ISSN 2007-9737

# Impact of Aggregation and Compression on Cluster-Based Wireless Sensor Networks

Mario E. Rivero-Angeles[1], Izlian Y. Orea-Flores[2]

[1] Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Mexico

[2] Instituto Politécnico Nacional,
Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas,
Mexico

{mriveroa, iorea}@ipn.mx

**Abstract.** In Wireless Sensor Networks (WSNs), a clustered architecture is generally used to reduce energy consumption irrespectively of the application of the system. We prove in this work that, a clustered network only reduces energy consumption if aggregation or compression functions are enabled. Furthermore, a clustered network would consume much more energy if clustering techniques without the use of aggregation/compression (or a low aggregation coefficient) due to the extra consumption in the cluster formation phase. As an additional feature, a general energy consumption model based on the notion of energy units is developed that can be easily extrapolated to either theoretical or experimental values. Hence, the proposed analytical framework is valid for any commercial node or energy consumption model.

**Keywords.** Wireless sensor networks (WSN), aggregation, clustering, energy consumption.

## 1 Introduction

Nodes in a Wireless Sensor Network (WSN), are limited in resources, processing, and usually battery powered. As such, the operations they can perform are also limited and the system lifetime becomes a major performance metric to consider in the design of the network. To increase the system lifetime, clustering algorithms are usually implemented where nearby nodes are grouped to reduce the communication range of nodes, effectively reducing energy consumption.

Clustering schemes have two phases: a) Cluster Formation (CF) phase where nodes transmit a small packet to participate in the data reporting (usually using a random access protocol where collisions are possible and highly probable at the beginning of this phase) and b) Steady State (SS) phase where nodes are identified and assigned to specific clusters. In this phase, nodes in each cluster are assigned a specific schedule to orderly transmit their data to the cluster head in a collision-free manner. The time required for all CMs to transmit their data to their respective CH is called a *frame*.

For each cluster, a node is selected to become a cluster head (CH) while the rest of the nodes in that group become cluster members (CM). Cluster members transmit to the cluster head, which is typically closer than the sink node and consequently the power used to reach a CH is typically lower than the power used to reach the sink (the term *typically* is used since not all clustering algorithms guarantee such uniform CH distribution or not always).

The cluster heads then transmit the gathered information to the sink node. Hence only a few long-range (high energetic) transmissions are made compared to a non-clustered network where all nodes have to transmit their data using long-range transmissions.

Since CHs consume more energy than CMs, this role is re-assigned periodically in order to uniformly distribute the energy across the system. This time is commonly referred to as a *round*. In many works, the use of such clustered architectures is straightforward by assuming that energy consumption would be lower than in a non-clustered WSN irrespective of the aggregation/compression functions [1, 2, 3, 6-8].

Some works use aggregation or compression as an *additional* technique to further reduce energy without explicitly acknowledging that with no aggregation or compression, the cluster technique cannot reduce energy by itself [4, 5]. Indeed, many of these works overlook the use of aggregation or compression by either assuming high aggregation capabilities or by simply ignoring if these capabilities can even be used. Rather, these works usually focus on the adequate selection fo the nodes that can act as cluster heads at certain periods or the best procedure to form the clusters, among other open issues in this research area.

Aggregation functions are simple operations that the cluster head can perform such as obtaining the maximum or minimum values in the cluster or obtaining the average value among other possible operations [9, 10]. It is common to consider that the cluster head only transmits a single packet to the sink that comprises all the information produced in the cluster. Compression, on the other hand, is related to the set of operations that reduce or eliminate redundancy in the packets reducing also the packet size [11, 12, 13]. Hence, in this case, the cluster head sends a long packet to the sink, but shorter than the concatenation of the packets form all the CMs. Aggregation and compression can be used simultaneously in some applications.

Building on this, we intend to prove in this work that using clustering irrespective of the level of aggregation or compression is not efficient since in the SS, energy consumption per frame is very close to the case of direct transmissions to the sink, but adding the energy used to form clusters, the clustered architecture is by far much higher than the non-clustered case.

Furthermore, there are applications where aggregation cannot be used. For instance, consider the case where nodes are sending audio, video, or photographs. Then, the cluster heads that receive these files, cannot obtain the maximum, minimum, or average value of photographs or videos. Compression can be performed, but surely, the packet that the CHs transmit to the sink would be much higher than a single packet, commonly considered in cluster-based WSNs. Clustering, however, has many other benefits besides energy reduction, such as, scalability, fault tolerance, and load balancing, among others.

In this work, we focus on continuous monitoring WSNs, where all nodes transmit their data to their respective CH in each frame. As such, the energy consumption calculated in the following section is per frame, i.e., once the clusters are formed and the system is in the SS phase. Additionally, we assume that all nodes can reach the sink in a single transmission. Then, the surveilled area is restricted to applications where commercial nodes have a practical range to reach the sink node. For applications where the communication range to reach the sink node is not possible in a single hop, a multihop scheme should be considered which we believe falls outside the scope of this work since specific routing protocols have to be considered and each of these protocols has their particular functions and capabilities.

To this end, the analytical model is first described in detail. Then, we propose to use energy units that normalize the energy of any consumption model or even practical measurements without modifying the analytical framework presented in this paper. Then relevant numerical results are shown to prove that clustering *per-se* does not reduce energy compared to direct transmissions to the sink, unless some level of aggregation and compression is used.

## 2 Mathematical Model and Results

In the cluster formation phase, where all nodes have to send a short packet (mainly its ID and other relevant information such as energy level, and position among others) there is no aggregation process. As such, only the steady-state is relevant for the purpose of this paper.

In this work, an energy consumption model based on energy units is developed. These energy units can be easily converted into joules watts or any other unit required according to the commercial devices or the analytical model considered for each particular case. Specifically, it is assumed that:

— The energy consumed to transmit a packet inside a cluster is $E_{Tx}^{SS}$ energy units.

— A packet reception inside the cluster consumes $E_{Rx}^{SS}$ energy units.

— Nodes in the sleep mode, or low energy consumption mode, consume $(E_{sleep}$ energy units per time slot.

— Finally, the transmission of a packet from a cluster head to the sink consumes $E_{Tx}^{CH-Sink}$ energy units.

Building on this, the energy consumed in a frame, $E_f$, i.e., the period where all the cluster members send their data to its assigned cluster head can be expressed as follows:

$$E_f = [E_{Tx}^{SS} + E_{Rx}^{SS} + (E_{sleep} \times (N_{CM} - 1))]N_{CM}. \quad (1)$$

In this expression we can see that in each time slot, the corresponding CM sends it data, consuming $E_{Tx}^{SS}$ energy units, the CH that receives this packet consumes $E_{Rx}^{SS}$ energy units and the rest of the nodes ($N_{CM} - 1$ nodes) consume $E_{sleep}$ energy units. All cluster members transmit its packet in each frame.

Then, the cluster head forms the packet to be transmitted to the sink and consumes energy as follows:

$$E_s = [E_{Tx}^{CH-Sink} + (E_{sleep} \times N_{CM})]N_{CH}. \quad (2)$$

Indeed, each CH consumes the energy to transmit the packet while the rest of the nodes remain in the sleep mode. Note that the energy consumed by the sink is not considered since we assume that the sink is located in a strategic location with access to an electric outlet.

The total energy consumed by frame is then:

$$E_{total} = E_f + E_s. \quad (3)$$

Note that the aggregation and/or compression procedure is performed by the cluster head before transmitting the information that contains all the information of the cluster to the sink. Indeed, if the cluster head performs neither aggregation nor compression, then it has to transmit the complete packet from each of its cluster members including its packet to the sink.

On the other hand, if the cluster head performs a high amount of aggregation or compression or both, then the cluster head can send a single packet that encompasses all the relevant data from the cluster. To study the impact of the aggregation and compression the parameter $Agg$ $(0 \leq Agg \leq 1)$ is added to account for the level of compression or aggregation performed in the system as follows:

$$E_{Tx}^{CH-Sink} = Agg \times Pf \times (N_{CM} + 1) \times E_{bit}^{sink}, \quad (4)$$

where $Pf$ is the number of bits per packet and $E_{bit}^{sink}$ is the energy required to transmit a single bit from any node in the system to the sink. Note that, when no aggregation or compression is performed, $Agg = 1$. Conversely, when a high level of aggregation/compression is performed by the cluster heads, then $Agg < 1$. For the particular case when a single packet is transmitted, i.e., all information form the cluster members and the cluster head is encompassed in a single packet, then, $Agg = \frac{1}{N_{CM}+1}$.

Now, we illustrate the flexibility of the proposed mathematical analysis by considering the energy consumption model presented in [9]. Here, the authors consider that for each transmitted bit, there are two main energy consumption sources: The first one, $E_{elec}$, is due to the electronic circuits of the node and the second one, $E_{amp}$, is due to the communication system. The latter includes the use of the amplifier, coding and modulation functions among others, and the former includes the use of memory and processing functions. Hence, the energy consumption for each packet transmitted is given as follows:

$$E_{tx}(d, Pf) = (E_{elec} * Pf) + [E_{amp} * Pf * (d^l)], \quad (5)$$

**Table 1.** System Parameters

| Parameter | Value |
|---|---|
| Number of bits per packet ($Pf$) | 120 bits |
| Energy to power the electronic circuits per bit ($E_{elec}$) | $50 \times 10^{-9}$ [J] |
| Energy to transmit a bit ($E_{amp}$) | $100 \times 10^{-12}$ [J] |
| Energy to transmit a packet inside the cluster | 1 energy unit |
| Energy to receive a packet inside the cluster | 0.83 energy units |
| Energy to transmit a packet to the sink | 5625.8 energy units |
| Energy in sleep mode ($E_{sleep}$) | 0.01 energy units |
| Communication range inside a cluster ($d_c$) | 10 meters |
| Communication range to reach the sink node ($d_n$) | 150 meters |

where $d$ is the communication range, i.e., the distance between the transmitter and the receiver, and $l$ is the path loss coefficient. Without loss of generality, we consider that $l = 2$ for *in-cluster* communications where small communication distances occur ($d_c = 10$ meters) while $l = 4$ for long-range communication distances in the case of transmissions to the sink ($d_n = 150$ meters).

Note that the exact values of $l$ depend on the height of the antennas, the characteristics of the terrain, the height were the sink and the nodes are placed among many other parameters, the interested reader can read [14] for further details. For the reception process, the amplification power does not depends on the communication range. As such, a packet reception consumes the following energy:

$$E_{rx}(Pf) = (E_{elec} * Pf). \tag{6}$$

Building on this, we can now calculate the energy units used in (3) and (2) by normalizing to the energy consumed to send a bit inside the cluster (normalizing to any other parameter is also possible and would render the same results) as follows:

— The energy consumed to transmit a single bit inside a cluster (a cluster member communicating to its cluster head) is considered to be 1 energy units ($E_{Tx}^{SS} = \frac{E_{tx}(d_c,Pf)}{E_{tx}(d_c,Pf)}$),

— The energy to receive a packet inside a cluster is 0.83 energy units ($E_{Rx}^{SS} = \frac{E_{rx}(Pf)}{E_{tx}(d_c,Pf)}$),

— The energy to transmit a packet from any node to the sink is 5625.8 energy units ($E_{bit}^{sink} = \frac{E_{tx}(d_n,Pf)}{E_{tx}(d_c,Pf)}$).

It is important to note the flexibility of the energy consumption model proposed in (3) and (2) using energy units. For illustration purposes, we considered the model used in [9]. However, any other analytical model could have been used instead. Furthermore, practical measurements or commercial devices could have been easily used instead.

Indeed, by measuring the energy consumption in the transmission of a bit in a practical setting, or by considering the datasheets of a commercial node, it is straightforward to compute the energy values calculated above. In this work, the energy consumption in the sleep mode was calculated through multiple experiments using a Raspberry II node in the laboratory and then by normalizing the practical value by the energy to transmit inside a cluster, as explained above.

From this, it is now clear the benefits of the proposed energy consumption model based on energy units, where any other theoretical or experimental value can easily be introduced without modifying the proposed framework for calculating energy consumption per frame. Specifically, equations (2)-(6) are valid for any other device that consumes different amounts of energy or even considering more complex theoretical models that consider more variables, such as obstacles the height of the antennas, or obstacles in the communication trajectory for example.

As expected, the energy required to transmit a packet inside a cluster is lower than the energy required to receive a packet and considerably lower than the energy required to transmit a packet to the sink. In fact, this is the rationale for using a clustered architecture. However, it is a general conception that clustering by itself can reduce energy. In contrast, we argue in this work that, clustering does not reduce energy consumption if no aggregation or compression is used. Indeed, as we show later in this section, a cluster-based WSN with no aggregation/compression, with the conditions assumed in this work, consumes basically the same amount of energy than a non-clustered WSN where nodes transmit directly to the sink. To prove this statement, let us first derive the energy consumption in such non-clustered WSN.

For the WSN with $N$ nodes where no clustering is used, all nodes orderly transmit directly to the sink node. The aforementioned schedule can be provided by the sink at the beginning of the system operation just like the cluster heads assign the schedule to their cluster members at the beginning of the round.

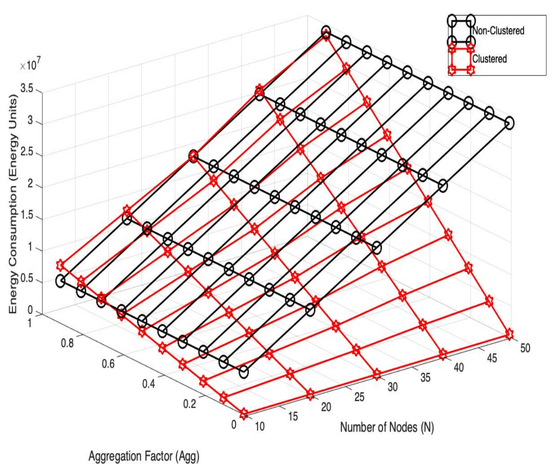


**Fig. 1.** Energy Consumption (Energy Units) per Frame for Different Number of Nodes and Aggregation Factor

Note that no aggregation can be done in this case, then, all the $N$ nodes transmit their complete packet to the sink. As such, the energy consumption per frame can be calculated as:

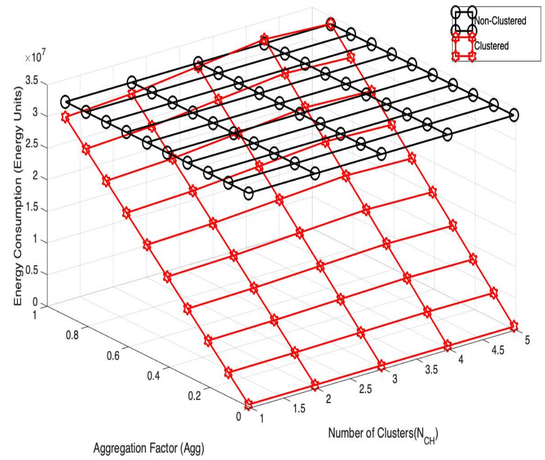


**Fig. 2.** Energy Consumption (Energy Units) per Frame for Different Number of Nodes and Clusters

$$E_{NC} = N * Pf * E_{bit}^{sink}. \quad (7)$$

For the next numerical results, the parameters shown in Table I were considered. Also, an ideal distribution of the cluster heads is assumed such that all clusters have the same number of cluster members. Many clustering algorithms that aim at providing such uniform distribution are found in the literature [1]. Considering a uniform CH distribution corresponds to the best-case scenario where the energy used for communications inside the clusters is minimized. For clustering schemes where this uniform distribution is not considered, the energy consumption would be higher. Hence, if there are $N_{CH}$ clusters in the system, then each cluster has $\lfloor N/N_{CH} \rfloor$.

Now, the energy consumption per frame for different number of nodes, aggregation factor, and the number of clusters is shown in Figs. 1 and 2 respectively. For the later, the number of nodes is fixed to 50, and for the former, the number of clusters is fixed to 3. From these results, it is clear that as the aggregation factor decreases (high amount of aggregation or compression), the energy consumption decreases drastically in the clustered architecture. However, when no aggregation or compression is used, i.e., the cluster heads send all the complete packets of their cluster members to the sink, the energy consumption is very close, or even higher than the non-clustered network. And the highest amount of energy reduction occurs for high aggregation factors, which led us to believe that the benefits of clustering are the possibility to perform aggregation or compression but not the clustering per se.

As such, in applications where no aggregation is allowed, i.e., the network administrator needs to obtain the complete values of all nodes in the system, the use of a clustered architecture is not advised or justified and even counter-advised since much energy is consumed in the cluster formation process.

Other relevant results are that the number of clusters where the system consumes the highest energy levels is 4, for all the configurations considered. And as expected, energy consumption increases when the number of nodes increases.

## 3 Conclusion

In this letter, the impact of the aggregation/compression in cluster-based WSNs is investigated. Through a simple mathematical analysis, it is proven that clustering in WSNs does not reduce energy consumption unless aggregation/compression capabilities are enabled at the cluster heads. Then, for certain applications where no aggregation is possible, for instance, if nodes are required to send photographs or video files to the sink, and obtaining the minimum, maximum, average or other similar parameter is not straightforward, then using a clustered structure does not provide any direct benefit in terms of energy reduction.

Conversely, for applications such as temperature, humidity, and pollution monitoring, among many others, it is clear that the cluster heads can easily obtain relevant information by enabling aggregation functions. Furthermore, high aggregation coefficients can be applied, obtaining and important energy reduction by using a clustered network.

## Acknowledgments

This paper was partially supported by IPN SIP project number 20200065.

## References

1. **Afsar, M.M., Tayarani-N, M.H. (2014).** Clustering in sensor networks: A literature survey. Journal of Network and Computer applications, Vol. 46, pp. 198–226. DOI: 10.1016/j.jnca.2014.09.005.

2. **Gupta, H.P., Rao, S.V., Yadav, A.K., Dutta, T. (2014).** Geographic routing in clustered wireless sensor networks among obstacles. IEEE Sensors Journal, Vol. 15, No. 5, pp. 2984–2992.DOI: 10.1109/JSEN.2014.2385734.

3. **Wu, X., Zeng, X., Fang, B., Yang, L., Zhang, W. (2017).** An energy-balance and game-theory-based cluster formation method for wireless sensor networks. International Journal of Distributed Sensor Networks, Vol. 13, No. 8, pp. 1550147717720792. DOI: 10.1177/1550147717720792.

4. **Aoun, B., Boutaba, R. (2006).** Clustering in WSN with latency and energy consumption constraints. Journal of Network and Systems Management, Vol. 14, No. 3, pp. 415–439. DOI: 10.1007/s10922-006-9039-4.

5. **Kumar, D., Aseri, T.C., Patel, R.B. (2011).** EECDA: Energy efficient clustering and data aggregation protocol for heterogeneous wireless sensor networks. International Journal of Computers Communications and Control, Vol. 6, No. 1, pp. 113–124. DOI: 10.15837/ijccc.2011.1.2206.

6. **Boubdallah, N., Rivero-Angeles, M.E. (2009).** Continuous monitoring using event-driven reporting for cluster-based wireless sensor networks. IEEE Transactions on Vehicular Technology, Vol. 58, No. 7, pp. 3460–3479. DOI: 10.1109/TVT.2009.2015330.

7. **Romo Montiel, E., Rivero-Angeles, M.E., Rubino, G., Molina-Lozano, H., Menchaca-Mendez, R., Menchaca-Mendez, R. (2017).** Performance Analysis of cluster formation in wireless sensor networks. Sensors, Special Issue: Smart Communication Protocols and Algorithms for Sensor Networks, Vol. 17, pp. 1–33. DOI: 10.3390/s17122902.

8. **Barrera-Figueroa, V., Rivero-Angeles, M.E., Menchaca-Mendez, R., Romo-Montiel, E., Menchaca-Mendez, R. (2018).** Design and performance analysis of segmented wireless sensor networks. IEEE Sensors Letters, Vol. 2, No. 4, pp. 1–4. DOI: 10.1109/LSENS.2018.2880676.

9. **Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H. (2002).** An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on Wireless Communication, Vol. 1, No. 4, pp. 660–670. DOI: 10.1109/TWC.2002.804190.

10. **Dagar, M., Mahajan, S. (2013).** Data aggregation in wireless sensor network: A survey. International Journal of Information and Computation Technology, Vol. 3, No. 3, pp.167–174. DOI: 10.1016/j.procs.2016.07.393.

11. **Srisooksai, T., Keamarungsi, K., Lamsrichan, P., Araki, K. (2012).** Practical data compression in wireless sensor networks: A survey. Journal of Network and Computer Applications, Vol. 35, pp. 37–59. DOI: 10.1016/j.jnca.2011.03.001.

12. **Incebacak, D., Zilan, R., Tavli, B., Barcelo-Ordinas, J.M., Garcia-Vidal, J. (2015).** Optimal data compression for lifetime maximization in wireless sensor networks operating in stealth mode. Ad Hoc Networks, Vol. 24, pp.134–147. DOI: 10.1016/j.adhoc.2014.07.019.

13. **Nguyen, M., Teague, K. (2016).** Energy-efficient data collection in clustered wireless sensor networks employing distributed DCT. arXiv preprint arXiv:1701.02606. DOI: 10.5121/ijwmn.2016.8501.

14. **Lee, W.CY. (1995).** Mobile cellular telecommunications: Analog and digital systems. McGraw-Hill Professional.

*Article received on 21/12/2020; accepted on 14/08/2021. Corresponding author is Mario E. Rivero-Angeles.*