# Comparison of Protocols for Fast Data Transmission in Binary Tree WSNs

Emanuele Goldoni and Paolo Gamba
University of Pavia
Dipartimento di Ing. Industriale e dell'Informazione
Via Ferrata 1 - 27100 Pavia, Italy
Email: {name.surname}@unipv.it

***Abstract*—Convergecast – the collection of data from many sources to a single sink – is a common operation in Wireless Sensor Networks. Early-warning systems and alarms, but also habitat monitoring and precision agriculture are examples of applications where data are collected from outlying nodes to the root through a direct spanning tree. The simplest approach forwards as-is, without performing intermediate processing or compression on transit data. However, saving energy is crucial for prolonging the lifetime of sensor nodes. Hence, some functions might be applied to use only a fraction of maximum schedulable load. Finally, the application of compressive sampling theory to data gathering in WSNs has recently emerged as a viable solution for reducing communication costs and extending network lifetime without introducing intensive in-network computation.**

**This work proposes an analysis of three main techniques for many-to-one data collection, namely Raw Convergecast, Aggregated Convergecast and Compressed Data Gathering. The achievable lower bounds for the schedule length and energy consumption of these three methods are derived for a binary tree topologies, and a numerical comparison of the results is presented.**

## I. Introduction

Many-to-one collection of data is a common operation in many scenarios, from early-warning systems to habitat monitoring. This process in Wireless Sensor Netoworks is referred as Convergecast: the gateway collects data from the nodes through a direct spanning tree, and most of data processing and decisioning is performed out of the network.

The simplest approach forwards raw-data as-is, without performing intermediate processing or compression on transit traffic. However, saving energy is crucial for prolonging the lifetime of sensor nodes. Since the wireless communication module is usually accounted for most of the power consumption, an effective way of reducing power consumption is to reduce the amount of transmitted data. Hence, some functions might be applied to aggregated and compress data, so that the actual payload uses only a fraction of the maximum allow load.

For example, [1] presents two algorithm for minimum energy data aggregation convergecast in WSNs, and their performances are investigated in the same article. Similarly, in [2] Durmaz Incel et al. study raw-data and aggregated convergecast in WSNs where nodes communicate using TDMA to minimize the schedule length. The authors also evaluate the performance of various channel assignment methods, addressing the fundamental limitations due to interference and half-duplex transceivers. A comprehensive survey of contention-free time division based protocols for data collection over tree-based routing topologies is presented in [3].

More recently, the application of compressive sampling (CS) theory to data gathering in WSNs has been envisioned as a useful technique to improve the performance of wireless sensor networks, extending network lifetime without introducing intensive in-network computation [4], [5].

This work proposes an analysis of three main techniques for many-to-one data collection, namely Raw Convergecast, Aggregated Convergecast, and CS-based Data Gathering. The achievable lower bounds for the schedule length and energy consumption of these three methods are derived for a complete binary tree topologies, and a numerical comparison of the results is presented.

In the next Section, the three data gathering techniques are presented, showing for each of them an analysis of the performance bounds. Section III presents a comparison of the protocols, showing the results in terms of delay and energy obtained through a numerical simulator. Finally, some concluding remarks and perspective on future research works are shown in the Section IV.

## II. Data Gathering Methods

In WSNs, only a few nodes are usually able to communicate directly with the gateway, but each device can be exploited as a relay device to forward data towards the gateway. Specifically, we consider a scenario, where multi-hops path are logically organized as a binary tree. This is not an oversimplifying hypothesis: many routing protocols designed for the WSN domain calculate network paths so that transmissions are routed and forwarded along such a hierarchical topology [6].

We also assume that nodes are not able to transmit and receive at the same time, since they are equipped with half-duplex transceivers. Hence, a collision-free TDMA schedule is needed, in order to avoid that a packet is sent while the intended receiver is still transmitting. There exist also interferences among concurrent transmissions, but combining power control with the usage of multiple channels can provide an effective way to cope with collisions.

In this Section we aim at computing the achievable lower bounds for the schedule length of three different data collection methods: raw convergecast, aggregated convergecast, and Compressed Data Gathering (CDG), that is data gathering based on Compressed Sampling.

We define $n_0$ as the total number of packets received by the gateway – which is the root node of the binary tree – for a convergecast iteration. We assume that all nodes generate the same amount $P$ of data, and the actual capacity of all wireless links in the network is equal to $C$. Given a generic level $l$ of the binary tree, each node at this level of the routing path takes $t_{rxl}$ to receive data from the child nodes, while its total transmission time is referred as $t_{txl}$. The amount of traffic generated by this node, expressed in number of packets, is $\lambda_l$. Looking at the whole tree, the number of time slots needed to schedule all the transmissions is $N$, and the time required to complete the convergecast process is $H$.

### A. Raw-data Convergecast

In raw data convergecast, packets are individually forwarded toward the sink and no intermediate processing is performed. Each packet is scheduled by the intermediate nodes independently from other packets – differently from aggregated convergecast, out-of-order delivery is allowed.

The minimum number $N$ of slots required in this configuration corresponds to the number of nodes in the tree minus the root node [7]. And since each node sends one packet,

$$N = n_0$$

where

$$n_0 = 2^{h+1} - 2 = 2\left(2^h - 1\right)$$

A node a level $l$ will send to the father its packet, plus all packets generated by all child nodes, up to the leaf nodes at level $h$:

$$\lambda_l = 1 + \sum_{i=l+1}^{h} 2^{i-l} = 1 + \frac{1}{2^l}\sum_{i=l+1}^{h} 2^i$$

and the summation can be rewritten, so that

$$\lambda_l = 1 + \frac{1}{2^l}\left(2\left(2^h - 2^l\right)\right) = 2^{h-l+1} - 1$$

Since no intermediate processing or aggregation is performed, all packets have the same size. Hence, radio usage both for transmission and reception can be easily computed as:

$$T_{txl} = \lambda_l \frac{P_l}{C_l} = \lambda_l \frac{P}{C} = \left(2^{h-l+1} - 1\right)\frac{P}{C}$$

$$T_{rxl} = \sum_{i=l+1}^{h} 2^{i-l}\frac{P}{C} = \left(2^{h-l+1} - 2\right)\frac{P}{C}$$

And the time required for gathering data from all the sources is directly related to number of time slots and their duration:

$$H = N\frac{P_l}{C} = 2\left(2^h - 1\right)\frac{P}{C}$$

.

### B. Aggregated Convergecast

Data aggregation is frequently used in convergecast to reduce the volume of transmitted messages, thus increasing the energy-efficiency. Each intermediate node perform an in-network aggregation to reduce the payload before forwarding it. Aggregated convergecast is useful when there is a strong correlation between adjacent sensors, or the network is used to received a summarized value like true/false or the maximum reading.

It is usually assumed that there is a compression factor $\gamma$, which is application-dependent and it represents the data fraction actually received by the gateway. More formally, if a node $j$ receives a packet of size $data_i$ from a node $i$, the size of the aggregated packet sent by $j$ after the reception will be

$$data_j + (1-\gamma)data_i$$

If the compression is maximum ($\gamma = 1$), such as for the extraction of the $min$ value, the packet size is constant. On the contrary, no aggregation is performed and data are appended without compression if $\gamma = 0$. The latter case, also referred as *concatenation*, represents the worst-case for aggregated convergecast and will be considered in our analysis.

Since node's data and data coming from the lower level are aggregated in a single packet before transmission, the gateway will receive only 2 packets

$$n_0 = 2$$

and each node sends only one packet

$$\lambda_l = 1$$

From [8] we also know that the number of slots needed for this kind of convergecast is

$$N = 2(h-1) + 2 = 2h$$

Packet size increases after each step along the path, and its size comes from the aggregation of new raw data $P$ plus all the packets generated by children:

$$P_l = P + \sum_{i=l+1}^{h} 2^{i-l}P = \left(2^{h-l+1} - 1\right)P$$

Hence, transmission time can be computed as

$$T_{txl} = \lambda_l\frac{P_l}{C} = 1\frac{P_l}{C} = \frac{\left(2^{h-l+1} - 1\right)P}{C} = \frac{\left(2^{h-l+1} - 1\right)P}{C}$$

Similarly, the node at level $l$ will receive 2 aggregated packets from its pair of children, remaining active in reception mode for a time

$$T_{rxl} = 2\frac{\left(2^{h-l} - 1\right)P}{C} = \frac{\left(2^{h-l+1} - 2\right)P}{C}$$

And the time required for gathering data from all the sources comes from the summation of all intervals:

$$H = 2\sum_{l=1}^{h}\frac{P_l}{C_l} = \frac{1}{C}2\sum_{l=1}^{h}P_l = \frac{1}{C}2\sum_{l=1}^{h}\left(2^{h-l+1} - 1\right)P$$
$$= 2\left(2^{h+1} - 2 - h\right)\frac{P}{C}$$

### C. Compressed Sensing

The idea of applying compressive sampling theory to data gathering in WSNs was originally proposed in [5]. The approach substitutes individual sensor readings with a few weighted sums of all the readings, which are received and used by the gateway to restore original data. Hence, the sink receives a weighted sum of all the samples, and the process is repeated $M$ times using sets of different weights. At the end, the sink will receive $M$ weighted sums coming from the $N$ nodes of the network. Solving a set of $M$ linear equations with $N$ unknown variables is an ill-posed problem when $M < N$. However, in most cases, readings are not independent variables because of temporal and spatial correlation. Hence, there might exist a transform domain in which the signal is sparse. If the signal can be represented by $M$ non-zero coefficients – where $M << N$ – the signal is considered as $M$-sparse.

The gateway will receive $M$ packets from the right-hand side of three, and $M$ from the left-hand subtree.

$$n_0 = 2M$$

The collection is similar to those previously described in Section II-B, but the process is repeated M times. Hence,

$$N = M\left(2\left(h-1\right)+2\right) = 2hM$$

and each node will transmit a total of

$$\lambda_l = M$$

packets, all equally sized

$$P_l = P$$

Combining all these information, the transmission, reception on total schedule times can be easily computed:

$$T_{txl} = \lambda_l \frac{P_l}{C_l} = M\frac{P}{C}$$

$$T_{rxl} = 2M\frac{P_l}{C_l} = M\frac{P}{C}$$

$$H = N\frac{P_l}{C_l} = 2hM\frac{P}{C}$$

## III. Performance Analysis

We implemented the results of the previous Section in a numerical simulator to compare the performances of the three methods. Specifically, we focused on two main parameters: delay and energy consumption.

Delay is crucial for critical applications, where alarms need the be received within a defined deadline. Delay in our analysis is related to the $H$ value described above: the time required to complete the convergecast process is also the time at which the last sample is received by the gateway.

Figure 1 depict the normalized maximum delays associated to raw or aggregated convergecast against CDG using different values of $M$. Looking at the big picture, the delay with traditional convergecast methods tends to increase quickly with the addition of new levels in the tree. On the contrary,

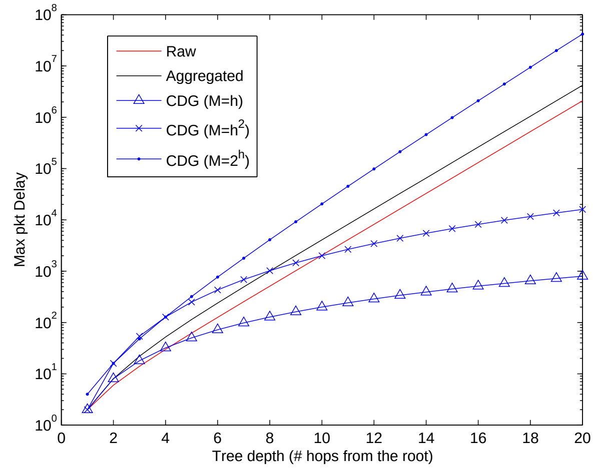


Fig. 1. Duration of the convergecast process.

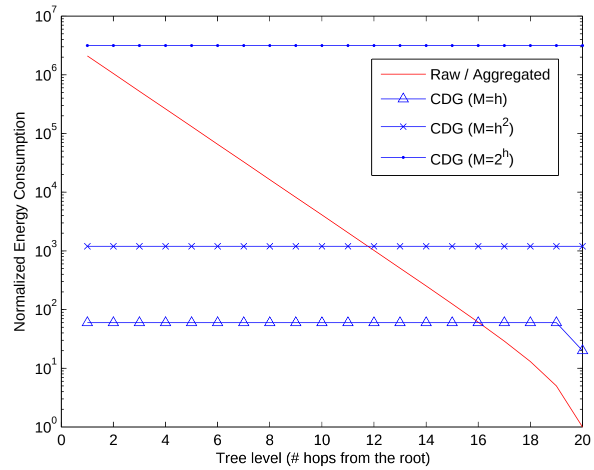


Fig. 2. Normalized energy consumption per bit.

compressed sampling can be even worst in some case, but might be the right choice for big networks if an appropriate value of $M$ is found.

Another crucial aspect for WSNs is lifetime. Different models for energy consumption in WSNs have been proposed in the years, but they do not agree each other: figures for power usage depend on too many factors. However, in the majority of these models, the communication module is usually accounted for most of the power consumption. In this analysis, we assume that the RF module consumes the same amount of energy for reception and for transmission of data. Figure 2 shows the normalized amount of energy consumed by nodes located in the tree at different levels.

According to our analysis, raw and aggregated convergecast have the same efficiency. Moreover, the so-called funneling effect [9] results in an increase in energy consumption when moving toward the gateway. Nonetheless, an increase in transit traffic intensity means that the nodes next to the gateway, which have heavy load, will die sooner. On the other hand, compressed sampling spread the consumption fairly throughout all the levels of tree. However, again, its efficiency is highly dependent on $M$. For most practical cases, given $M$ it is possible to identify a turning point: convergecast is the best approach below a certain level of the tree, while CDG becomes preferable for shorter trees.

## IV. Conclusion

Saving energy is crucial for prolonging the lifetime of sensor nodes, and delay is a fundamental metric for many critical WSN applications.

In this work, we proposed an analysis of three main techniques for performing many-to-one data collection in Wireless Sensor Networks, namely Raw Convergecast, Aggregated Convergecast and Compressed Data Gathering.

Preliminary results showed that raw and aggregated convergecast have similar performances both in terms of timeliness and power consumption. On the contrary, CDG exhibits high variability and its results highly depend on $M$. Since the value of $M$ is related to the monitored parameter and its dynamic, the choice of the technique used to collected data should be carefully carried on a case-by-case basis.

An extension of the comparison including real-wold datasets of measurements is under development. In future works, we will extend the analysis to different topologies, starting from N-ary trees.

## References


[1] S. Upadhyayula and S. Gupta, “Spanning tree based algorithms for low latency and energy efficient data aggregation enhanced convergecast (dac) in wireless sensor networks,” *Ad Hoc Networks*, vol. 5, no. 5, pp. 626 – 648, 2007.

[2] O. Durmaz Incel, A. Ghosh, B. Krishnamachari, and K. Chintalapudi, “Fast data collection in tree-based wireless sensor networks,” *Mobile Computing, IEEE Transactions on*, vol. 11, no. 1, pp. 86 –99, jan. 2012.

[3] O. Incel, A. Ghosh, and B. Krishnamachari, “Scheduling algorithms for tree-based data collection in wireless sensor networks,” in *Theoretical Aspects of Distributed Computing in Sensor Networks*, ser. Monographs in Theoretical Computer Science. An EATCS Series, S. Nikoletseas and J. D. Rolim, Eds. Springer Berlin Heidelberg, 2011, pp. 407–445.

[4] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak, “Compressed sensing for networked data,” *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 92 –101, march 2008.

[5] C. Luo, F. Wu, J. Sun, and C. W. Chen, “Compressive data gathering for large-scale wireless sensor networks,” in *Proceedings of the 15th annual international conference on Mobile computing and networking*, ser. MobiCom ’09. New York, NY, USA: ACM, 2009, pp. 145–156.

[6] E. Goldoni and P. Gamba, “Prism: A novel protocol for real-time synchronous acquisitions in wsns,” *International Journal of Sensors Wireless Communications and Control*, vol. 1, no. 2, pp. 80–87, 2011.

[7] H. Z. Pablo Soldati and M. Johansson, “Deadline-constrained transmission scheduling and data evacuation in wirelesshart networks,” in *Proc. of the European Control Conference (ECC)*,, aug 2009.

[8] V. Dimakopoulos and N. Dimopoulos, “Leaf communications in complete trees,” in *Euro-Par’96 Parallel Processing*, ser. Lecture Notes in Computer Science, L. Boug, P. Fraigniaud, A. Mignotte, and Y. Robert, Eds. Springer Berlin Heidelberg, 1996, vol. 1123, pp. 347–352.

[9] G.-S. Ahn, S. G. Hong, E. Miluzzo, A. T. Campbell, and F. Cuomo, “Funneling-mac: a localized, sink-oriented mac for boosting fidelity in sensor networks,” in *Proceedings of the 4th international conference on Embedded networked sensor systems*, ser. SenSys ’06. New York, NY, USA: ACM, 2006, pp. 293–306.