センサーネットワークにおけるデータ集約の性能解析モデル

トビアス ホスフェルト† ライプニッツ 賢治††

† University of Würzburg, Insitute of Computer Science
Am Hubland, 97074 Würzburg, Germany
†† 大阪大学 大学院情報科学研究科 先進ネットワークアーキテクチャ講座
〒 565-0871 大阪府吹田市山田丘1-5
E-mail: †hossfeld@informatik.uni-wuerzburg.de, ††leibnitz@ist.osaka-u.ac.jp

あらまし 本稿では、無線センサネットワークにおいて、データ集約を行うセンサ端末の解析モデルを提案する。各センサ端末はセンシング装置と無線送受信機を持ち、限られた容量のバッテリで駆動する。センサネットワークでは、電力消費の削減が重要な課題である。そのため、各センサ端末は、他のノードから受信したセンシング情報をバッファリングし、上流のノードからの情報を全て受信した後に、その情報を集約して次のセンサ端末に転送する。本稿では、マルコフ連鎖を用いて、データ集約方式の性能解析モデルを提案する。
キーワード センサーネットワーク, データ集約, マルコフ連鎖モデル, 性能解析

Performance Modeling of Data Aggregation in Wireless Sensor Networks

Tobias HOSSFELD† and Kenji LEIBNITZ††

† University of Würzburg, Insitute of Computer Science
Am Hubland, 97074 Würzburg, Germany
†† Osaka University, Graduate School of Information Science and Technology,
1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
E-mail: †hossfeld@informatik.uni-wuerzburg.de, ††leibnitz@ist.osaka-u.ac.jp

Abstract In this paper we propose an analytical model of a node in a wireless sensor network, which is performing data aggregation and forwarding. Each sensor node is equipped with a sensing device and RF transceiver, as well as a battery unit of only limited capacity. Due to its stringent energy consumption requirements, received sensing data from other nodes is buffered at the considered node and transmitted to its next hop toward the sink only after a batch of packets from all its upstream nodes has been received. We derive an analytical model for investigating data aggregation strategies using a Markov chain analysis.
Key words Sensor networks, data aggregation, Markov chain model, performance analysis

1 Introduction

In the recent years, wireless sensor networks (WSN) [1] consisting of sensor nodes have become readily available as off-the-shelf products. Usually, a sensor node consists of the actual sensing unit, an RF transceiver, and is powered by a battery, see Fig. 1. Since the nodes are conceived as low-cost devices, they are deployed in large numbers with hundreds or thousands of nodes performing their respective task, such as monitoring the environment (e.g. temperature, humidity, wildlife), intrusion detection, or other applications within an ambient information infrastructure.

Conservation of energy is due to the limited power source a major issue as it directly translates into the lifetime of the entire network before an operator can manually replace expired nodes with new ones. In order to reduce the consumed energy, sensor nodes often employ a sleep scheduling mechanism in which redundant nodes switch to a sleep mode which requires significantly less energy [2]. The most power is required during RF transmissions [3] and, therefore, also data aggregation policies have been proposed, in which the limited computation power of the sensor nodes is utilized to preprocess all data packets a node receives from its neighbors prior to forwarding. Depending on the considered application, the information received from the node's neighbors nodes is usu-

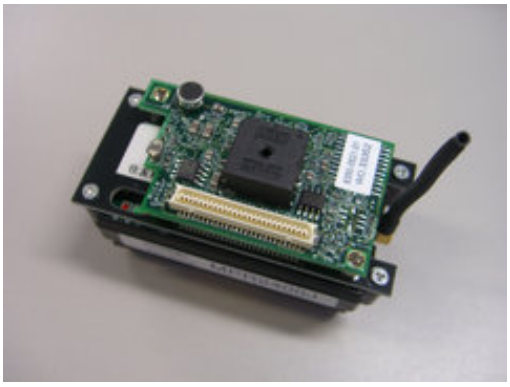

Fig. 1 Wireless MicaZ sensor mote

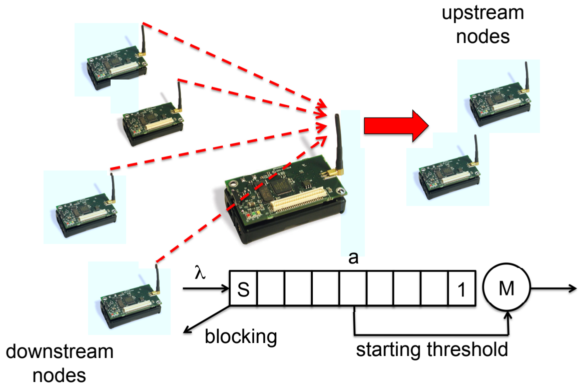

Fig. 2 Considered sensor network scenario

ally correlated with that of its own data and compression techniques can be used to reduce the amount of data transmitted over the wireless channel. For instance, if a cluster of nodes is monitoring the temperature within a room, the measurements will all yield very similar values lying around a mean, which may be sufficient for forwarding and which is updated on each hop along the packet's path to the sink.

In this paper we provide a queuing theoretic framework for analyzing the data aggregation in a sensor network. We only consider a single arbitrary node in this work, which receives data from its downstream neighboring nodes, performs data aggregation when it receives a sufficient number of data packets, and then forwards this data to its upstream neighbors toward the sink node. The motivation of using data aggregation is to reduce the energy consumption used for wireless transmission by fusing data packets into one single packet. The underlying assumption is that processing the data packets requires much less energy than transmission. Especially, in scenarios with high density of nodes, due to the employment of CSMA/CA in IEEE 802.15.4 the number of collisions on MAC layer will drastically improve, so reducing the number of transmissions may have a highly beneficial impact [4].

The remainder of this paper is organized as follows. We first discuss our envisaged scenario in greater in detail in Section 2 and also briefly summarize some existing work on data aggregation in WSN. Following that, we introduce our analytical model in Section 3, which is supplemented by numerical evaluations. Finally, Section 4 concludes this paper with an outlook on future work.

2 Data Gathering and Aggregation in Sensor Networks

In this paper we do not restrict ourselves to any specific application of WSN, but simply assume that a generic data gathering scheme is applied. This section summarizes the scenario we are considering and briefly discusses related work.

2.1 Scenario Description

Each node reads at certain time intervals the measurement data from its sensory unit and keeps this data in its buffer together with other measurement samples it receives from neighboring nodes. Once the considered node has collected enough data or a timer expires, it broadcasts an aggregated data packet to its upstream neighbors toward the sink node, see Fig. 2. In this paper, we assume that the node is aware of the topology of its surrounding nodes by prior exchange of messages and for the sake of simplicity omit the timer in the following analysis.

In this way, data is propagated in a hop-by-hop manner from the boundary of the network toward the sink node. The application of data aggregation is beneficial on the energy consumption as previous studies have indicated that the energy required for fusing is at about 5 nJ/bit, whereas that for transmission and reception lies about 10 times higher [5].

Finally, it should also be remarked that the buffer size at each sensor node is rather small due to hardware limitations. In the case that a batch arrives that can not be fully stored in the buffer, the buffer is filled until its capacity is reached and the remaining packets are discarded. All packets are also discarded if the buffer is found full upon arrival.

2.2 Related Work

Data aggregation in WSN has been studied extensively in the past. At this point, we would like to discuss some of the papers that are relevant to our investigation. In [6], the authors consider a data-centric sensor network in contrast to traditional end-to-end routing schemes. A heuristic approach is used to construct a data aggregation tree and the impact of the node placement is studied in terms of energy costs and delay. In [7], security is taken into account in sensor networks and SDAP (Secure Hop-by-hop Data Aggregation Protocol) is proposed, which partitions the aggregation tree into groups for reducing the importance of high-level nodes in the aggregation tree.

The paper by Fan et al. [8] provides a nice overview of

$$
Q = \begin{pmatrix}
d_0 & d_1 & d_{i-b-1} & d_{i-b} & d_a & d_j & d_{S-b-1} & d_{S-b} & \sum_{k=S}^{\infty} d_k \\
d_0 & d_1 & d_{i-b-1} & d_{i-b} & d_a & d_j & d_{S-b-1} & d_{S-b} & \sum_{k=S}^{\infty} d_k \\
d_0 & d_1 & d_{i-b-1} & d_{i-b} & d_a & d_j & d_{S-b-1} & d_{S-b} & \sum_{k=S}^{\infty} d_k \\
d_0 & d_1 & d_{i-b-1} & d_{i-b} & d_a & d_j & d_{S-b-1} & d_{S-b} & \sum_{k=S}^{\infty} d_k \\
d_0 & d_1 & d_{i-b-1} & d_{i-b} & d_a & d_j & d_{S-b-1} & d_{S-b} & \sum_{k=S}^{\infty} d_k \\
0 & d_0 & d_{i-b-2} & d_{i-b-1} & d_{a-1} & d_{j-1} & d_{S-b-2} & d_{S-b-1} & \sum_{k=S}^{\infty} d_k \\
0 & 0 & 0 & d_0 & d_{a-i+b} & d_{j-i+b} & d_{S-i-1} & d_{S-i} & \sum_{k=S}^{\infty} d_k \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & d_0 & \sum_{k=S}^{\infty} d_k
\end{pmatrix} \tag{1}
$$

various data aggregation strategies, ranging from tree-based or cluster-based aggregation and they deviate from other approaches by considering a structure-free method, which combines a spatial convergence of packets meeting at a certain node utilizing Data-Aware Anycast with temporal convergence through Randomized Waiting. An entropy-based viewpoint of data aggregation is given in [9], which focuses on the transmission of redundant information and the joint entropy of the correlated information sent by different sources is estimated.

Most of the papers dealing with data aggregation consider a graph theoretic view, whereas we consider a queuing theoretic approach for the problem of data aggregation in WSN. The following section will provide a description of our proposed model.

3 Analytical Modeling of Data Aggregation

Let us now return to the scenario as shown in Fig. 2. Our considered node receives packets following a Poisson process with rate λ from all downstream neighboring nodes. Since each downstream neighbors may already have aggregated packets to send, the arrival process at our observed node consists of a superposition of Poisson streams with batch arrivals of packets that may be generally distributed. We assume that data aggregation can be modeled as a batch or bulk service process, where two thresholds trigger the service unit to become active for aggregation. The service time may be considered as general independent with threshold a at which the service starts and parameter b, which is the maximum size of the batch that can be simultaneously processed. Due to memory limitation requirements the queue length S is finite.

Models such as this can be found in the literature, e.g. in [10]～[12], and a nice overview of models with different arrival and service processes for bulk service queues with finite buffers is presented in [13]. In this paper, we extend the model by Gold and Tran-Gia [10] by taking also batch ar-

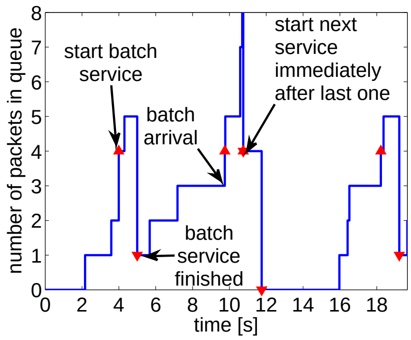

Fig. 3 Evolution of state space over time

rivals into account for the case of Poisson arrivals, which can be expressed in Kendall's notation as $M^G/G^{[a,b]}/1-S$.

An example of the state space dynamics of the $M^G/G^{[a,b]}/1-S$ system is shown in Fig. 3. At time step $t=4$, the buffer occupancy exceeds the service threshold of $a=4$ and service is started, which finishes by time $t=5$. Batch arrivals occur for example at time $t=4$ and $t=10$ and since the completion of the service at $t=11$ still results in a buffer occupancy of at least the threshold a, another service phase is instantly entered after the processing delay. In this specific example, we consider a deterministic processing time of 1 s.

In Section 3.1, we derive the corresponding steady state probabilities following the analysis in [10]. A simple example of how to apply the performed approach follows in Section 3.2 for the $M^{GEOM_1}/D^{[a,S]}/1-S$ queueing sytem. Finally, we investigate the paramater sensitivity of the general system with respect to key performance characteristics like sojourn time, blocking probability or probability of completely processed batches in Section 3.3.

3.1 Analysis of $M^G/G^{[a,b]}/1-S$

Let $X(t)$ denote the number of packets in the queue at time t. At the end of each service phase, a Markov chain can be embedded. The matrix Q of transition probabilities of this Markov chain follows that in [10] and is given in Eqn. (1).

The entries d_i correspond to the probabilities of having i arrivals during a service phase.

$$d_i = \int_{t=0}^{\infty} f_H(t) \sum_{j=1}^{i} a_j(t) \sum_{\substack{\{(i_1,\cdots,i_j):\\ \sum_{k=1}^{j} i_k = i \wedge i_k > 0\}}} \prod_{i_k} \mathrm{P}\left(N = i_k\right) \, dt \tag{2}$$

Here, the term $a_j(t) = \frac{(\lambda t)^j}{j!} e^{-\lambda t}$ is the probability of having j batch arrivals at time t, $f_H(t)$ is the probability density function of the service time for a batch, and N is the random variable for the number of total packet arrivals.

To obtain the steady state probabilities of the Markov chain can then be simply reduced to finding the eigenvalues of the matrix Q with respect to the normalizing condition in Eqn. (3).

$$\sum_{i=0}^{S} x(i) = 1 \tag{3}$$

Since an embedded Markov chain has been considered so far, the state probabilities are only valid at the regeneration instants when a service ends. However, using the Markov chain state probabilities, the state probabilities at an arbitrarily chosen observation epoch can also be derived. Let X^* denote the random variable for the number of packets in the queue at an arbitrary point in time. Then we use Eqns. (7)–(14) from [10] to obtain the joint probabilities for the number of packets in the queue for a specific service type at an arbitrary instant in time.

As we consider batch arrivals, we need to refine Eqn. (10) in [10] and derive the arrival probabilities during the forward recurrence time. Furthermore, we have for type $4i$ the following mean $EType$ in Table 1 of [10]: $\frac{a-i}{\lambda\, E[N_b]}$; the interval length may follow any distribution, but only the mean value are of interest for our analysis.

3.2 Example: $M^{GEOM_1}/D^{[a,S]}/1-S$

As an example, we assume now that the service time in the considered systems is deterministic and equals to $\bar{t}$. In the system, the service period is initialized when at least a packets are in the waiting queue. The service unit allows to process all packets in the queue, i.e., up to S packets can be processed simultaneously with S being the capacity of the entire waiting queue.

Let N_a denote the number of batch arrivals during a service time period $\bar{t}$ which follows a Poisson process with rate λ. Then, the probability for having i arrivals follows a Poisson distribution as well, i.e.,

$$\mathrm{P}\left(N_a = j\right) = \frac{(\lambda\bar{t})^j}{j!} e^{-\lambda\bar{t}} \,. \tag{4}$$

Let N denote the total number of packets arriving during a service time period $\bar{t}$. Further, let N_b denote the size of a single batch and it follows a geometric distribution with parameter q shifted by one. Hence, N_b can be described as the sum of a geometric distribution and a deterministic distribution with mean 1, $N_b \sim GEOM_1(q) = GEOM(q) + D(1)$. Then, for y batch arrivals during $\bar{t}$, the total number of packets is

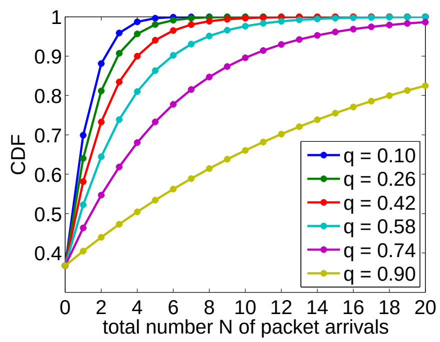

Fig. 4 Total number N of packet arrivals during a service period for various q in the $M^{GEOM_1}/D^{[a,S]}/1-S$ system

$$\begin{aligned} N &= \sum_y N_b \\ &= \sum_y GEOM(q) + D(1) \\ &= NEGBIN(y,q) + D(y) \end{aligned} \tag{5}$$

which is the sum of a negative binomial distribution with parameters y and q, and a deterministic distribution with mean y. The probability of i packets arriving within y batch arrivals follows as

$$\begin{aligned} \mathrm{P}\left(N = i \,|\, N_a = y\right) &= \sum_{j=0}^{i} x_1(i-j)\, x_2(j) = x_1(i-y) \\ &= \begin{pmatrix} i-1 \\ i-y \end{pmatrix} (1-q)^y q^{i-y} \end{aligned} \tag{6}$$

with the probability distribution function $x_1(i)$ for the $NEGBIN(y,q)$ and $x_2(i)$ for $D(y)$, respectively. Then, we obtain the probability distribution function of the total number of packets N within a service period as follows.

$$\mathrm{P}\left(N = 0\right) = \mathrm{P}\left(N_a = 0\right) = e^{-\lambda\bar{t}} \tag{7}$$

$$\begin{aligned} \mathrm{P}\left(N = i\right) &= \sum_{j=1}^{i} \mathrm{P}\left(N_a = j\right) \mathrm{P}\left(N = i \,|\, N_a = j\right) \\ &= \sum_{j=1}^{i} \frac{(\lambda\bar{t})^j}{j!} e^{-\lambda\bar{t}} \begin{pmatrix} i-1 \\ i-j \end{pmatrix} (1-q)^j q^{i-j} \end{aligned} \tag{8}$$

Figure 4 shows the cumulative distribution function of the total number of packets for $\lambda = 1\,{}^1\!/\!{}_s$ and $\bar{t} = 1\,\mathrm{s}$. These numerical values $d_i = \mathrm{P}\left(N = i\right)$ can now be used to generate the matrix Q and to derive the steady state probabilites as eigenvalues of Q.

In order to compute the state probabilities not only at the embedding points, but at arbitrary time instants, we need to compute the d_i^* values, corresponding to the probabilities of

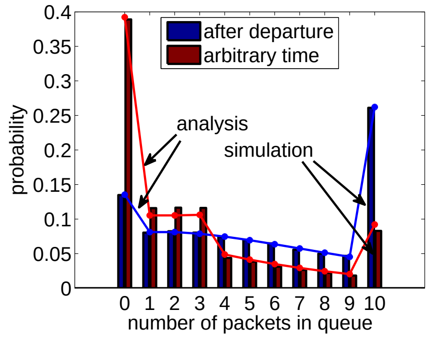

Fig. 5 Comparison of steady state probabilities from simulation and analysis

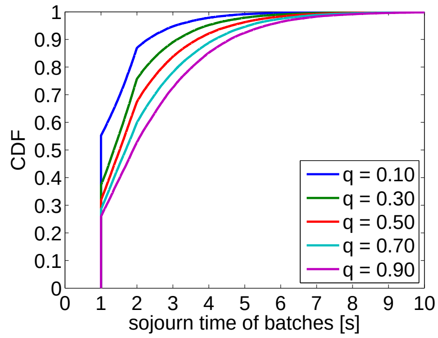

Fig. 6 Sojourn time of batches in the $M^{GEOM_1}/D^{[a,S]}/1-S$ system

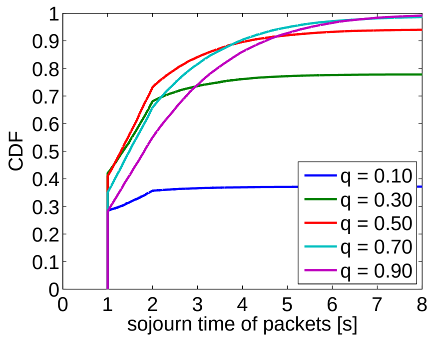

Fig. 7 Sojourn time of packets in the $M^{GEOM_1}/D^{[a,S]}/1-S$ system

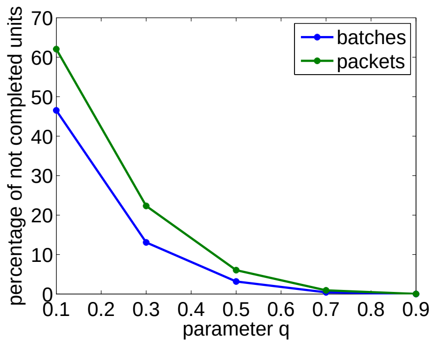

Fig. 8 Percentage of the not completed batches in the $M^{GEOM_1}/D^{[a,S]}/1-S$ system

arrival of packets during the forward recurrence time. This is given in Eqn. (9).

$$\begin{aligned} d_i^* &= \sum_{j=1}^{i} \left(\begin{array}{c} i-1 \\ i-j \end{array} \right) (1-q)^j \, q^{i-j} \int_{t=0}^{E[t]} \frac{1}{E[t]} \, \frac{(\lambda t)^j}{j!} \, e^{-\lambda t} \, dt \\ &= \left[\sum_{j=1}^{i} \left(\begin{array}{c} i-1 \\ i-j \end{array} \right) (1-q)^j \, q^{i-j} \frac{-\lambda^j \, e^{-\lambda t}}{E[t]} \sum_{n=0}^{j} \frac{t^n}{n! \, \lambda^{j-n+1}} \right]_{t=0}^{E[t]} \end{aligned} \tag{9}$$

3.3 Parameter Sensitivity Study

In the following, we will discuss some of the features that can be seen in the system with respect to the sensitivity toward the system parameters. We consider the arrival of packets in batches with a batch arrival rate of $\lambda = 1\,{}^1\!/\mathrm{s}$ and where the batch size is determined by the geometric distribution with parameter q. As described before, we have discrete and deterministic service time for packets with a processing time of $H = 1\,\mathrm{s}$. The activation threshold for service corresponds to the number of downstream neighboring nodes from which the sensor node receives its data and which is expressed by a. We use as thresholds for the service process the lower value of $a = 4$, which activates the service and the upper value of $b = S = 10$, i.e., all entries in the queue are processed.

To evaluate the performance in more complex scenarios and to validate our analysis, we additionally implemented a discrete event simulation of the queueing system. In the experiments given below, we used a very large number of 10^5 batches for obtaining the statistical results from the simulation. Figure 5 shows that the state probabilities obtained from simulation and analysis achieve a good match at the embedding points directly following a departure from the system, as well as at arbitrary time instants.

Figure 6 shows the CDF of the sojourn time of batches for the $M^{GEOM_1}/D^{[a,S]}/1-S$ system. Since we have a discrete service time of at least 1 s, this is also the minimum sojourn time in the system. As the batch size decreases (indicated by growing values of q), the sojourn time of entire batches also increases as batches have to wait longer until the threshold a is reached.

In Fig. 7, the CDF of the sojourn time of packets is shown. It should be noted that these values given here are not normalized, so the CDF may not reach 1, if packets are lost due to blocking. Together with Fig. 8, we are able to see the blocking probability of batches and packets as function of parameter q. For large batch sizes (small q), there is a significantly higher probability of uncompleted units. However, this decreases exponentially with q.

4 Conclusion and Outlook

In this paper we proposed an analytical model for evaluating data aggregation of a wireless sensor node. The model is based on a batch arrival/batch service queuing model and we used the analysis in [10] as the basis, which we extended for our purposes. We derived probability distributions for the number of packets during a service period and studied the influence of the parameters on the system's performance in the case of a discrete time model with geometric packet arrivals. Furthermore, we investigated the sojourn time of batches and packets as well as the percentage of uncompleted batches and packets in dependence of the batch size of packets.

In the future, we would like to extend the study to investigate the data dissemination delay and energy consumption in a multi-hop sensor network, where each node performs data aggregation. This can be combined with using spatial stochastic processes for characterizing the random node layout similarly to [14].

Acknowledgment

The authors would like to thank Yuki Koizumi for his help in preparing this manuscript. Part of this research is supported by the *"Special Coordination Funds for Promoting Science and Technology: Yuragi Project"* of the Ministry of Education, Culture, Sports, Science and Technology in Japan.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] R. Subramanian and F. Fekri, "Sleep scheduling and lifetime maximization in sensor networks: fundamental limits and optimal solutions," in *Proc. of 5th international conference on Information processing in sensor networks (IPSN '06)*, (Nashville, TN), pp. 218–225, 2006.

[3] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," in *Proc. of INFOCOM '02*, (New York, NY), pp. 1567–1576, June 2002.

[4] K. Leibnitz, N. Wakamiya, and M. Murata, "Modeling of IEEE 802.15.4 in a cluster of synchronized sensor nodes," in *Proc. of 19th International Teletraffic Congress (ITC-19)*, (Beijing, China), pp. 1345–1354, August 2005.

[5] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in *Proc. of the Fifth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom '99)*, (Seattle, Washington), pp. 174–185, August 1999.

[6] B. Krishnamachari, D. Estrin, and S. B. Wicker, "The impact of data aggregation in wireless sensor networks," in *22nd International Conference on Distributed Computing Systems (ICDCSW)*, pp. 575–578, 2002.

[7] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: A secure hop-by-hop data aggregation protocol for sensor networks," in *Proc. of MobiHoc '06*, (Florence, Italy), May 2006.

[8] K.-W. Fan, S. Liu, and P. Sinha, "Structure-free data aggregation in sensor networks," *IEEE Transactions on Mobile Computing*, vol. 6, pp. 929–942, Aug. 2007.

[9] L. Galluccio, S. Palazzo, and A. T. Campbell, "Efficient data aggregation in wireless sensor networks: an entropy-driven analysis," in *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, (Cannes, France), September 2008.

[10] H. Gold and P. Tran-Gia, "Performance analysis of a batch service queue arising out of manufacturing and system modelling," *Queueing Systems*, vol. 14, no. 2, pp. 413–426, 1993.

[11] J. C. W. van Ommeren, "Loss probabilities in batch-arrival bulk-service queues with finite capacity," *Communications in statistics. Stochastic models*, vol. 13, no. 2, pp. 371–379, 1997.

[12] U. C. Gupta and P. V. Laxmi, "Analysis of the MAP/G^{a,b}/1/N queue," *Queueing Systems*, vol. 38, no. 2, pp. 109–124, 2001.

[13] A. D. Banik, U. C. Gupta, and M. L. Chaudhry, "Finite-buffer bulk service queue under markovian service process," in *Proc. of Valuetools '07*, (Nantes, France), October 2007.

[14] K. Leibnitz, N. Wakamiya, M. Murata, and M.-A. Remiche, "Analysis of energy consumption for a biological clustering method in sensor networks," in *Proc. of 3rd International Workshop on Measurement, Modeling, and Performance Analysis of Wireless Sensor Networks (SenMetrics '05)*, (San Diego, CA), July 2005.