



The University of Manchester Research

# A Comparative Study of Data Aggregation Approaches for Wireless Sensor Networks

**DOI:** 10.1145/2988272.2988285

# **Document Version**

Accepted author manuscript

# Link to publication record in Manchester Research Explorer

# Citation for published version (APA):

Naureen, A., & Zhang, N. (2016). A Cómparative Study of Data Aggregation Approaches for Wireless Sensor Networks. In *Proceedings of the 12th ACM Symposium on QoS and Security for Wireless and Mobile Networks* https://doi.org/10.1145/2988272.2988285

#### **Published in:**

Proceedings of the 12th ACM Symposium on QoS and Security for Wireless and Mobile Networks

#### Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

#### **General rights**

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

#### **Takedown policy**

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



# A Comparative Study of Data Aggregation Approaches for Wireless Sensor Networks

Ayesha Naureen ayesha.naureen@manchester.ac.uk

Ning Zhang ning.zhang@manchester.ac.uk

School of Computer Science The University of Manchester Manchester, United Kingdom

# ABSTRACT

In Wireless Sensor Networks (WSNs), data aggregation has been used to reduce bandwidth and energy costs during a data collection process. However, data aggregation, while bringing us the benefit of improving bandwidth usage and energy efficiency, also introduces opportunities for security attacks, thus reducing data delivery reliability. There is a trade-off between bandwidth and energy efficiency and achieving data delivery reliability. In this paper, we present a comparative study on the reliability and efficiency characteristics of different data aggregation approaches using both simulation studies and test bed evaluations. We also analyse the factors that contribute to network congestion and affect data delivery reliability. Finally, we investigate an optimal trade-off between reliability and efficiency properties of the different approaches by using an intermediate approach, called Multi-Aggregator based Multi-Cast (MAMC) data aggregation approach. Our evaluation results for MAMC show that it is possible to achieve reliability and efficiency at the same time.

# Keywords

Wireless Sensor Networks; Data Aggregation; Base Station; Aggregators; Efficiency; Reliability

# 1. INTRODUCTION

Data aggregation is widely applied in hierarchical WSNs to reduce energy consumptions in data collections. A number of data aggregation methods have been proposed in literature to achieve reliable and energy efficient data aggregation. These methods, depending on the number of aggregators used, may be classified into three data aggregation approaches as: Single-Aggregator based Uni-Cast (SAUC) approach, Multi-Aggregator based Uni-Cast (MAUC) approach and Multi-Aggregator based Broad-Cast (MABC) approach. With SAUC, each node sends its data along one path to its next-hop neighbour and data aggregation is per-

*Q2SWinet'16, November 13-17, 2016, Malta, Malta* © 2016 ACM. ISBN 978-1-4503-4504-0/16/11...\$15.00 DOI: http://dx.doi.org/10.1145/2988272.2988285 formed at a single location by a single aggregator in the network. This aggregator is typically the BS. With the MAUC and MABC approaches, on the other hand, multiple aggregators spread out on multiple hops are used. The difference is that with the MAUC approach, each node transmits its data along a single path to exactly one aggregator at each hop, whereas with the MABC approach, each node broadcasts its data along multiple paths to multiple aggregators at each hop. Among the three approaches, MAUC consumes the least energy as it requires fewer transmissions in the network. However, it is prone to deliberate and accidental loss of individual and aggregated data. With MABC, data delivery reliability is better than both SAUC and MAUC as each node uses more aggregators at each hop.

There are rooms for improvement in terms of optimizing data delivery reliability and energy efficiency of the data aggregation process. This paper presents a comparative study of the SAUC, MAUC and MAMP data aggregation approaches. This study allows us to investigate an ideal approach for data aggregation that may provide both energy efficiency as well as data delivery reliability.

The rest of this paper is organized as follows. Section 2 provides a critical analysis on the energy consumption and data delivery reliability of the three data aggregation approaches. Section 3 investigates the factors contributing to network congestion and evaluates the energy consumption and data delivery reliability characteristics of an intermediate approach against the existing data aggregation approaches. Finally, Section 4 concludes the paper.

# 2. EFFICIENCY AND RELIABILITY ANAL-YSIS

Intuitively, the MAUC approach should introduce less communication overhead than its counterparts, i.e. SAUC and MABC. The SAUC approach requires each node to transmit its data separately to the BS, which is expensive in terms of bandwidth costs. The MABC approach requires each node to broadcast its data to multiple aggregators and this is also an expensive option with respect to bandwidth costs. In comparison with these two approaches, the MAUC approach should be cheaper in terms of bandwidth costs as it only requires each node to transmit its data to one aggregator at each hop. This should lead to fewer transmissions, thus less energy consumption, in comparison with the SAUC and MABC approaches.

We have used both simulation and real WSN test bed to evaluate energy consumptions of the three data aggregation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions @acm.org.



Figure 1: Energy Consumption Levels for the Three Aggregation Approaches

approaches (SAUC, MAUC and MABC). We have implemented the three approaches on TinyOS [3], using NesC. The CTP (Collection Tree Protocol) [1] is used to count the number of nodes in each of the approaches. CTP is essentially a unicast routing protocol. We have modified the CTP protocol to support broadcasting in case of MABC.

As mentioned above, two evaluation methods are used. The first method is by simulation for which we have chosen the Cooja simulator [6]. The values of the parameters used in the simulation are set as follows. The payload size of the message is fixed to 27 bytes for each of the three approaches. The simulated network consists of 12 nodes, with four leaf nodes (nodes 4, 5, 6, 8), four 1-hop upstream neighbours (nodes 3, 7, 9, 10) and three 2-hop upstream neighbours (nodes 2, 11, 12) of the leaf nodes, and one BS (node 1). The maximum hop count is 3. The length of a simulation run for each result collection is 90 seconds. The Unit Disk Graph Medium (UDGM) radio model is used in the simulation, where each node can only communicate with the nodes present within its communication radius. To get energy consumption values for each of the approaches, we have used LEDs (light-emitting diodes) and printf debugging facility to count the number of messages received and transmitted by each node, and the statistics from CC2420 transceiver datasheet that show current consumption in receive and transmit mode as 18.8 mA and 17.4 mA respectively [2].

The energy consumption values for each of the nodes in the network for each of the three data aggregation approaches are recorded and plotted in Figure 1(a). From the results shown in the figure, we can make four observations. The four observations are:

- The first is that, the MABC approach generates the highest level of energy consumption among the three approaches, and this is the case for almost all the nodes in the network. This observation is within our expectation, as the broadcasting approach is both bandwidth and energy expensive.
- The second observation is that the MAUC approach consumes the least level of energy, thus the most efficient data aggregation approach. This observation too is within our expectation, because at each hop only one node is aggregating data for its downstream neighbours.
- The third observation is that the energy consumption level per node fluctuate similarly for some nodes in both SAUC and MAUC approaches. For example, nodes 7,9 in Figure 1(a) experience higher energy consumption levels than others. This is because

these high energy consuming nodes receive and transmit data from the downstream neighbours (nodes 4,5,6,8) in case of SAUC and receive, aggregate and transmit aggregated data from the downstream neighbours (nodes 4,5,6,8) in case of MAUC.

• The fourth observation is that, different from our intuitive thinking, the SAUC approach consumes similar level of energy consumption as the MAUC approach except for the nodes closest to the BS (i.e. nodes 2, 11, 12) and the BS (node 1). This is because the more close a node is to the BS, more data are received, and more transmissions need to be done thus corresponding to higher level of energy consumptions. The high energy consumption at BS is because it receives and processes 11 individual transmissions (data from each node in the network) for SAUC, as compared to just 3 transmissions (aggregated data from nodes 2, 11, and 12) for MAUC.

The second evaluation method is to test the three approaches on a sensor network test bed, the FlockLab test bed [4]. We used the same payload size of 27 bytes for each packet and ran the tests for 600 seconds each in a network of 29 nodes. Different from the simulation experiments, the energy consumptions in our test bed experiments indicate the energy consumed not only by the message transmissions but also by the data computations. Initial experimental results showed that LEDs and printf debugging facilities were energy intensive, so it was hard to observe the differences in the respective energy consumptions of the three approaches. To overcome this problem, we modified our experimental implementations in two ways. Firstly, we used CTP on top of LPL (Low Power Listening) [5]. Secondly, instead of using LEDs and printf for debugging, we logged data received at the serial port on the BS for debugging. After these modifications, we are able to see the respective power consumptions of the three approaches, which have been plotted in Figure 1(b). From these experimental results, we can make two observations consistent with the simulation results shown in Figure 1(a) and two other observations. The four observations are:

- The MAUC approach consumes the least energy among the three approaches.
- There are energy peaks for certain nodes (e.g. nodes 2, 12, 25) with the SAUC and MAUC approaches. As explained earlier, the energy peaks are due to the fact that these nodes receive and transmit data from the downstream neighbours in case of SAUC, and receive, aggregate and transmit aggregated data from the downstream neighbours in case of MAUC, which causes additional energy consumptions at these nodes.

Table 1: Reliability Characteristics for the Three Data Aggregation Approaches

$DataAggregation \\ Approach$	$\begin{array}{c} DataDelivery\\ Ratio(DDR) \end{array}$	
	Cooja	Flocklab
SAUC	1	1
MAUC	1	0.828
MABC	0.675	0.759

(a) Case I

(b) Case II

$DataAggregation \\ Approach$	$\begin{array}{c} Data Delivery \\ Ratio(DDR) \end{array}$	
	Cooja	Flocklab
SAUC	0.825	0.414
MAUC	0.925	0.689
MABC	0.675	0.759

- However, the energy peaks are much less prominent for the nodes (2, 12, 25) in the MAUC approach as compared to the SAUC approach. This indicates that computation of aggregates at the intermediate nodes (aggregators) consumes much less energy than the energy consumed when the intermediate nodes have to receive and transmit data for each of the downstream neighbours.
- There are no energy peaks in the MABC approach, and the average energy consumption for each node in the network is almost identical, maintaining at the highest level of 20 mA. This may indicate that, with the increased number of nodes in the network, the MABC approach generates more messages (than the other two approaches) and almost all the nodes in the network have to perform the role of an aggregator.

With regard to data delivery reliability, the more the number of aggregators per hop, the higher is the probability of successful data delivery to the BS. With the SAUC approach, there is no aggregation at the intermediate nodes but it does require one intermediate node per hop to receive data from the downstream neighbours and forward it towards the BS. The data delivery reliability, in this case, depends on the reliability and honesty of each of the intermediate nodes along the path. If there is a single intermediate node along the path, which is compromised or not trustworthy, then the data will not be, or false data will be, delivered to the BS. Similarly with the MAUC approach, as only one aggregator is used at each hop, to ensure a successful data delivery at the BS, every aggregator along the path must be reliable and honest. With the MABC approach, on the other hand, as multiple aggregators are used at each hop, the probability of successful data delivery to the BS is much higher than the SAUC and MAUC approaches. As long as, at least, one of the multiple aggregators at each hop remains honest or reliable, correct aggregation results can be obtained by the BS. This means that, in terms of data delivery reliability, the MABC approach is more preferable.

We investigated data delivery reliability for the three approaches, SAUC, MAUC and MABC, using both Cooja simulator and Flocklab test bed. The investigation is carried out under two network reliability cases. Case I assumes that the underlying network is trustworthy i.e. all nodes are working reliably and there are no failed and/or compromised nodes in the network. Case II assumes that the underlying network contains failed and/or compromised nodes. We measure the data delivery reliability in terms of DDR (Data Delivery Ratio), which is defined as the ratio of successful data collections at the BS.

Initially, we ran the simulation for a network size of 12 nodes on the Cooja simulator. The DDR values for the three

approaches are all 1 (i.e. 100% of the data aggregates were received by the BS). We then increased the network size to 40 nodes on the Cooja simulator and collected results from these simulation runs. We also tested data delivery reliability for the three approaches on the Flocklab test bed for a network of 29 nodes. The results from the simulation and test bed are presented in Table 1(a). From these results, we can see that MABC fails to deliver a 100% DDR value as we expected. We were able to confirm the same observations on the Flocklab test bed as well, for a network of 29 nodes. This is due to the fact that, as additional nodes are added into the network, more packets are injected into the network, and with MABC, the increase in the traffic level can easily make the network congested. If the network is congested, the probability of packet collisions over wireless media also increases and leads to more data loss. This means that, to improve data delivery reliability, MABC should be broadcasting data to a fewer number of neighbours.

In Case II, we have introduced failed nodes in the network of 40 nodes on Cooja simulator and in the network of 29 nodes on Flocklab test bed. Table 1(b) summarises the DDR values for the three approaches. Comparing the results in Table 1(b) with those in Table 1(a), we can see that the DDR values for SAUC and MAUC decrease while the DDR values for MABC remain the same with failed and/or compromised nodes in the network. This means that, under our experiment setting, the reliability performance of the SAUC and MAUC approaches are mostly affected by the failed and/or compromised nodes in the network, whereas the reliability performance of the MABC approach is largely affected by packet collisions resulting from the congested network. This means that, in the design of a data aggregation method, the task of controlling the communication overheads and preventing the network from being congested is as important as tackling malicious nodes or node failures.

#### 3. DISCUSSION

In the reliability analysis of Section 2, we observe a difference in the simulation results and test bed results for both Case I and Case II. This is because our simulation, operating in an ideal network environment, experiences only congestions and collisions introduced by our network configuration, with no noise or interference from the environment. On the other hand, the Flocklab test bed, operating in an actual network environment, experiences congestions and collisions introduced by our network configuration, as well as the noise and interference from the environment.

We investigate two factors that may contribute to network congestion and affect data delivery reliability in MABC. These two factors are the number of nodes in the network and the placement of nodes in the network. Firstly, we con-

#### Table 2: Determining the Factors Contributing to Network Congestion

(a) Impact of the Number of Nodes

$DataAggregation \\ Approach$	Data Delivery Ratio $(DDR)$ for 'n' Nodes			
	20	40	60	80
MAUC	1	0.95	0.783	0.712
MABC	0.95	0.675	0.55	0.488

sider a network of n' sensor nodes in a 10 x 10 grid on Cooja simulator. Table 2(a) compares the DDR values for MAUC and MABC for 20, 40, 60 and 80 nodes placed in the grid, operating in a trustworthy network environment (i.e. Case I). From the Table, we see that for just 20 nodes network, MAUC and MABC show the same DDR values, indicating a non-congested network. However, as more nodes are placed in the grid, the DDR values for MAUC degrade linearly whereas the DDR values for MABC degrade exponentially. Secondly, we consider a network of 40 nodes placed linearly and elliptically in a 10 x 10 grid on Cooja simulator. Table 2(b) gives the DDR values for MAUC and MABC for linear and elliptic placement of the nodes in a trustworthy network environment (i.e. Case I). We ran the simulation for 45 seconds, which was the time required by MAUC, in both linear and elliptic placement, to deliver a 100 % DDR. From the results, we see that MABC gives a better DDR when the nodes are placed in a linear manner. This is because with linear placement, fewer nodes are within the communication radius of a node, leading to reduced traffic levels in the network and providing better data delivery reliability. Our results from Table 2 verify that, unlike MAUC, the DDR of MABC is highly dependent on the number of nodes in the network and their placement in the network.

From the analyses in Section 2, we have shown that the SAUC approach is neither efficient nor reliable, the MAUC approach is efficient but not reliable in case of failed and/or compromised nodes in the network and the MABC approach is not efficient and only provides reliability when the network is not congested. To achieve optimal efficiency and data delivery reliability, we have investigated the use of two aggregators along each hop to deliver data to the BS. This new approach is called the Multi-Aggregator based Multi-Cast (MAMC) data aggregation approach. Figure 2(a) and 2(b) compare the energy consumption levels and the reliability characteristics for MAUC, MAMC and MABC in an untrustworthy network environment (i.e. Case II). From the results, we can see that MAMC exhibits similar level of energy consumptions as MAUC and achieves similar reliability level as MABC. This indicates that there is a possibility to achieve efficiency and reliability at the same time by using an intermediate approach such as MAMC.

## 4. CONCLUSION

In this paper, we have carried out a comparative study of three data aggregation approaches, SAUC, MAUC and MABC. Our evaluation results show that MAUC consumes the least energy, compared to SAUC and MABC. However, it is prone to deliberate and accidental loss of individual and aggregated data. Our evaluation also showed that MABC provides conditional data delivery reliability i.e. it only provides a higher level of data delivery reliability under the condition that the network should not be congested. Through

#### (b) Impact of the Placement of Nodes





(a) w.r.t. Energy Consumption Levels

$DataAggregation \\ Approach$	$DataDelivery \\ Ratio(DDR)$	
	Cooja	Flocklab
MAUC	0.925	0.689
MAMC	0.95	0.828
MABC	0.675	0.759

(b) w.r.t. Reliability

Figure 2: Comparison of MAMC with MAUC and MABC

this study, we have discovered that there are rooms for improvement in terms of optimizing both efficiency and reliability in data aggregation. As a first step, we have tested the idea of an intermediate data aggregation approach, MAMC. As our next step, we will pursue the design of a data aggregation approach that provides reliability and cost-effectiveness at the same time.

#### 5. **REFERENCES**

- R. Fonseca, O. Gnawali, K. Jamieson, S. Kim, P. Levis, and A. Woo. The collection tree protocol (ctp). *TinyOS TEP*, 123:2, 2006.
- [2] T. Instruments. Cc2420 datasheet. *Reference* SWRS041B, 2007.
- [3] P. Levis, S. Madden, J. Polastre, R. Szewczyk, K. Whitehouse, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, et al. Tinyos: An operating system for sensor networks. In *Ambient intelligence*, pages 115–148. Springer, 2005.
- [4] R. Lim, F. Ferrari, M. Zimmerling, C. Walser, P. Sommer, and J. Beutel. Flocklab: A testbed for distributed, synchronized tracing and profiling of wireless embedded systems. In *Information Processing* in Sensor Networks (IPSN), 2013 ACM/IEEE International Conference on, pages 153–165. IEEE, 2013.
- [5] D. Moss, J. Hui, and K. Klues. Low power listening. *TinyOS Core Working Group*, *TEP*, 105, 2007.
- [6] F. Osterlind, A. Dunkels, J. Eriksson, N. Finne, and T. Voigt. Cross-level sensor network simulation with cooja. In *Local computer networks, proceedings 2006* 31st IEEE conference on, pages 641–648. IEEE, 2006.