Proactive Data Dissemination in Wireless Sensor Networks with Uncontrolled Sink Mobility

Massimo Vecchio, Aline Carneiro Viana, Artur Ziviani, Roy Friedman

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Proactive Data Dissemination in Wireless Sensor Networks with Uncontrolled Sink Mobility

Massimo Vecchio — Aline C. Viana — Artur Ziviani — Roy Friedman

N° 6820

January 2009

Thème COM

apport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6820--FR+ENG

Proactive Data Dissemination in Wireless Sensor Networks with Uncontrolled Sink Mobility

Massimo Vecchio* , Aline C. Viana † , Artur Ziviani‡ , Roy Friedman§

Thème COM — Systèmes communicants
Équipes-Projets Asap

Rapport de recherche n° 6820 — January 2009 — 17 pages

Abstract:
This paper investigates Wireless Sensor Networks (WSNs) with mobile sinks that collect data from sensors by following uncontrolled trajectories. In particular, the paper focuses on proactive data dissemination strategies in which the trajectory of the mobile sink is unknown to the sensors. These strategies attempt to obtain a good trade-off between the number of sensors the mobile sinks has to visit in order to collect representative data of all sensors, and the communication effort required by the sensors. All the investigated techniques also avoid the use of multi-hop routing, due to its high cost. Some of them are based on random walks whereas others are based on a combination of probabilistic forwarding and probabilistic storage. An analysis of the various methods in terms of the performance trade-off between the efficiency of the data gathering and communication overhead is also presented.

Key-words: mobile sink, probabilistic data dissemination, random walk-based dissemination, performance analysis

* IMT Institute for Advanced Studies Lucca
† INRIA Saclay - Ile de France sud
‡ National Laboratory for Scientific Computing (LNCC).
§ Technion.

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université
4, rue Jacques Monod, 91893 ORSAY Cedex
Téléphone : +33 1 72 92 59 00

Dissémination proactive de donnés dans les réseaux de capteurs sans fil composés par de puit mobile suivant des trajectoires non définies

Résumé :

Cet article étudie les réseaux de capteurs sans fil (WSNs) avec des puits mobiles qui collectent des données provenant de capteurs, en suivant des trajectoires non définies. En particulier, le document se concentre sur des stratégies de dissémination de données proactives dans lequel la trajectoire du puit mobile est inconnue par les capteurs. Ces stratégies ont comme objective d'obtenir un bon compromis entre le nombre de capteurs le puit mobile doit visiter afin de recueillir une quantité représentative de donnés de l'ensemble des capteurs, et l'effort de communication requis par les capteurs. En raison du coût élevé, toutes les techniques étudiées évitent l'utilisation de routage multi saut. Certains d'entre eux sont basés sur des marches aléatoires, alors que d'autres sont basés sur une combinaison de probabilité d'acheminement et de stockage. Une analyse des différentes méthodes en terme du compromis entre l'efficacité de la collecte de données et la communication est également présentés.

Mots-clés : puit mobile, Dissémination proactive de donnés, dissémination basée sur des marches aléatoires, analyse de performance

1 Introduction

We consider scenarios of Wireless Sensor Networks (WSNs) in which the sink node is mobile [12,5]. Data management in mobile sink wireless sensor networks (MSWSNs) include data dissemination and data collection [3]. Data dissemination is related to how data is distributed in the network either to proactively store the monitored data through the network for it to be retrieved later by the mobile sink or to reactively send the data towards the mobile sink. Data collection is related to how the mobile sink gathers the monitored data made available by the sensors in the network. In other words, depending on how data is managed, the sink may either follow a predetermined trajectory with controlled mobility (i.e. the sink must visits some predefined nodes to retrieve a representative view of the monitored data) or move freely in the network following an uncontrolled mobility pattern and still harvest a representative view of the monitored data as long as it visits a minimum number of nodes, no matter which one. In short, data management proposals in MSWSN may be characterized by their data dissemination strategy (proactive or reactive) combined with the kind of trajectory (controlled or uncontrolled) the mobile sink takes through the network to collect the data, as we describe in further details in Section 2.

In this paper, we focus on proactive data dissemination strategies that allow a mobile sink to effectively gather a representative view of the monitored region covered by n sensor nodes by visiting, in an uncontrolled trajectory, *any* m nodes, where $m << n$. This approach does not require a priori knowledge of all network nodes, does not use multi-hop routing, or any mechanism to track the sink, while it allows the mobile sink to follow any trajectory while collecting the data. In a previous work [14], we propose a proactive data dissemination strategy based on random walks (RWs) to achieve this decoupling of the data dissemination management from the mobile sink's trajectory. Using this previous approach, a mobile sink can typically gather information from approximately 90% of the nodes by visiting about a few number of any nodes (about 7% of them) in the WSN. Although these results are quite encouraging in terms of efficiency in data collection, the use of RWs to uniformly distribute the monitored data through the network incur in a high communication overhead. An open issue is to achieve an alternative solution to proactively disseminate the data in an efficient way (i.e. with a reduced communication overhead) so that a mobile sink following an uncontrolled trajectory is able to effectively retrieve a representative view of the network. Here, we make a first step towards this solution.

Our goal in this paper is thus to investigate and analyze the trade-off between the performance in data gathering and the performance in communication overhead of different strategies of proactive data dissemination that allow the mobile sink to move in an uncontrolled way to harvest the monitored data. We consider different proactive data dissemination strategies based on both random walks and probabilistic forwarding—including flooding—with a storing probability at each traversed node. To the best of our knowledge, this is the first paper to fully analyze this trade-off in the context of WSNs with mobile sinks. Analyzing and understanding this trade-off is crucial to create a basis to build upon in order to conceive and develop new approaches for proactive data dissemination that allow uncontrolled mobility of the sink, thereby leading to efficient data gathering with a reasonable communication overhead—i.e. levering off the involved mutual effects of the analyzed trade-off.

The remainder of this paper is organized as follows. In Section 2, we briefly present related work in the context of WSNs with mobile sinks. Section 3 introduces the proactive data dissemination strategies we consider in our performance analysis. Our performance analysis are presented in Section 4. Finally, Section 5 summarizes the outcomings of our analysis and discusses possible future work.

2 Wireless Sensor Network with Mobile Sinks

Data management proposals in wireless sensor networks with mobile sinks may be characterized by their data dissemination strategy (proactive or reactive) combined with the data collection strategy adopted by the mobile sink—i.e. the kind of trajectory (controlled or uncontrolled) the mobile sink takes through the network to collect the data. In this section, we briefly overview the related work in this context considering the combinations of different data dissemination and data collection strategies:

- Reactive data dissemination with controlled sink mobility – in this approach the mobile sink follows a controlled, and thus predictable, trajectory as it traverses the monitored field. Sensors react to the presence of the mobile sink by disseminating (actually routing) data towards the current position of the sink or to other nodes that are located close to the predicted sink trajectory. Examples of such an approach are the proposals by Jea et al. [10], Luo et al. [13], and Basagni et al. [2].

- Reactive data dissemination with uncontrolled sink mobility – in this approach the mobile sink follows a free, i.e., uncontrolled, trajectory as it traverses the sensor field to collect data. As the sink mobility is uncontrolled, the sink only collects data from sensors along its trajectory or from sensors that route their data towards the sink. Sensors thus establish some scheme to track the sink forwarding the data towards it as the sink moves throughout the monitored field. Hwang et al. [9] build a dynamic sink oriented tree and Yang et al. [15] adopt a swarm intelligence approach for reactively routing the data towards one or multiple mobile sinks that follow uncontrolled trajectories.

- Proactive data dissemination with controlled sink mobility – in this approach sensors disseminate their monitored data towards a subset of sensors that play the role of storage nodes. This subset of nodes typically form a virtual structure within the WSN that make the data available to be retrieved later by the mobile sink. The mobile sink should then visit some nodes belonging to the formed virtual structure to collect a representative view of the monitored data, i.e. the mobile sink has its trajectory determined (controlled) by the shape of the virtual structure. Such proactive data dissemination solutions with a rendez-vous virtual structure basically differ on how the virtual structure is formed within the WSN. Some examples of virtual structures for this purpose are a grid structure such as TTDD [11] and a line-based one [7]. Ben Hamida and Chelius [8] provide a recent survey on proactive data dissemination strategies that use virtual structures in MSWSN.

- Proactive data dissemination with uncontrolled sink mobility – in this approach sensors proactively distribute their monitored data to the other sensors in the WSN. Ideally, the data dissemination yields a uniform data distribution throughout the WSN in a way that allows a sink following an uncontrolled trajectory to retrieve a representative view of the monitored field by visiting a relatively small number of any nodes in the network. An example of such an approach is a previous work of ours [14] in which the data dissemination is performed based on random walks (RWs) in the WSN.

Reactive proposals dealing with sink mobility, both controlled and uncontrolled, must dedicate a significant amount of resources to track the sink and to forward on-the-fly the data to be collected towards the mobile sink. Proactive proposals based on rendez-vous virtual structures must keep knowledge about the nodes relative position. This is needed to forward data towards the nodes that take part of the virtual structure. Furthermore, nodes in the virtual structure and nodes close to them are subject to a higher demand in terms of storage and transmission resources, leading to an imbalance in energy consumption among the sensor nodes within the WSN.

In contrast with most previous work in MSWSN, we focus in solutions based on a proactive data dissemination approach to deal with uncontrolled sink mobility. This approach does not require a priori knowledge of all network nodes, does not use multi-hop routing, or any mechanism to track the sink, while it allows the mobile sink to follow any trajectory while collecting the data, leading to a full decoupling of the data dissemination from the sink trajectory management. This paper contributes with a full analysis of the trade-offs involved in adopting different proactive data dissemination strategies to deal with uncontrolled sink mobility.

3 Proactive Data Dissemination in MSWSN

Following the proactive data dissemination's features, each node in the network is provided with a *partial view* regarding some other nodes (including for itself). In this way, each node acts as a storage node for some other nodes in the WSN, but not for all of them. The goal is thus to limit load in each individual storage node and to enable each node to determine its own group of storage nodes independently of other nodes, without any implication on the randomness of each group. By slight abuse of terminology, we interchangeably use the term partial view both for the actual information stored at a given node p and for the IDs of the nodes whose information is stored at p.

Some common properties of all the studied proactive data dissemination approaches are: (i) data dissemination management is decoupled from the mobile sink's trajectory; (ii) data availability is improved by replicating nodes sensed data in selected storage nodes in the network; (iii) no reply is returned to the node that initiates the dissemination; (iv) nodes do not have to be aware of the size of the partial views; (v) no multiple-hop routing is required, and (vi) no a priori knowledge of all network nodes is required. Although sharing these common properties, each data dissemination approach has particular mechanisms for *forwarding* and *storing* information—the two main dissemination components. As discussed later, these two components dictate the trade-off between data collection efficiency and communication overhead.

3.1 Random walks based data dissemination

The goal of this approach is to ensure that the partial view of each node will correspond to a uniform sample of the entire information (or entire set of nodes) in the system. To construct these partial views, a *forwarding strategy* is used where an appropriate number of maximum degree Random Walks (MD RW) [1] is started, where each RW carries the information and ID of the node that started it. The *storing strategy* establishes then that the data carried by the RW is stored at the node in which the RW stops, which is known as *reverse sampling* as defined in the RaWMS proposition in [1]. As proved in [1], this ensures that each node in the system ends up serving as the *storage node* for a uniform sample of the other nodes. This data dissemination strategy will be referred hereafter in this paper as *RaWMS-based*.

A RW is started every δ time units at the source node by randomly selecting the next neighbor to send the message; the same process is repeated at each node along the RW until the distance $d = \frac{n}{2}$ is reached. Each storage node is thus set with an average *view* size of $k = \sqrt{n}$ (only used for the performance estimation, since nodes are not aware of the k value). Since there is no way to a priori know where a RW will end, two or more RWs may end in the same node. Thus, to guarantee a target partial view size k within distinct nodes, each source node starts $n \ln(\frac{n}{n-k})$ RWs, for n nodes in the network (refer to [14] for more details).

Following the complexity analysis performed in [1], the communication complexity of this approach is determined by the number of RWs each node should start, i.e., $nb\_RW = n \ln(\frac{n}{n-k})$, and the length of each RW, i.e. d. This results in a total communication complexity of $n \times nb\_RW \times d = \theta(n \times n \ln(\frac{n}{n-k}) \times d)$. For the special case of $k = \theta(\sqrt{n})$, we get $nb\_RW \approx \frac{nk}{n-k} \approx \sqrt{n}$. Hence, the communication complexity equals $\theta(\frac{n^2}{2}\sqrt{n})$, for $d = \frac{n}{2}$, or still $O(n^2\sqrt{n})$.

3.2 Probabilistic data dissemination

In order to investigate different performance and overhead related metrics, this section describes different probabilistic forwarding and storing strategies. In particular, a good data collection by a mobile sink with uncontrolled mobility depends on an efficient distribution of data over nodes in the network. Efficiency in data distribution needs, however, to be combined with low communication overhead. Thus, contrarily to the efficient uniform distribution of data assured by the described deterministic approach, here we focus on a less efficient but also less costly probabilistic data dissemination. We will see that, with a lower communication overhead, they allow the data to be gathered with a slightly decrease in representativeness when compared with the deterministic approach.

Fixed forwarding and storing probabilities: This probabilistic data dissemination is composed of a *forwarding probability* $f = 1$ (i.e., flooding) and of a *storing probability* $s = \frac{\sqrt{n}}{n}$. Thus, each node sends its own information and ID to all neighbors. Neighbors then simply forward each received message if not already sent; at the same time each node stores the forwarded information with a fixed a priori probability s. Considering nodes with limited buffer and similarly to the RaWMS-based approach, the storing probability is set to allow each node to remember at most $\sqrt{n}$ information. In this case, considering an ideal flooding where all nodes receives all the disseminated information, each node remembers the received information with probability $\frac{\sqrt{n}}{n}$. Note that the communication complexity here equals $O(n^2)$.

Variable forwarding and fixed storing probabilities: This data dissemination approach employs a combination of probabilistic forwarding with deterministic corrective measures, as described in [4]. The deterministic corrective measures include deterministic gossiping as well as timer based corrections of the probabilistic process. The forwarding probability is variable and adaptively set based on the observed number of nodes in each one-hop neighborhood. Thus, each node p sends its own information and

ID to the set of direct neighbors $N(p)$, but forwards the received messages with probability $f = \frac{\beta}{|N(p)|}$. β is called the reliability factor and represents the average number of nodes in each one-hop neighborhood that retransmit each received message m. As established by Drabkin et al. [4], a $\beta = 2.5$ guarantees that approximately 90% of all nodes will receive m. On the other hand, the storing probability is fixed and equals $s = \frac{\sqrt{n}}{n}$. As stated before, this probability is set to allow each node remembering at most $\sqrt{n}$ information. In this case, using $f = \frac{\beta}{|N(p)|}$ to disseminate the information to almost all nodes, each node that receives this information remembers it with probability $\frac{\sqrt{n}}{n}$. The communication complexity of this approach depends on the network density and is determined by the number of forwarding each node performs. This results in a total communication complexity of $(f \times n) \times n = O(\frac{\beta}{|N(p)|} \times n^2)$

3.3 Strategies advantages and drawbacks

The RaWMS-based dissemination approach guarantees the replication of nodes information in well selected storage nodes in a well balanced manner. This is achieved by the uniformly chosen storage nodes and by the uncorrelation between the views of neighboring nodes. This way, whenever the mobile sink visits a given node, it can collect a uniform sample of the information of all nodes in the system. Hence, when compared to the other approaches, the RaWMS-based approach gets a high performance in terms of representativeness: it only requires the visit of *any* $2.3\sqrt{n}$ nodes to collect information from about 90% of the nodes (as implied by the analysis of [6] and results from [14]). Nevertheless, this is achieved at the expense of generating a high communication overhead, i.e. $O(n^2\sqrt{n})$ point-to-point messages.

A lower communication overhead can be obtained with the approach using $f = 1$ and $s = \frac{\sqrt{n}}{n}$, i.e. $O(n^2)$ broadcast messages[1]. Although generating good data distribution that gets closer to RaWMS-based approach (as shown in Section 4), uniform distribution is not the focus of this approach. In this way, the mobile sink has to visit randomly selected sensor nodes. Moreover, as shown in Section 4, the number of visited nodes required to get a significant representative view of the network is similar to RaWMS-based, i.e. $2.3\sqrt{n}$.

Finally, the last discussed dissemination approach, i.e. $f = \frac{\beta}{|N(p)|}$ and $s = \frac{\sqrt{n}}{n}$, allows the lowest communication overhead among the investigated approaches, i.e. $O(\frac{\beta}{|N(p)|} \times n^2)$ if no deterministic corrective measures are used. Nevertheless, since uniform distribution is not assured and not all nodes receive the forwarded information, randomly selected sensor visits are also required. More specifically, supposing the delivery probability is 0.90 (the case for $\beta = 2.5$), by visiting x nodes, the probability of missing information is $(1 - 0.90)^x$, assuming that the failure to receive a message is truly independent. In other words, by visiting x nodes, the probability of missing some amount of information is given by $1 - (1 - 0.90)^x$. Although seeming to be irrelevant, we will see in Section 4 that this missing probability has a huge impact in how information is distributed in the network, and consequently, in the efficiency of data gathering by the mobile sink. In this approach, it is important to emphasize the need for random sink visits, since when any probabilistic dissemination protocols are invoked, it is more likely that nodes close to the dissemination source will have the message stored than remote nodes. Hence, it is important that the nodes visited by the

[1] Notice that in some wireless MAC protocols, a broadcast can be more expensive than a point-to-point message. For example, WiFi NICs tend to send broadcast messages at 1-2 Mbps, whereas point-to-point messages can enjoy the full bandwidth of the channel, up to 54Mbps in 802.11g.

sink will be chosen at random. On the other hand, if the $f = \frac{\beta}{|N(p)|}$ probability is set to deliver 99% of the message, then the nodes that miss the message are likely to be uniformly randomly distributed in the network, in which case it is less important for the mobile sink to visit random nodes.

4 Performance Analysis

This section describes the experiments we have conducted to assess both the performance and the accuracy of the investigated approaches. The results presented here helped us to take confidence with the studied dissemination strategies, their parameters and cost, and how they are influenced by different network densities. As we will show, the $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$ approach presents the best results in terms of communication overhead, while the RaWMS and $f = 1$, $s = \frac{\sqrt{n}}{n}$ ones have the best efficiency in data gathering by a sink with uncontrolled mobility.

4.1 Evaluation methodology

The experiments have been done using a discrete event simulator implemented using Matlab. Note that as we are mostly interested in understanding the impact of the dissemination behavior of the different data dissemination approaches, our simulator deliberately does not model all the details of a realistic MAC protocol. Instead, we considered a simplified MAC layer where neither messages losses nor collisions happen. However, the techniques we investigate can accommodate messages loss through standard salvation techniques, and here we would like to focus on their ability to disseminate data uniformly and efficiently. This is mainly due to the fact that the presented dissemination approaches either are independent of the order of message arrival or implement corrective measures for the probabilistic process. A realistic MAC protocol would have the major effect of introducing arbitrary delays on the dissemination process due to messages re-transmission and timeout. A detailed study of its impact on our investigated approaches constitutes a future work.

Experimental setup: Our simulations involve scenarios with 200 nodes placed at uniformly random locations in a 2-dimensional grid. The average number of nodes in the communication range of any node was set to a target average density d_{avg} (i.e. average number of neighbors per node). By varying d_{avg}, we built two classes of *connected* density-based topologies, refereed here as: *Sparse* ($d_{avg} = 6$) and *Dense* ($d_{avg} = 24$). In each class and when it is not specified the contrary, results correspond to an average of 30 different grid networks topologies composed of 200 nodes. We performed all the experiments per class, averaging the results so as to have a global view of the parameters to be tuned, with respect to the different d_{avg}.

Simulation composition: Each simulation comprises two tasks: the *data dissemination* and *the data collection* tasks. In the former, the three discussed dissemination approaches in Section 3 were implemented. They are referred in the graphs as: RaWMS; $f = 1$, $s = \frac{\sqrt{n}}{n}$; and $f = \frac{\beta}{|N(p)|}$, $s = \frac{\sqrt{n}}{n}$. The data collection task is then used to investigate the efficiency of data gathering by the mobile sink (see Section 4.3 for more details). For this, the sink performs x visits and the amount of data collected per visit is recorded.

Metrics: To evaluate efficiency and cost, we use three metrics: (i) the efficiency in data gathering, which is the accumulated amount of collected information after a node is visited by the sink; (ii) the message overhead, which is the forwarding count of messages in the network per node; (iii) the quality of data distribution, which is the number of nodes that store information about each node p (replication level), and the uniformness of the distribution of knowledge about nodes in the network.

4.2 Understanding the impact of β in the view size of nodes

As stated before, in order to fairly compare the probabilistic dissemination approaches with the RaWMS-based approach, we limit the view size of nodes to around $\sqrt{n}$ entries and set the storing probability to $s = \frac{\sqrt{n}}{n}$. Nevertheless, this probability is strongly dependent on how information is propagated in the network, and consequently, on how well the forwarding probability performs. Contrarily to the approach with fixed forwarding probability (i.e. $f = 1$, where all nodes will receive the message with a high probability), the dissemination efficiency gotten with the variable forwarding probability, $f = \frac{\beta}{|N(p)|}$, depends on the appropriate set of the reliability factor β. Therefore, for each class of density-based topologies, we performed different experiments in order to determine the β value inducing an average nodes' view size around $\sqrt{n}$.

Figure 1 depicts the average number of nodes in the network for different view sizes and for *Dense* network topologies. In particular, for different values of β, we left the nodes' view size parameter unbounded, performed an one-shot dissemination, and analyzed the final nodes' view size distributions. It can be observed that the average nodes' view size can be easily controlled by just tuning β. We can conclude that by setting β equals to 1, we can tune the average view size to a value of around $\sqrt{n}$ in dense topologies.

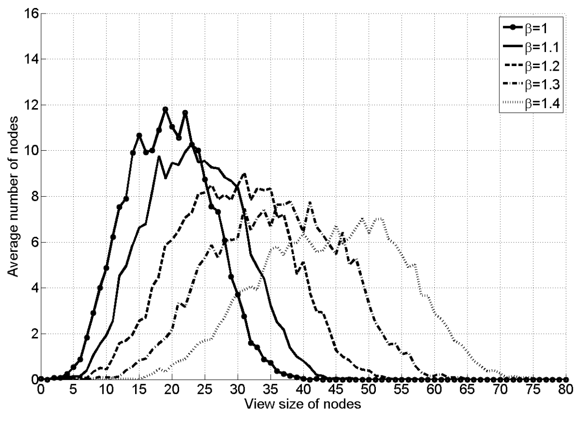
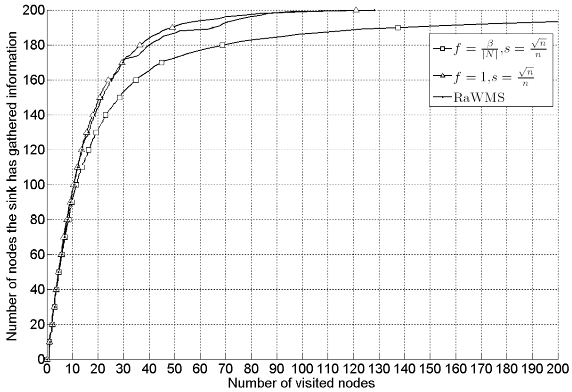
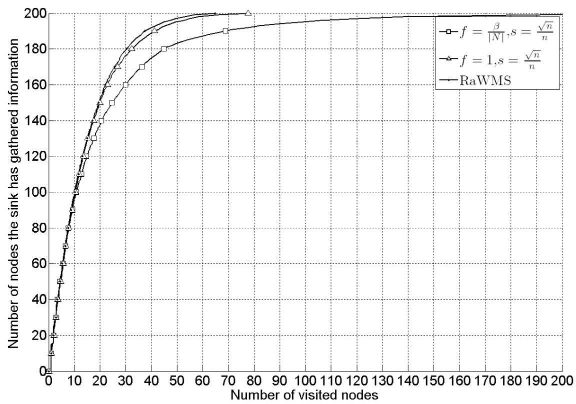


Figure 1: $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$, *Dense* topologies ($d_{avg} = 24$), and 200-node network.

We also calculate the proper β value to be used in experiments with *Sparse* topologies in a similar way. Likewise, a view size's control was observed by just tuning β. The results show that a view size of around $\sqrt{n}$ entries can be achieved in sparse networks by setting β equals to 2.3, which matches expectations following the analysis in [6].



(a) Sparse topology.



(b) Dense topology.

Figure 2: Amount of collected information per visited node in 200-node (a) *Sparse* topology ($d_{avg} = 6$) and (b) *Dense* topology ($d_{avg} = 24$). Comparison between RaWMS; $f = 1$, $s = \frac{\sqrt{n}}{n}$; and $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$.

Once the particular values of β have been determined for dense and sparse topologies, they will be used in the next sections for the evaluation of the previously described metrics related to the dissemination strategies.

4.3 Efficiency in data gathering

The experiments in this section were run with the dissemination and the gathering tasks. During the dissemination task and according to the dissemination strategy being evaluated, each node simply disseminate and store the information they generate/receive. Then, during the gathering task a mobile sink is firstly placed in a random position in the network. After visiting the nodes in this position, the sink then chooses the next node to visit, trying to avoid to visit again an already visited node (for this reason we refer to some particular visiting strategies introduced in [6]). When the sink visits a node, it gathers this node's data and all the information in its view. This procedure is repeated until the sink has collected all the network information.

Figure 2(a) and 2(b) show the amount of accumulated collected information the mobile sink gathers per visited node, in 200-node networks with $d_{avg} = 6$ and $d_{avg} = 24$, respectively. The results indicate that RaWMS and $f = 1$, $s = \frac{\sqrt{n}}{n}$ have similar performance and are slightly affected by the network density. In particular, after visiting *any* $2.3\sqrt{n} \approx 32$ nodes, the sink is able to collect information from about 90% of the nodes (87% in sparse topologies), as implied by the RaWMS analysis in [6]. These results confirm that, as RaWMS, the $f = 1$, $s = \frac{\sqrt{n}}{n}$ allow the mobile sink to achieve a high representative view of the information concerning the deployment region by only visiting a relative small number of nodes in a sparse or dense network.

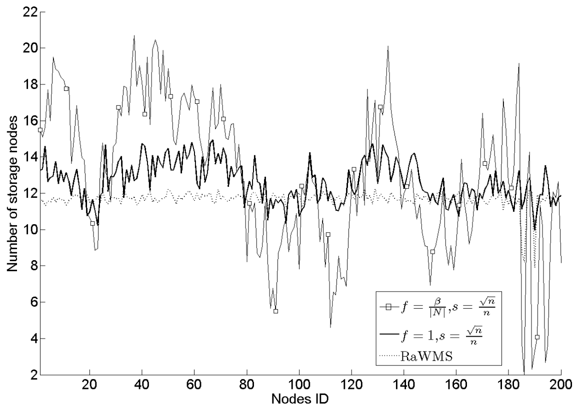
Even if a decrease of about 10% in the data collection efficiency is observed with the $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$ strategy, the results still remain quite satisfactory: 82% of nodes' information ($78,5\%$ for sparse networks) is achieved after visiting any $2.3\sqrt{n} \approx 32$ nodes. The decrease in performance is mainly due to the way information is distributed in the network in this dissemination strategy. Inline with this trend, Section 4.4 discusses in details the quality of information distribution resulted from each dissemination strategy.

4.4 Quality of data distribution

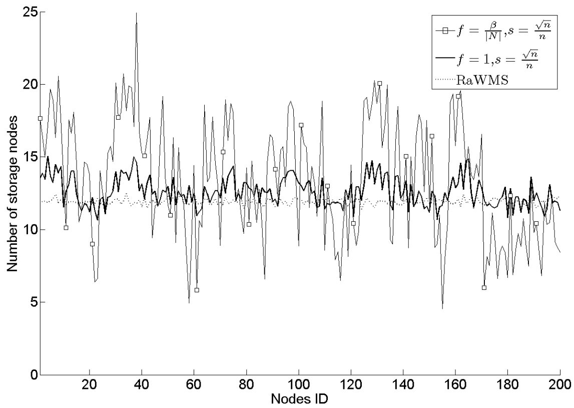
Using two set of experiments, we investigate here the quality of the data dissemination yield by each approach over the network. For this, the experiments were performed selecting at random one network topology from each set of dense and sparse topologies. On the selected topologies, each data dissemination approach was performed 30 times (simulation runs) and the results averaged for being later plotted in the graph. The same was performed for different topologies of each class and similar results were observed so the results shown here are representative for the whole relative class they belong.

Representing the first set of experiments, Figure 3 shows the average amount of distinct storage nodes associated to each node ID for each dissemination approach, allowing for the level of replication of information in the network. Compared to the total number of nodes in the network, this average amount is kept low for all the studied dissemination strategies: a maximum amount of 12% of storage nodes per node is used. As expected, the RaWMS-based approach presents a quite stable behavior, which proves the regular and uniform information distribution performed by each node. In addition, the network densities barely affects this stable behavior. Although not being as stable as RaWMS-based approach, the $f = 1$, $s = \frac{\sqrt{n}}{n}$ approach still guarantees an almost regular amount of storage nodes to each node ID, which also shows the good distribution resulting from this approach. This similar behavior in data distribution explains the good data gathering performance of the $f = 1$, $s = \frac{\sqrt{n}}{n}$ approach.

On the other hand, the $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$ approach assigns each node with a different amount of storage nodes in the network, besides being more affected by the network density. In particular, when the density is low (i.e. $d_{avg} = 6$), some nodes store their information in very few storage nodes in the network (from 8 to 2 storage nodes). This reflects the effect of a double-based probabilistic data dissemination approaches: even if a node, after playing the probability $f = \frac{\beta}{|N|}$, decides to forward a received information, it may decide not to store it, after playing the $s = \frac{\sqrt{n}}{n}$ storing probability. All these factors, justify the decrease in data gathering efficiency of this approach, when compared to RaWMS-based and $f = 1$, $s = \frac{\sqrt{n}}{n}$ approaches, as discussed in Section 4.4.



(a) Sparse topology.

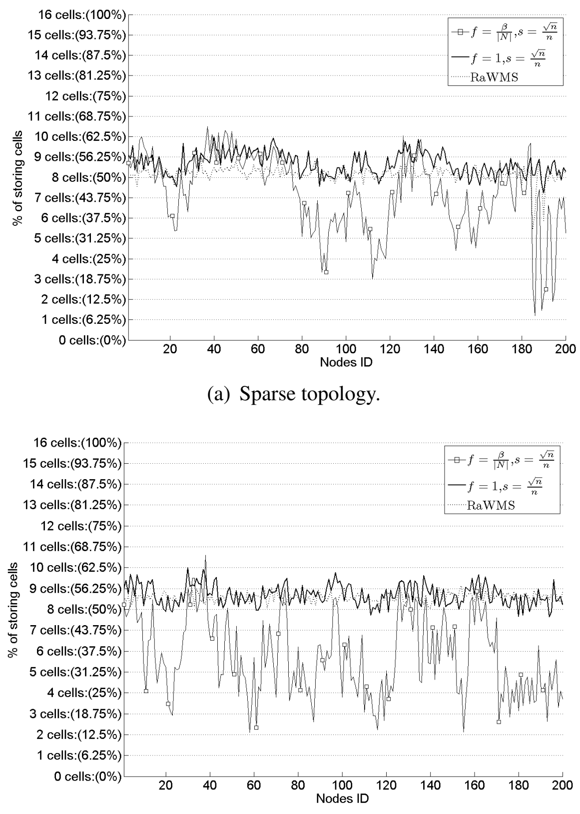


(b) Dense topology.

Figure 3: Number of storage nodes per nodes ID in a 200-node (a) *Sparse* topology ($d_{avg} = 6$) and (b) *Dense* topology ($d_{avg} = 24$). Comparison between RaWMS; $f = 1$, $s = \frac{\sqrt{n}}{n}$; and $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$.

Finally, the second set of experiment, shown by Figure 4, evaluate in how many different locations in the network, a node's information is contained, allowing for the level of data distribution in the network. We believe that this is an important experiment since it shows how well the dissemination approaches could be able to distribute the nodes' information over the network. For this, the network area of size 17×17 is divided into 16 cells (6 cells of 4×5, 1 of 5×5, and 9 cells of 4×4). Note that, since nodes are positioned in a grid-based network, the size of the cells also gives the maximum number of nodes that may be located in each of them. For each node in the network, we then evaluate in how many different cells its information is contained. The obtained values are averaged over the 30 simulation runs for each selected topology and for each dissemination approach.

We can observe that RaWMS has a very homogeneous information distribution and presents the same behavior for dense and sparse topologies: each node has its information distributed in approximately 50% of distinct locations in the network. As implied

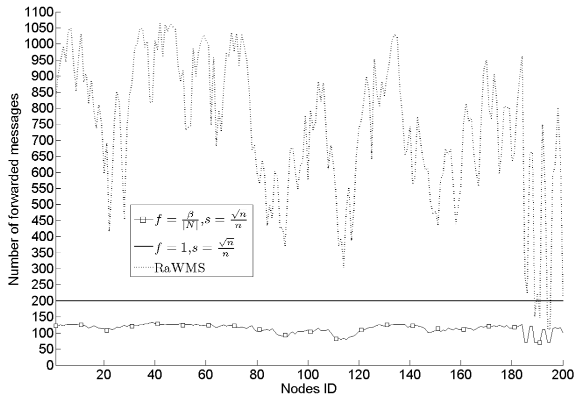


(a) Sparse topology.

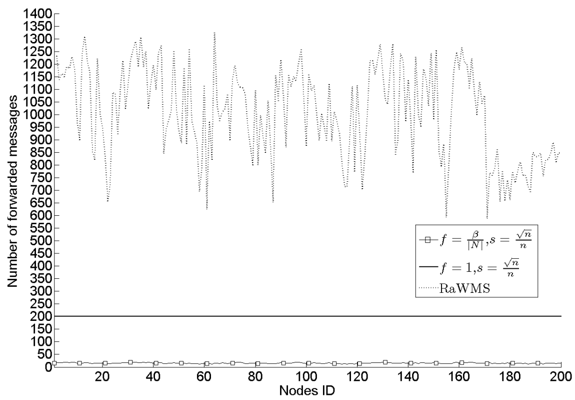
(b) Dense topology.

Figure 4: Relative location information distribution per nodes ID in a 200-node (a) *Sparse* topology ($d_{avg} = 6$) and (b) *Dense* topology ($d_{avg} = 24$). Comparison between RaWMS; $f = 1$, $s = \frac{\sqrt{n}}{n}$; and $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$.

by the analysis and results of [1], the RaWMS-based dissemination approach provides a uniform distribution of data over the network. On the other hand, the $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$ approach does not have the same trend. Nodes' information is distributed to fewer distinct locations in the network. These results confirm the discussion provided in Section 4.4. Messages in probabilistic dissemination protocols are more likely to achieve nodes close to the dissemination source than by remote nodes, leading to a less uniform data distribution in the network. Nevertheless, it is interesting to note the capability of adapting to density of this approach. More specifically, in sparse networks, nodes have less neighbors and consequently, nodes' information has a higher probability to reach farther locations than in dense networks. This is reflected on the higher average number of distinct cells nodes store their information, see Figure 4(a). Finally, the less efficient information distribution of this approach also justifies its decrease in data gathering performance discussed in Section 4.3.



(a) Sparse topology.



(b) Dense topology.

Figure 5: Communication overhead per node ID in a 200-node (a) *Sparse* topology ($d_{avg} = 6$) and (b) *Dense* topology ($d_{avg} = 24$). Comparison between RaWMS; $f = 1$, $s = \frac{\sqrt{n}}{n}$; and $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$.

4.5 Communication overhead

This section investigates the communication overhead incurred by each node during the data dissemination. Figure 5 depicts the number of forwarded messages per node in sparse and dense topologies. Similar to Section 4.4, the experiments were performed selecting at random one network topology from each set of sparse and dense topologies, with results reflecting the average of 30 simulation runs on the selected topologies. In both densities, results show the lower communication overhead generated by the adaptive probabilistic dissemination approach, i.e. $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$. It also serves as proof of how adaptive to neighborhood density this approach is, since in sparse networks more messages are forwarded in order to guarantee the delivery reliability imposed by β (as defined in Section 4.2, $\beta = 1$ in dense topologies and $\beta = 2.3$ in sparse topologies).

The results in Figure 5 also present a clear view of the ordering in terms of communication overhead between the data dissemination approaches: RaWMS has a much higher communication overhead than $f = 1$, $s = \frac{\sqrt{n}}{n}$, which in turn presents higher overhead than $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$, as discussed in Section 3. As implied by the analysis in [6] and confirmed by Figure 5(b), the worst case in RaWMS occurs in dense networks, where each node may forward at most $n \times \frac{n}{2}$ point-to-point messages. Concerning communication overhead, these results prove that the RaWMS-based approach is more viable in sparse topologies. The $f = 1$, $s = \frac{\sqrt{n}}{n}$ approach has the overhead of flooding, i.e, n broadcast messages generated per node, as shown in Figures 5(b) and 5(a). Finally, contrarily to RaWMS and in terms of communication overhead, the $f = \frac{\beta}{|N|}$, $s = \frac{\sqrt{n}}{n}$ is more adaptable to dense topologies. Because of its capability of being density adaptable, less number of messages are generated and forwarded by each node in dense networks. According to discussions in Section 3 and considering only the probabilistic forwarding is used, the worst case in this approach occurs in sparse topologies and equals to $\frac{\beta}{|N|} \times n$ messages at most being sent by each node. Since our variable probabilistic approach also implement some deterministic corrective measures, this explains the enhanced number of messages.

5 Summary and Outlook

We have investigated in this paper both the performance and the accuracy of three different proactive data dissemination approaches. The approaches use random walks or probabilistic mechanisms to disseminate around $\sqrt{n}$ data at each node through the WSN, with a controlled overhead. Consequently, a mobile sink can follow any trajectory through the deployment region and still get a highly representative view of the monitored region, while visiting only a small number of nodes. This decouples the data dissemination management from the mobile sink's trajectory. The obtained results allowed us to analyze the performance trade-off between the efficiency in data gathering and communication overhead that emerge from each dissemination approach. According to our results, the variable probabilistic approach leads to a satisfactory efficient data gathering with a reasonable communication overhead. We believe, however, that an improvement in data gathering can be achieved if more adaptable approaches could be invoked. For instance, by knowing that messages in probabilistic dissemination protocols are more likely to be gotten by nodes close to the dissemination originator than by remote nodes, we could think about metrics that are more adaptable to this issue. Some interesting metrics we are currently investigating are: the distance in hops to the information originator and the number of nodes that have already stored the disseminated information in the adaptation process. This constitutes the direction of our future works.

References

[1] Ziv Bar-Yossef, Roy Friedman, and Gabriel Kliot. RaWMS - Random Walk based lightweight Membership Service for wireless ad hoc networks. *ACM Trans. Comput. Syst.*, 26(2):1–66, June 2008.

[2] Stefano Basagni, Alessio Carosi, Emanuel Melachrinoudis, Chiara Petrioli, and Z. Maria Wang. Controlled sink mobility for prolonging wireless sensor networks lifetime. *Wireless Networks Journal (WINET)*, 14(6):831–858, December 2008.

[3] I. Chatzigiannakis, A. Kinalis, and S. Nikoletseas. Efficient data propagation strategies in wireless sensor networks using a single mobile sink. *Computer Communications*, 31(5):896–914, 2008.

[4] Vadim Drabkin, Roy Friedman, Gabriel Kliot, and Marc Segal. RAPID: Reliable Probabilistic Dissemination in wireless ad-hoc networks. In *Proceedings of the 26th IEEE International Symposium on Reliable Distributed Systems (SRDS)*, October 2007.

[5] Eylem Ekici, Yaoyao Gu, and Doruk Bozdag. Mobility-based communication in wireless sensor networks. *IEEE Communications Magazine*, 44(6), July 2006.

[6] R. Friedman, G. Kliot, and C. Avin. Probabilistic quorum systems in wireless ad hoc networks. In *Proceedings of the 9th IEEE Conference on Dependable Systems and Networks (DSN)*, June 2008.

[7] Elyes Ben Hamida and Guillaume Chelius. A line-based data dissemination protocol for wireless sensor networks with mobile sink. In *Proceedings of IEEE International Conference on Communications (ICC)*, May 2008.

[8] Elyes Ben Hamida and Guillaume Chelius. Strategies for data dissemination to mobile sinks in wireless sensor networks. *IEEE Wireless Communications Magazine*, 15(6), December 2008.

[9] Kwang-il Hwang, Jeongsik In, Yeo-hong Yun, and Doo-seop Eom. Dynamic sink oriented tree algorithm for efficient target tracking of multiple mobile sink users in wide sensor field. In *Proceedings of the 6oth IEEE Vehicular Technology Conference - VTC2004-Fall*, pages 4607–4610, 2004.

[10] David Jea, Arun Somasundara, and Mani Srivastava. Multiple controlled mobile elements (data mules) for data collection in sensor networks. In *Proceedings of the 1st IEEE/ACM International Conference on Distributed Computing in Sensor Systems (DCOSS05)*, pages 244–257, Marina del Rey, CA, USA, June 2005.

[11] H. Luo, F. Ye, J. Cheng, S. Lu, and L. Zhang. TTDD: Two-tier data dissemination in large-scale wireless sensor networks. *Wireless Networks Journal (WINET)*, 11(1):161–175, January 2005.

[12] J. Luo and J. P. Hubaux. Joint mobility and routing for lifetime elongation in wireless sensor networks. In *Proceedings of the IEEE INFOCOM 2005*, Miami, FL, USA, March 2005.

[13] J. Luo, J. Panchard, M. Piorkowski, M. Grossglauser, and J. P. Hubaux. Mobiroute: Routing towards a mobile sink for improving lifetime in sensor networks. In *Proceedings of the 2nd IEEE/ACM International Conference on Distributed Computing in Sensor Systems (DCOSS06)*, pages 480–497, San Francisco, CA, USA, June 2006.

[14] Aline Carneiro Viana, Artur Ziviani, and Roy Friedman. Decoupling data dissemination from mobile sink's trajectory in wireless sensor networks. *IEEE Communications Letters*, 2009. To be published.

[15] H. Yang, F. Ye, and B. Sikdar. SIMPLE: Using swarm intelligence methodology to design data acquisition protocol in sensor networks with mobile sinks. In *Proceedings of the IEEE INFOCOM 2006*, April 2006.

Contents

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 Orsay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399