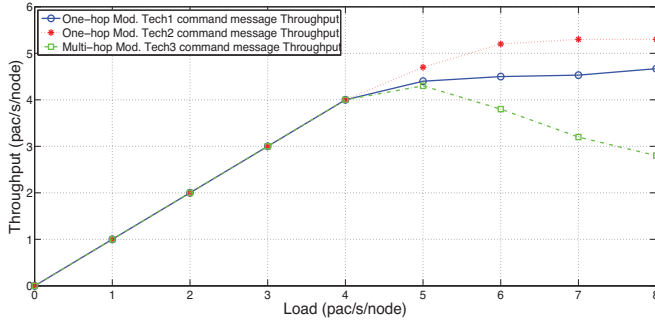

Fig. 7. Command message throughput vs. offered load for an 8 clients network for the modified techniques with sink command messages enabled.

D. Packet loss per node of the three techniques with no sink command messages

Figure 8 shows the packet loss per node against the node id for a network with 8 clients with the sink being node 1 for the three modified techniques for different values of offered load with no sink command messages. Values of offered load are 6, 7, and 8 packets/s/node. The reason for showing only these values is because the loss is almost zero for the lower values of offered load. We can see from the figures that while modified Technique 3 performs worse than the other techniques for these values of offered load, the loss is almost constant for

all the nodes in the system. Which means that there is a degree of fairness in the network as all the nodes transmit the same amount of data. This is different for the other techniques where nodes closer to the sink have low values of packet loss compared with nodes that are far from the sink.

We can also observe from the figures that as the offered load increases, the loss difference between the two branches of the network increases as well, with the branch to be polled second having higher losses. This is due to the fact that as the load increases, the time taken to poll the nodes decreases. And so, nodes on the second branch will have less time to reply which leads to higher losses.

From what we can see from these results, Technique 3 with data aggregation in a multi-hop network, performs close to the other techniques in a one-hop network and no aggregation up to a value of the offered load after which its performance degrades severely due to the congestion.
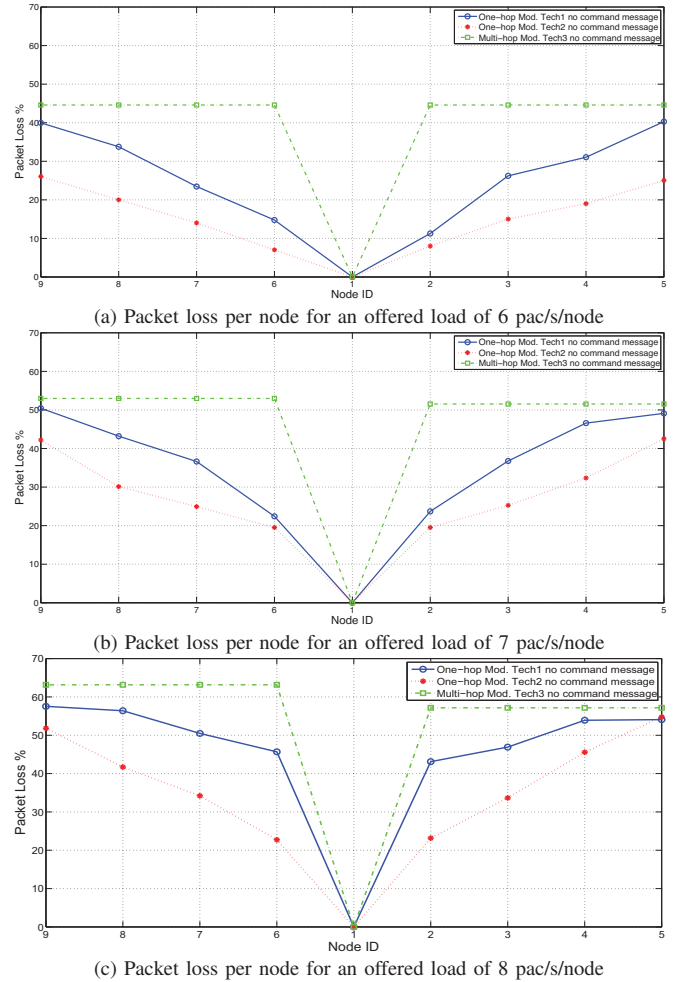

(a) Packet loss per node for an offered load of 6 pac/s/node

(b) Packet loss per node for an offered load of 7 pac/s/node

(c) Packet loss per node for an offered load of 8 pac/s/node

Fig. 8. Node packet loss of a network with 8 client nodes for different values of offered load.

E. Testbed experiments for Technique3

In order to validate results obtained from the simulation of modified Technique 3, a testbed was created. The testbed consisted of eight clients and a sink placed in a line with the

sink in the middle. The transmission power was reduced and nodes were placed relatively far apart in order to ensure that each node has only two direct neighbors in order to create a multi-hop network. Additionally, receiver sensitivity was reduced in order to ignore any packets coming from nodes who were not direct neighbors.

Figure 9 shows the throughput of both Technique 3 and modified Technique 3 against the offered load with command message disabled. It also shows results obtained from both the simulation and the testbed experiments.

As we can see from the figure, the testbed results confirm the results obtained from the simulation in both the original and modified Technique 3. The testbed results are slightly worse in terms of the throughput due to the interference which is found in nature from other devices operating in the real world and was not taken into consideration while performing the simulation. Additionally, the figure clearly shows the improvement the aggregation gives to the performance of the technique when compared with the original technique with no aggregation. This improvement comes from the reduced number of messages that are needed to be transmitted with the use of aggregation.
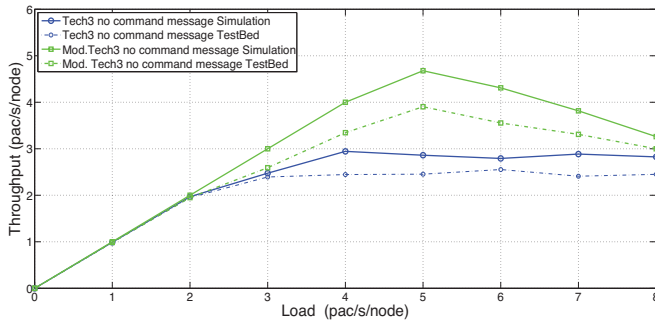

Fig. 9. Throughput vs. offered load for an 8 clients network for modified and original Technique3 with no sink command message.

VI. Conclusions and Future Work

In this paper, we studied the effect of data aggregation and sink command messages on the performance of an IEEE 802.15.4 WSN network employing a polling based data collecting technique using the Contiki COOJA simulator, and validating some of the results with the use of a testbed using Sky motes. The study was done for a fixed sized network with 8 client nodes and different values of the offered load, estimating the network throughput and packet loss for each value of the offered load.

This evaluation enabled us to conclude that the selected technique with data aggregation allows a set of nodes in a column to communicate among themselves with low probability of packet loss and high values of average throughput up to a considerable value of the offered load. Results showed that from the three proposed techniques, Technique 3 with data aggregation in a multi-hop network performs close to the other two techniques with no aggregation in a one-hop network. We have selected to use a multi-hop network in order to reduce transmission ranges as well as interference levels among the different nodes. And with modified Technique 3 performing

close to the others in terms of throughput in a multi-hop network we were able to show that it is the best choice for the requirements of the network.

However, as the offered load increases beyond a certain point, the performance of modified Technique 3 degrades due to congestion of the network which in turns adds to the packet loss. Additionally, enabling the sink command messages also degrades the performance a little due to the added number of messages to be sent. On the other hand, sink command messages might be needed if the sink is required to communicate with the clients, sending different commands which can be used in the next cycle. Additionally, results obtained from the simulation of both Technique 3 and modified Technique 3 were validated by the use of the testbed. The work done in this paper focused on trying to improve the work done in [5] by introducing data aggregation and sink command messages to the previously introduced technique. The following research aspects we will tackle are the inter column communication, as well as the self configuration problem, in order to create a totally scalable and self-configurable network.

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," in *Computer Networks*, vol. 52, no. 12, 2008, pp. 2292 –2330.

[2] M. Dohler, D. Barthel, R. Maraninchi, L. Mounier, S. Aubert, C. Dugas, A. Buhrig, R. Paugnat, M. Renaudin, A. Duda, M. Heusse, and R. Valois, "The ARESA project: Facilitating research, development and commercialization of WSNs," in *The 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks. SECON '07*, June 2007, pp. 590 –599.

[3] A. Bachir, M. Dohler, T. Watteyne, and K. Leung, "MAC essentials for wireless sensor networks," *IEEE Communications Surveys Tutorials*, vol. 12, no. 2, pp. 222 –248, 2010.

[4] "SELF-PVP," in *http://www.cmuportugal.org/tiercontent.aspx?id=3374*.

[5] M. M. Abdellatif, J. M. Oliveira, M. Ricardo, and P. Steenkiste, "Impact of data collecting techniques on the performance of a wireless sensor network," *Proceedings of the ISWCS 2012, the Ninth International Symposium on Wireless Communication Systems*, August 28-31 2012.

[6] F. Osterlind, A. Dunkels, J. Eriksson, N. Finne, and T. Voigt, "Cross-level sensor network simulation with COOJA," in *Proceedings of the 31st IEEE Conference on Local Computer Networks*, November 2006, pp. 641 –648.

[7] "ContikiOS," in *http://www.contiki-os.org/*.

[8] H. Zhai and Y. Fang, "A distributed packet concatenation scheme for sensor and ad hoc networks," in *Military Communications Conference, 2005. MILCOM 2005. IEEE*. IEEE, 2005, pp. 1443–1449.

[9] N. S. Patil and P. R. Patil, "Data aggregation in wireless sensor network," *IEEE International Conference on Computational Intelligence and Computing Research*, 28-29 December 2010.

[10] V. Pandey, A. Kaur, and N. Chand, "A review on data aggregation techniques in wireless sensor network," *Journal of Electronic and Electrical Engineering*, vol. 1, no. 2, pp. 01–08, 2010.

[11] F. Wang and J. Liu, "Networked wireless sensor data collection: Issues, challenges, and approaches," *IEEE Communications Surveys Tutorials*, vol. 13, no. 4, pp. 673 –687, 2011.

[12] "Tmote Sky," in *http://www.snm.ethz.ch/Projects/TmoteSky*.

[13] "IEEE802.15.4," in *http://www.ieee802.org/15/pub/TG4.html*.