Efficient data aggregation in wireless sensor networks: an entropy-driven analysis

(Invited Paper)

Laura Galluccio and Sergio Palazzo
DIIT, University of Catania
Catania, Italy
{Name.Surname}@diit.unict.it

Andrew T. Campbell
Computer Science Department, Darthmouth College
Hanover, New Hampshire
{campbell@cs.dartmouth.edu }

***Abstract*— Sensor networks are characterized by limited energy, processing power, and bandwidth capabilities. These limitations become particularly critical in the case of event-based sensor networks where multiple collocated nodes are likely to notify the sink about the same event, at almost the same time. The propagation of redundant highly correlated data is costly in terms of system performance, and results in energy depletion, network overloading, and congestion. Data aggregation is regarded as an effective technique to reduce energy consumption and prevent congestion. In this paper, we derive a number of significant insights concerning the data aggregation process, which have not been discussed in the literature so far. We first estimate the conditions under which aggregation is a costly process as compared to a no-aggregation approach, by considering a realistic scenario where processing costs related to aggregation of data are not neglected. We also consider that aggregation should preserve the integrity of data, and therefore, the entropy of the correlated data sent by sources can be considered in order to both decrease the amount of redundant data forwarded to the sink and perform an overall lossless process. Our framework can be used to investigate the tradeoff between the increase in data aggregation required to reduce energy consumption, and the need to maximize information integrity.**

Key Words: *Wireless Sensor Networks, Data Aggregation, Entropy, Energy Consumption.*

I. Introduction

Wireless sensor networks are composed of devices characterized by limited battery, processing and storage capabilities [1].

This issue becomes especially important in case of event-based applications, where sensors monitor a given phenomenon and send notifications and measurements back to one or more sink nodes. A significant amount of redundant data is likely delivered to the sink(s), particularly in case of dense networks, thus, wasting precious bandwidth and energy resources. Also, due to the so-called *funneling effect* [2], the closer a node is to the destination (i.e., the sink), the more demands are made on its energy resources and the higher is the traffic it has to manage and relay. For this reason aggregation in sensor networks is regarded as an effective technique to reduce energy consumption ([3], [4]) as well as preventing congestion [5]. This is also witnessed by the large amount of works investigating energy savings using data aggregation. As an example, in [4] an appropriate tradeoff between delay and energy is investigated in the context of a distributed estimation algorithm where the final result depends on the aggregation performed by few nodes. In [6] the impact of nodes' density on energy-efficient aggregation tree construction is also considered. In [7] interdependence between routing and data compression is also explored.

When performing aggregation, usually the cost of processing is disregarded. Instead, this could contribute to make the aggregation process even more costly than no-aggregation. Also one can expect that, the higher the number of packets aggregated, the higher the advantage of using aggregation with respect to not using it. However, aggregation cannot be increased indefinitely unless precious information is lost from the system. and attention should be paid to metrics used to fuse data together, thus avoiding to weigh multiple times the same readings leading to unreliable estimates at the sink(s). To make aggregation more efficient, in case of event-based applications, spatial correlation of data monitored by nodes in close proximity can be used.

Together with choice of the appropriate aggregation metric, also building an optimal data aggregation tree is a non trivial minimum Steiner tree NP-hard problem [8] for which approximated solutions have been recently proposed in the literature [9], [10], [11], [12].

However, in the previous literature, no emphasis is given to the way aggregation should be performed; also, aggregation is related to neither information integrity nor delay constraints. The main focus of this paper is on determining a tradeoff between the need to reduce the power consumption and preserving information integrity in the aggregation process. Accordingly, a set of conditions under which aggregation can be less expensive than no aggregation is identified. The latter can be used by network designers to appropriately investigate the conditions when aggregation can increase system performance and to design network protocols able to increase network lifetime and fidelity in data delivery. The rest of this paper is organized as follows. In Section II a power consumption analysis is developed which allows to derive the conditions when aggregation is a costly process with respect to no-aggregation. Also, the maximum value of the aggregation function which guarantees to preserve data integrity at destination is estimated. In Section III some performance results are discussed and, finally, in Section IV some conclusions and considerations on future work are drawn.

II. System Model

Let us consider an event-based application scenario. In this case, nodes are required to monitor unexpected events, e.g., fire, carbon-monoxide density or degree of contamination exceeding a given threshold, and report the monitored information to the sink(s) which disseminate(s) queries through an interest-like [3] approach. We also assume that routing trees have been established to deliver data from a sensor source to the sink.[1] Observe that, when different metrics are monitored throughout the network, only homogeneous data will be aggregated. So if a node manages different types of data, we assume it can aggregate packets carrying the same data items, e.g., only pollution notifications or only earthquake strength measurements.

We consider moderate-to-heavy traffic loads so that the Kleinrock independence approximation [13] reasonably holds. This is definitely a realistic assumption in sensor networks for environmental monitoring where conditions suddenly change in the area of interest and most of the sensors, at least for a certain time interval, keep on transmitting new data following the detected event. Let us suppose that a node, say k, receives upstream traffic from N_k one-hop neighbor nodes. This traffic needs to be relayed to other nodes towards the sink(s). In order to reduce energy consumption, thus increasing network lifetime, packets sent by these N_k nodes could be effectively aggregated so as to take advantage of the correlated traffic generated by high density sensor networks when nodes are located in the proximity of each other. In fact, the correlation coefficient between traffic generated by a pair of nodes is a function of their physical proximity: close nodes are likely to produce more correlated traffic.

If a node k performs aggregation, it will be denoted hereafter as *aggregator*. Aggregation will clearly result in a reduction in the energy which is depleted by relaying redundant data. However, aggregation implies performing more complex operations with respect to simply relaying traffic; this can lead to an increase in the overall energy consumption.

In the next sections, we develop a power consumption analysis and an entropy estimation for this model. More specifically, in Section II-A we will investigate the conditions when aggregation is a costly process in terms of energy as compared to no aggregation; then, in Section II-B, we will develop a framework for characterizing when aggregation can be performed while preserving the integrity of the information delivered to the sink(s).

A. Power Consumption Analysis

Let us evaluate the power consumption, $\mathcal{C}_k$, at node k, comparing both the case when this node acts as a forwarder only, i.e. simply relaying other nodes' packets and/or generating its own traffic, and the case it acts as an aggregator. In the first case,

$$\mathcal{C}_k = (P_{TX} + P_{RX}) \cdot \sum_{j=1}^{N_k} \lambda_j + P_{TX} \cdot r_k \qquad (1)$$

where

- λ_j is the packet emission rate at each of the j one-hop neighbors of the considered forwarder node, k;
- P_{TX} and P_{RX} are the packet transmission and reception powers, respectively;
- r_k is the rate of data generated by node k.

In the case when node k is an aggregator, the power consumption can be written as follows

[1]Though different approaches can be used in building trees, the analysis made in the rest of the paper is independent of how each specific tree has been established. Likewise, the analytical derivation is no way affected by the particular distributed algorithm that nodes use to exchange information needed to estimate the power consumption and the information entropy associated to aggregation.

$$\mathcal{C}_k = \left(\frac{P_{TX}}{\alpha_k(t)} + P_{Cod}\right) \cdot \left(\sum_{j=1}^{N_k} \lambda_j + r_k\right) + P_{RX} \cdot \sum_{j=1}^{N_k} \lambda_j + P_{Circ} \tag{2}$$

where

- P_{Circ} is the power needed to keep on the aggregation circuitry, independently of the amount of data to be aggregated;
- $\alpha_k(t)$ is the *aggregation function* at node k, where $\alpha_k(t) \geq 1$ represents the number of packets whose information can be aggregated by node k into one packet, at time t: different choices are possible since $\alpha_k(t)$ could either remain constant and be not reactive to network dynamics, or be variable according to the current network status;
- P_{Cod} is the power employed for coding and processing a single packet to perform data aggregation.

The additional power needed to perform aggregation cannot be neglected. In fact, typically, the current drain at a sensor device depends on the amount of instructions being performed. More specifically, the higher is the amount of data to be aggregated, the higher the number of instructions to be performed, the higher the power consumption. As an example, typical values are 650 μA/MIPS at 2V for a Microchip PIC24FJ64GA004 sensor device [14].

From comparison of the cost in eqs. (1) and (2) we observe that, when

$$\frac{P_{Cod}}{P_{TX}} < 1 - \frac{1}{\alpha_k(t)} - \frac{P_{Circ}}{P_{TX}} \cdot \frac{1}{r_k + \sum_{j=1}^{N_k} \lambda_j} \tag{3}$$

the process of transmission without aggregation results more costly in terms of energy consumption as compared to the case when aggregation is used.

Let us also consider the total power consumption $\mathcal{C}_T$ per transmitted packet at nodes upstream a generic node towards the sink [2]. The node k is supposed to be H hops far away from the sink. As discussed above, we assume that a tree-path from each node to the sink is available and, consequently, we take into account both the case when there are no other aggregator nodes upstream, along the path to the sink, and the case when there are A aggregator nodes. In the first case, the total power consumption per each packet transmitted by k is related to the transmission and reception power employed by each of the $(H-1)$ nodes along the path to the sink, and can be evaluated as

$$\mathcal{C}_T = (P_{TX} + P_{RX}) \cdot (H-1) \tag{4}$$

In the second case, instead, we have to consider the transmission power at each of the $(H-1-A)$ nodes along the path to the sink which are not aggregators, the reception power at each of the $(H-1)$ nodes, the transmission power at the A aggregators, and the coding power required by the A aggregators.

Consequently, if follows that

$$\begin{aligned} &\mathcal{C}_T = (P_{TX} + P_{RX}) \cdot (H-1) + A \cdot P_{Cod} + \\ &+ \textstyle\sum_{i=1}^{A} \frac{P_{TX}}{\alpha_i(t)} - A \cdot P_{TX} + A \cdot P_{Circ} \end{aligned} \tag{5}$$

From comparison of the cost in eqs. (4) and (5) we can figure out that when

$$\frac{P_{Cod}}{P_{TX}} < \left(1 - \frac{1}{A} \cdot \sum_{i=1}^{A} \frac{1}{\alpha_i(t)} - \frac{P_{Circ}}{P_{TX}}\right) \tag{6}$$

transmission of a single packet is more costly when no other aggregators are met upstream towards the sink than when A aggregators can be found.

Accordingly, if we want to satisfy both eqs. (3) and (6)

$$\begin{aligned} &\tfrac{P_{Cod}}{P_{TX}} < min\left\{1 - \tfrac{1}{A} \cdot \textstyle\sum_{i=1}^{A} \tfrac{1}{\alpha_i(t)} - \tfrac{P_{Circ}}{P_{TX}}, 1 - \tfrac{1}{\alpha_k(t)} + \right. \\ &\left. - \tfrac{P_{Circ}}{P_{TX} \cdot (r_k + \sum_{j=1}^{N_k} \lambda_j)}\right\} \end{aligned} \tag{7}$$

Thus, in this condition, performing aggregation results more energy efficient than not aggregating at all. According to (7) a node which is assumed to know the identities of aggregators available along its path to the sink and their aggregation function values, can foresee if the aggregation process will be energy efficient or not. Furthermore, if many aggregation trees are available, eq. (7) can be used to allow a node to properly select one path with respect to another one so as to make the aggregation performed by the upstream nodes an energy efficient process.

Looking at eqs. (2) and (5), it is evident that, the higher the value of the aggregation function, the lower the amount of power needed to perform transmission. However, higher aggregation could be costly in terms of loss of information. So, in the next section, we will analyze in detail to what extent aggregation can be performed preserving information integrity.

[2]Nodes upstream w.r.t. a node k are nodes met along the path going from node k up to the sink.

B. Entropy estimation

In this section we focus on the information aggregation process which can be performed at an aggregator node, so as to reduce the amount of data traveling towards the sink, while preserving the integrity of the information.

According to the goal being pursued, on the one hand the aggregation process should be lossless because no information has to be lost; on the other hand, the process should be such that useless data packets are not worth being forwarded.

By using aggregation, we expect that node's k emission rate, $\lambda_k(t)$, should be lower than $\sum_{j=1}^{N_k} \lambda_j(t)$, i.e.

$$\lambda_k(t) \leq \sum_{j=1}^{N_k} \lambda_j(t) \tag{8}$$

To evaluate $\lambda_k(t)$ let us model the data generated by each neighbor of node k as a time-continuous random variable X_j.

In order to appropriately design the data aggregation process at a generic node k, we preliminarely evaluate the differential joint entropy of the variable $\Gamma = (X_1, \dots X_{N_k})$, namely $h(\Gamma) = h(X_1, \dots X_{N_k})$, which differs from the entropy in that the random variables are not required to be discrete but are supposed to be continuous.[3]

As a property of the differential joint entropy, it follows that [15]

$$h(X_1, \dots X_{N_k}) \leq \sum_{j=1}^{N_k} h(X_j) \tag{9}$$

Accordingly, in a lossless process, a rate $r \geq h(X_1, \dots X_{N_k})$ is sufficient to accurately reconstruct variables X_1, X_2, $\dots X_{N_k}$. By recalling well known results from the information theory [15], $h(X_1, \dots X_{N_k})$ can be evaluated as $h(\gamma) = \int_{-\infty}^{+\infty} f_\Gamma(\gamma) \cdot log_2(\frac{1}{f_\Gamma(\gamma)}) d\gamma$ where

$$f_\Gamma(\gamma) = \frac{1}{2\pi^{N_k/2}\sqrt{detV}} \cdot e^{\left[-\frac{(\gamma-m)V^{-1}(\gamma-m)^T}{2}\right]} \tag{10}$$

being m the array of the average values for each of the assumed N_k Gaussian random variables and V the matrix of the covariances.

This result can be used for estimating the maximum improvement in terms of reduction in the amount of useless data that have to be transmitted using aggregation with respect to no-aggregation, provided that no information is lost. The improvement at the node k, denoted as θ_k, is given by the ratio between the difference in the data to be relayed by node k when no-aggregation is performed and when aggregation is applied and the total amount of data to be relayed with no-aggregation performed. More specifically,

$$\theta_k = \left[\sum_{j=1}^{N_k} h(X_j) - h(X_1, X_2 \dots X_{N_k})\right] / \sum_{j=1}^{N_k} h(X_j) \tag{11}$$

Whatever is the choice in the aggregation function, in order not to loose any information, $\alpha_k(t)$, i.e. the number of data packets which can be aggregated with no information reduction, should be not higher than the ratio between the sum of the single nodes' differential entropies and the joint differential entropy. In other words, it should be:

$$\alpha_k(t) \leq \frac{1}{1-\theta_k} \tag{12}$$

An increase in $\alpha_k(t)$ beyond this threshold would result in a loss of information. Therefore, in order to be aggressive in the process so as to reduce energy consumption while also preserving information integrity, the maximum value of the aggregation function at node k should be such that

$$\alpha_k^{MAX} = \frac{1}{1-\theta_k} \tag{13}$$

Finally, eq. (8) can be rewritten as

$$\lambda_k(t) = (1-\theta_k) \cdot \sum_{j=1}^{N_k} \lambda_j(t) \tag{14}$$

where the term $(1-\theta_k)$ takes into account the ratio between the data which should be sent using aggregation and those that are expected without aggregation.

III. Case Study

In this section, we investigate the joint differential entropy and, consequently, the maximum amount of aggregation which can be applied compatibly with the need of minimizing energy consumption while preserving information integrity. For this reason in Fig. 1(a) we show the values of the differential joint entropy obtained when considering a node k with a number of one-hop neighbors equal to $N_k = 2$. In this figure the mean value of the sensor measurements is denoted as m_X, and the measurements are assumed to have a

[3]Differential entropy represents the extension of the entropy concept, i.e. a measure of the degree of "surprise" associated to a random variable, to the case of continuous variables [15].

standard deviation σ_X, equal for all neighbors. Also, as soon as the number of one-hop neighbors increases, the improvement θ_k increases because the difference between the joint differential entropy and the sum of the single sources' entropies, as a function of the ratio σ_X/μ_X, increases. This is also evident when looking at Fig. 1 (b), where the maximum value of the aggregation function is shown.

Results in Fig. 1 have been obtained assuming a generic random topology and considering that sensor measurements have a Gaussian distribution. Furthermore, the covariance between two measurements is a decreasing function of the distance between the pair of nodes who generated them, as done in [16], i.e.

$$\sigma_{J_i,J_j} = Ae^{-\beta \cdot d_{i,j}} \tag{15}$$

As expected, in Fig. 1(b) it can be seen that, as soon as the number of neighbor nodes increases, a higher aggregation can be applied which allows to reduce the amount of redundant data packets sent by the aggregator node to the sink. Observe that an increase of one in the number of neighbors, allows to almost double the maximum value of the aggregation function.

In Fig. 2(a), we show the normalized power consumption at node k as a function of the normalized coding power and the number of neighbors, N_k. In this figure, which has been obtained assuming $\lambda_j = 2$ pkts/s, $P_{Cod} = P_{Circ}$, $r_k = 1$ pkts/s and $\alpha_k(t) = 2$, we show the region described in eq. (3). More specifically, the region where the dark curve overcomes the light curve is the one where performing aggregation results cheaper than not aggregating.

This region is met as soon as the ratio P_{Cod}/P_{TX} is approximately above 0.4. This is because when the coding cost becomes comparable to about half of the transmission cost, aggregation starts to become a costly process. When the aggregation function, $\alpha_k(t)$, increases, the region where aggregation is more energy efficient than no-aggregation increases as well. This figure can be used in conjunction with Fig. 1 for design purposes. In Fig. 2(b) we show the results derived in eq. (7) when considering that $\alpha_i(t) = \alpha \ \forall i$, $r_k = 1$ pkts/s, and $\lambda_j = 2$ pkts/s $\forall j$. Looking at the above figure we observe that, the higher is the aggregation value α, the higher the ratio $(P_{Cod}+P_{Circ})/P_{TX}$ and, thus, the lower the value of P_{TX} which allows to achieve both energy consumption reduction and decrease in the amount of data traveling throughout the network, while preserving integrity of data.

IV. Conclusions and Future Work

In this paper, we have investigated the use of data aggregation for improving energy efficiency in high density wireless sensor networks for event-monitoring applications. Under such conditions nodes are likely to send multiple correlated data to the sink, thus causing the propagation of redundant information throughout the network which in turn leads to both a waste of energy resources and bandwidth, and increase in network congestion. Aggregation is however a costly mechanism because additional processing is required which could imply, under certain conditions, higher power consumption with respect to traditional forwarding of data. Also, aggregation should preserve the data integrity, as the higher is the aggregation, the higher is the risk to miss important information. We have developed some analysis for the evaluation of the power cost of the aggregation process with respect to not performing aggregation. We have estimated the joint entropy of the correlated information sent by different sources; this has allowed us to determine a tradeoff between the need to perform an aggregation process which is energy efficient and lossless in the same time. The results discussed in this paper can be used to design appropriate aggregation processes which both preserve the integrity of the information and reduce energy consumption. Currently, we are working on a testbed implementation to compare analytical and experimental results. We are also studying the applicability of the work to multi-sink scenarios.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Comm. Magazine*, vol. 40, no. 8, Aug. 2002.

[2] G. 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," *ACM Sensys 2006, Boulder, CO*, Nov. 2006.

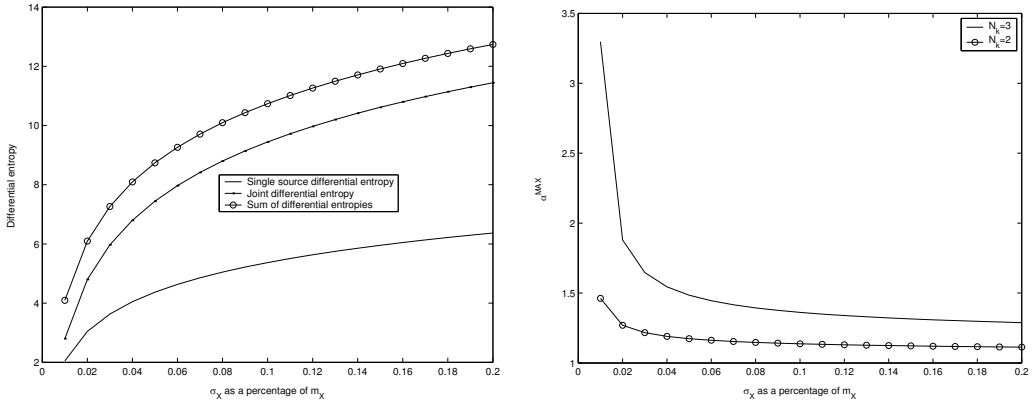
[3] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidenmann, and F. Siva, "Directed diffusion for wireless sensor networking," *ACM/IEEE Trans. on Netw.*, vol. 11, no. 1, Feb. 2002.

[4] A. Boulis, S. Ganeriwal, and M. Srivastava, "Aggregation in sensor networks: an energy-accuracy tradeoff," *IEEE SNPA 2003, Ankorage, AL*, May 2003.

[5] L. Galluccio, A. T. Campbell, and S. Palazzo, "Concert: aggregation-based congestion control for sensor networks," *ACM Sensys, San Diego, CA*, Nov. 2005.

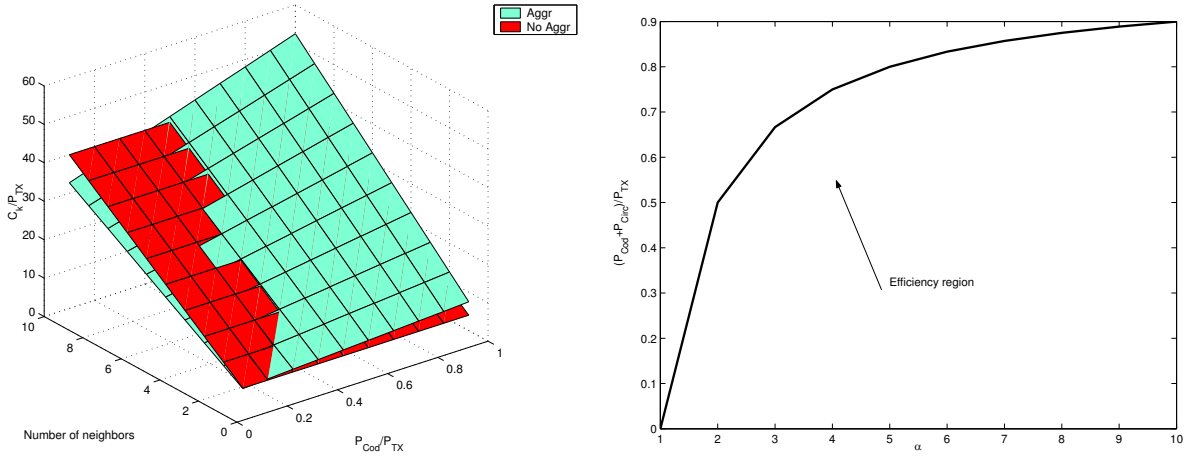
[6] C. Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidenmann, "Impact of network density on data aggregation in wireless sensor networks," *IEEE ICDCS 2002, Wien, Austria*, Jul. 2002.

[7] A. Scaglione and S. Servetto, "On the interdependence between routing and data compression," *ACM Mobicom 2002, Atlanta, GA*, Sept. 2002.



(a) (b)

Fig. 1. (a) Differential entropy when node k has two neighbors, i.e. $N_k = 2$. (b) Maximum value of the aggregation function as a function of the ratio between σ_X and m_X.



(a) (b)

Fig. 2. a) Normalized power consumption at node k in case of aggregation and no aggregation as a function of the number of neighbors and the normalized coding power when $r_k = 1$ pkts/s, $\mu_j = 2$ pkts/s $\forall j$ and $\alpha_k = 2$ pkts. b) Ratio $(P_{Cod} + P_{Circ})/P_{TX}$ as a function of the aggregation value, α, assumed constant at all nodes.

[8] B. Krishnamachari, D. Estrin, and S. Wicker, "The impact of data aggregation in wireless sensor networks," *DEBS, Wien, Austria*, Jul. 2002.

[9] D. Vass and A. Vidacs, "Distributed data aggregation with geographical routing in wireless sensor networks," *IEEE ICPS 2007, Istanbul, Turkey*, Jul. 2007.

[10] X. Su, "A combinatorial algorithmic approach to energy efficient information collection in wireless sensor networks," *ACM Trans. on Sens. Netw.*, vol. 3, no. 1, Mar. 2007.

[11] H. Gupta, V. Navda, S. R. Das, and V. Chowdhary, "Efficient gathering of correlated data in sensor networks," *ACM Mobihoc 2005, Urbana-Champaign, IL.*, May 2005.

[12] R. Cristescu, B. Beferull-Lozano, M. Vetterli, and R. Wattenhofer, "Network correlated data gathering with explicit communication: *np* completenss and algorithms," *ACM/IEEE Trans. on Netw.*, vol. 14, no. 1, Feb. 2006.

[13] L. Kleinrock, *Queueing Systems, Volume II: Computer Applications*. J. Wiley and Sons Eds, 1976.

[14] "http://www.microchip.com/paramchartsearch/."

[15] S. S. Haykin, *Communication Systems*, 4th ed. J. Wiley and Sons Eds.

[16] M. C. Vuran and I. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," *ACM/IEEE Trans. on Netw.*, vol. 14, no. 2, Apr. 2006.