Biased Random Walk Model to Estimate Routing Performance in Sensor Networks

Issam Mabrouki, Gwillerm Froc, Xavier Lagrange

Biased Random Walk Model to Estimate Routing Performance in Sensor Networks

Issam Mabrouki, Gwillerm Froc and Xavier Lagrange

Mitsubishi Electric TCL - e-mails: {mabrouki, froc}@tcl.ite.mee.com
RSM, ENST Bretagne - email: xavier.lagrange@enst-bretagne.fr

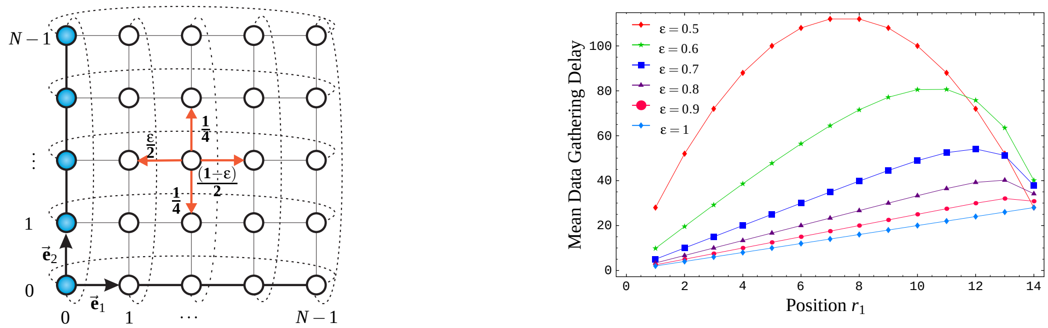
Les réseaux de capteurs sans fils sont constitués d'un grand nombre de nœuds assujettis à de sévères contraintes en terme d'énergie, de capacité de traitement et de communication. Dans ce contexte, afin de réduire la complexité, un des défis majeurs rencontrés dans ce type de réseau est le calcul des routes et la mise en œuvre de schémas de routage efficaces tout en minimisant la quantité d'information utilisée sur l'état du système. De nombreux travaux ont étudié ce compromis de façon qualitative ou grâce à des simulations. Nous proposons un modèle basé sur la théorie de la marche aléatoire pour estimer analytiquement ce compromis en considérant plus particulièrement l'influence du degré de connaissance de l'état du système que possède un noeud sur le temps moyen de collecte dans un réseau de capteurs sans fils.

Keywords: Réseaux de capteurs, routage, marche aléatoire, entropie, performance

1 Introduction

A wireless sensor network (WSN) is formed by a large number of sensor nodes deployed over an extended region. Sensor nodes, usually battery-operated, are simple and cheap, and cannot offer plentiful resource. The routing problem is an important issue to be considered here because it is significantly sensitive to the previous constraints for two reasons. First, searching a large space of possible routes –derived from having a large number of nodes– may prove computationally prohibitive for low complexity devices such as sensor nodes. Second, performing explicit route discovery/repair computations and maintaining explicit state information about available routes at the nodes is costly in terms of complexity and energy. Thus, the selection of a routing scheme depends on the *knowledge* available at the network nodes and the communication overhead that can be tolerated. Such knowledge provides nodes with a picture of the network that can be exploited to make decisions. There exist different routing approaches in this respect. In *centralized* approach, each node is provided with *full* topology information, then the shortest path algorithm can be applied. Large networks with reasonably stable nodes over time, where autonomous nodes do not know the full network graph (*e.g.*, the Internet), require a *distributed* routing approach. However, this approach induces significant communication overhead, which is problematic for large scale networks with high constraints (*e.g.*, WSN). A *localized* approach, where nodes make decisions solely based on *partial* information available from neighbors is rather suitable. These approaches leverage the intuitive idea that the more knowledge is available at the network nodes, the more efficient is the routing scheme but against the complexity and the energy expenditure.

Many research works have addressed this tradeoff but only from a qualitative view or by means of simulations [SB02]. In this paper we use the random walk theory to quantitatively estimate the influence of the requisite knowledge on the routing efficiency in WSN. This is motivated by the fact that making appropriate decisions to forward data depends readily on the amount of state information a node holds. Thus, without any state information, nodes would *blindly* forward data. This results in a packet wandering from node to node until reaching its destination. As analyzed in [FML07], the routing problem becomes then a problem of an *unbiased* random walk taking place on a graph. However, if some state information is available at nodes, the induced random walk would be *biased* with a favored direction to enhance performance.



(a) Torus lattice $\mathfrak{T}$

(b) Impact of the bias

Figure 1: (a) Torus lattice $\mathfrak{T}$ is formed by connecting the opposite ends of the square lattice. Hops are represented by red arrows. (b) Mean data gathering delay as a function of the position r_1 for different values of ε and for $N = 15$.

We consider in this paper the case of a periodic lattice network of finite square cells, each containing two kind of nodes: sensor and sink nodes. We choose this simple structure because it fits well the actual properties of some WSN (e.g., mesh or grid networks) while being useful to incorporate theoretical elements to modelize the impact of the requisite knowledge on the routing efficiency.

2 The Model: Definitions and Notation

Let $\mathfrak{C}$ denote a finite square cell of size $N \times N$. The envisioned network is then induced by replicating $\mathfrak{C}$ by rigid translations specified by vectors $N\vec{\mathbf{m}}$, where $\vec{\mathbf{m}} \in \mathbb{Z}^2$. With a large number of cells, we can assume that the formed network is infinite and hence, it is equivalent to a torus lattice $\mathfrak{T}$ formed by connecting the opposite ends of cell $\mathfrak{C}$. Every node $\vec{\mathbf{r}} \in \mathfrak{T}$ is labeled with (r_1, r_2), where r_1 and r_2 are integers such that $0 \le r_1, r_2 \le N-1$. At a given node $\vec{\mathbf{r}}$, let $\mathfrak{p}(\vec{\mathbf{r}}, \vec{\mathbf{s}})$ be a nonnegative real such that $\sum_{\vec{\mathbf{s}} \in \mathfrak{T}} \mathfrak{p}(\vec{\mathbf{r}}, \vec{\mathbf{s}}) = 1$. This defines a transition probability distribution over the nodes of $\mathfrak{T}$ at fixed $\vec{\mathbf{r}}$. When a packet reaches node $\vec{\mathbf{r}}$, the next hop to node $\vec{\mathbf{s}}$ occurs with probability $\mathfrak{p}(\vec{\mathbf{r}}, \vec{\mathbf{s}})$. The random sequence of nodes selected this way is a *random walk* on $\mathfrak{T}$, and $\mathfrak{p}(\vec{\mathbf{r}}, \vec{\mathbf{s}})$ is called the *transition function* of the random walk [Hug95]. By making different assumptions on the topology of the underlying network and on constraints imposed on $\mathfrak{p}(\vec{\mathbf{r}}, \vec{\mathbf{s}})$, we are able to explore a large space of possible routing schemes. In particular, let us consider N sink nodes corresponding to the set $\mathcal{C}$ of nodes $\vec{\mathbf{s}}_j = (0, j)$ where $j = 0, \cdots, N-1$. Set $\mathcal{C}$ is called the gathering border. The other nodes are members of $\mathcal{S}$, the set of sensor nodes. The data generated by all sensor nodes are collected by the sink nodes without any specific mapping between sensor nodes and sink nodes. At a given sensor node $\vec{\mathbf{r}}$, we assume that the next hop occurs only to the 4-nearest neighbors with probability $\frac{1}{2}(1-\varepsilon)$ to the right neighbor, $\frac{1}{2}\varepsilon$ to the left neighbor ($0 \le \varepsilon \le 1$), and $\frac{1}{4}$ to either top or bottom neighbor (Fig. 1(a)).

Key probabilities from which main results can be derived are $P_n(\vec{\mathbf{r}}, \vec{\mathbf{s}})$ the probability of being at node $\vec{\mathbf{s}}$ after n hops, given that a packet has been issued at node $\vec{\mathbf{r}}$, and $F_n(\vec{\mathbf{r}}, \vec{\mathbf{s}})$ the probability of arriving at node $\vec{\mathbf{s}}$ for the *first* time on the nth hop, given that the packet started at node $\vec{\mathbf{r}}$. They are called the *node occupation probability* and the *first-passage probability* respectively. We make use in this paper of the *generating function* formalism to deal with a sequence $\{c_n\}_{n \in \mathbb{N}}$ by capturing all these coefficients into a formal infinite series defined as $C(z) = \sum_{n=0}^{\infty} c_n z^n$ where the complex variable z is small enough to ensure the convergence of this series. $C(z)$ is called the generating function associated with $\{c_n\}_{n \in \mathbb{N}}$. Thus, we denote the generating functions associated with $P_n(\vec{\mathbf{r}}, \vec{\mathbf{s}})$ and $F_n(\vec{\mathbf{r}}, \vec{\mathbf{s}})$ as $P(\vec{\mathbf{r}}, \vec{\mathbf{s}}|z)$ and $F(\vec{\mathbf{r}}, \vec{\mathbf{s}}|z)$ respectively. $P(\vec{\mathbf{r}}, \vec{\mathbf{s}}|z)$ and $F(\vec{\mathbf{r}}, \vec{\mathbf{s}}|z)$ are related to each other according to a classical relation extensively used in random walk theory

$$P(\vec{\mathbf{r}}, \vec{\mathbf{s}}\,|\,z) = \delta_{\vec{\mathbf{r}}\vec{\mathbf{s}}} + F(\vec{\mathbf{r}}, \vec{\mathbf{s}}\,|\,z)\,P(\vec{\mathbf{s}}, \vec{\mathbf{s}}\,|\,z), \qquad \vec{\mathbf{r}}, \vec{\mathbf{s}} \in \mathfrak{T}. \tag{1}$$

The proof of this relation is based on the law of total probability and can be found in [Hug95].

3 Requisite Knowledge vs. Routing Efficiency

Before estimating analytically the influence of the requisite knowledge on the routing efficiency, let us give an interpretation of parameter ε. If H_ε denotes the *entropy* associated with the transition probabilities at a given sensor node, we can then write $H_\varepsilon = \frac{3}{2} - \frac{1}{2}\big(\varepsilon \log_2(\varepsilon) + (1-\varepsilon)\log_2(1-\varepsilon)\big)$. However, from information theory, H_ε describes how much uncertainty (or knowledge) is carried by a single transition. Therefore, parameter ε can be considered as an estimator of the amount of requisite knowledge available at a given sensor node. Note that for $\varepsilon = \frac{1}{2}$, H_ε is maximal which corresponds to an equally likely hops. Otherwise, the uncertainty is lower, which corresponds to transitions with a *favored* direction leading to performance enhancement.

3.1 Mean Data Gathering Delay: a Routing Efficiency Indicator

The *data gathering delay* of a packet issued from sensor node $\vec{\mathbf{r}}$ is the time or the number of hops it takes to reach $\mathcal{C}$ where it is trapped. This time is a random variable denoted by $D_\varepsilon(\vec{\mathbf{r}})$. We propose in the following to evaluate $E\left(D_\varepsilon(\vec{\mathbf{r}})\right)$, the *mean* data gathering delay, as an indicator of the routing efficiency. For that, let $G_n(\vec{\mathbf{r}})$ be the probability that a packet issued from sensor node $\vec{\mathbf{r}}$ will be trapped at $\mathcal{C}$ on the nth hop. Therefore, in terms of probability notation we have $Pr\{\mathrm{D}_\varepsilon(\vec{\mathbf{r}}) = n\} = G_n(\vec{\mathbf{r}})$. If $G(\vec{\mathbf{r}}|z)$ denotes the generating function associated with $G_n(\vec{\mathbf{r}})$, $E\left(D_\varepsilon(\vec{\mathbf{r}})\right)$ can be then expressed as

$$E\left(\mathrm{D}_\varepsilon(\vec{\mathbf{r}})\right) = \sum_{n=0}^{\infty} n\, G_n(\vec{\mathbf{r}}) = \partial G(\vec{\mathbf{r}}|z)/\partial z\,|_{z=1}. \tag{2}$$

Let us now make $G(\vec{\mathbf{r}}|z)$ explicit. Using the law of total probability, it can be possible to decompose the event that a packet issued from sensor node $\vec{\mathbf{r}}$ will be trapped at $\mathcal{C}$ on the nth hop, which has the probability $G_n(\vec{\mathbf{r}})$, into the N mutually exclusive events that the packet arrives at the sink node $\vec{\mathbf{s}}_j$ for the *first* time on the nth hop, which has the probability $F_n(\vec{\mathbf{r}},\vec{\mathbf{s}}_j)$. Thus, we obtain $G_n(\vec{\mathbf{r}}) = \sum_{j=0}^{N-1} F_n(\vec{\mathbf{r}},\vec{\mathbf{s}}_j)$. The n-dependence can be eliminated from this equation by multiplying both sides by z^n and summing over all n. Therefore, we obtain $G(\vec{\mathbf{r}}|z) = \sum_{j=0}^{N-1} F(\vec{\mathbf{r}},\vec{\mathbf{s}}_j|z)$. However, from (1), $F(\vec{\mathbf{r}},\vec{\mathbf{s}}_j|z) = P(\vec{\mathbf{r}},\vec{\mathbf{s}}_j|z)/P(\vec{\mathbf{s}}_j,\vec{\mathbf{s}}_j|z)$. Since a packet starting from sink node $\vec{\mathbf{s}}_j$ never leaves, we have $P_n(\vec{\mathbf{s}}_j,\vec{\mathbf{s}}_j) = 1$ for all n, which leads to $P(\vec{\mathbf{s}}_j,\vec{\mathbf{s}}_j|z) = 1/(1-z)$. Hence, we obtain

$$G(\vec{\mathbf{r}}|z) = (1-z)\sum_{j=0}^{N-1} P(\vec{\mathbf{r}},\vec{\mathbf{s}}_j|z). \tag{3}$$

Montroll *et al.* [MS73] evaluated the discrete Fourier transform of the sequence $\{P(\vec{\mathbf{r}},\vec{\mathbf{s}}_j|z)\}_j$ at points $0 \le k \le N-1$. Remarking that the sum involved in (3) is nothing but the value of this discrete Fourier Transform at the origin ($k=0$), (3) can be then rewritten as

$$G(\vec{\mathbf{r}}|z) = \frac{\rho_1(z)^{r_1} - \rho_2(z)^{r_1}}{\rho_1(z)^N - \rho_2(z)^N} + \left(\rho_1(z)\rho_2(z)\right)^{r_1}\frac{\rho_1(z)^{N-r_1} - \rho_2(z)^{N-r_1}}{\rho_1(z)^N - \rho_2(z)^N}, \tag{4}$$

where

$$\left\{\begin{array}{l}\rho_1(z)\\ \rho_2(z)\end{array}\right. = \frac{1 \pm \sqrt{1-4A(z)B(z)}}{2A(z)} \qquad \text{and} \qquad \left\{\begin{array}{l}A(z) = z(1-\varepsilon)/(2-z)\\ B(z) = z\varepsilon/(2-z).\end{array}\right.$$

By plugging (4) into (2), the mean data gathering delay can be simplified as

$$E\left(\mathrm{D}_\varepsilon(\vec{\mathbf{r}})\right) = \frac{2}{(2\varepsilon-1)}\left(r_1 - N\frac{1-(\frac{\varepsilon}{1-\varepsilon})^{r_1}}{1-(\frac{\varepsilon}{1-\varepsilon})^{N}}\right), \qquad \text{for } \varepsilon \ne \frac{1}{2}, 1. \tag{5}$$

By expanding (5) close to $\varepsilon = 1/2$ and $\varepsilon = 1$, it can be deduced that $E\left(\mathrm{D}_\varepsilon(\vec{\mathbf{r}})\right)$ is continuous for all $0 \le \varepsilon \le 1$. We retrieve that $E\left(\mathrm{D}_{1/2}(\vec{\mathbf{r}})\right) = 2\,r_1\,(N-r_1)$ and $E\left(\mathrm{D}_1(\vec{\mathbf{r}})\right) = 2\,r_1$ respectively. Moreover, note that the value of $G(\vec{\mathbf{r}}|z)$ at point $z=1$ equals to unit, which represents the probability that a packet issued from sensor node $\vec{\mathbf{r}}$ is ever trapped by $\mathcal{C}$. This means that the data gathering ensured by the random walk process is *certain*. Remark also that $E\left(\mathrm{D}_{1-\varepsilon}(\vec{\mathbf{r}})\right) = E\left(\mathrm{D}_\varepsilon(N\vec{\mathbf{e}}_1 - \vec{\mathbf{r}})\right)$, which means that the plots of $E\left(\mathrm{D}_\varepsilon(\vec{\mathbf{r}})\right)$ and $E\left(\mathrm{D}_{1-\varepsilon}(\vec{\mathbf{r}})\right)$ as functions of r_1 are symmetric about the axis $X = N/2$. It suffices then to assume that $1/2 \le \varepsilon \le 1$.

3.2 Analysis

Let us now give a physical interpretation of (5). Indeed, if we denote the displacement of a packet issued from sensor node $\vec{\mathbf{r}}$ *after* n hops by $\vec{\mathbf{M}}_n(\vec{\mathbf{r}})$, then $\vec{\mathbf{M}}_n(\vec{\mathbf{r}}) = \sum_{j=1}^{n} \vec{\mathbf{H}}_j(\vec{\mathbf{r}})$ where the random vector $\vec{\mathbf{H}}_j(\vec{\mathbf{r}})$ represents the displacement *on* the jth hop. Since we have the same transition probability distribution over all sensor nodes, the *mean* displacement after n hops is given by $E\left(\vec{\mathbf{M}}_n(\vec{\mathbf{r}})\right) = n\vec{\mu}$ where $\vec{\mu}$ denotes the *mean* displacement on a single hop. Vector $\vec{\mu}$ represents also the bias of the random walk and can be interpreted as the *mean* velocity of propagation of a packet. We have $\vec{\mu} = \frac{1}{2}(1-2\varepsilon)\vec{\mathbf{e}}_1$. Therefore, we deduce that for $1/2 \leq \varepsilon \leq 1$ the negative direction of the X-axis is favored by the walk. In other word, for $1/2 \leq \varepsilon \leq 1$ packets are attracted by the random walk towards the gathering border. A packet issued from sensor node $\vec{\mathbf{r}}$ travels then the distance r_1 before being trapped by the gathering border with a speed $\mid \vec{\mu} \mid = \frac{1}{2}(2\varepsilon - 1)$, hence, the time required by this packet before being trapped is simply the distance divided by the speed, that is, $2r_1/(2\varepsilon-1)$. This result could be retrieved from the first term of (5).

The impact of the requisite knowledge via the bias on the mean data gathering delay is illustrated in Fig. 1(b). We see that the higher are the values of ε (lower entropy), the lower is the delay. Moreover, it turns out that an attractive bias (*i.e.*, $1/2 < \varepsilon \leq 1$) towards the gathering border accelerates the data gathering process whereas a repulsive bias (*i.e.*, $0 \leq \varepsilon < 1/2$) backwards the gathering border decelerates the data gathering process. The case of $\varepsilon = 1/2$ (maximum entropy) corresponds to an *unbiased* walk where no direction is favored. This represents the worst case for the performance of the walk. Note also that for fixed ε such that $1/2 < \varepsilon < 1$, the mean data gathering delay increases with r_1 until it reaches a maximum value from which it decreases as r_1 increases. This observation can be explained by the fact that with the periodicity property of the torus lattice, the gathering border ($X = 0$) is replicated to infinity by rigid translations specified by vectors $mN\vec{\mathbf{e}}_1$ where m is an integer. In particular, the axis $X = N$ corresponds to a line of sink nodes. Hence, when r_1 increases, the attractive effect of the bias is compensated by the farness from the gathering border $X = 0$ until the nearness from the axis $X = N$ takes away and therefore, the mean data gathering delay goes down again.

4 Conclusion

In this paper we have related quantitatively the degree of knowledge to the routing performance and we have studied to what extend the state information available at network nodes can be minimized to reduce the complexity while ensuring an efficient routing scheme. This paradigm arises especially in the design of WSN where the localized approach is extensively embraced. With the aid of random walk theory, we have confirmed analytically the intuitive result that the larger the amount of state information, the more efficient the routing scheme. All details of this model will appear in the full version of this paper.

Acknowledgments

Authors would like to thank Stéphane Paquelet for helpful discussions about the reduction of complexity in large scale systems.

References

[FML07] G. Froc, I. Mabrouki, and X. Lagrange. Design and performance of wireless data collection networks based unicast random routing. 2007. Submitted.

[Hug95] B. D. Hugues. *Random Walks and Random Environments*, volume 1, pages 119–121. Oxford University Press, New York, 1995.

[MS73] E.W. Montroll and H. Scher. Random walks on lattices. continuous-time walks and influence of absorbing boundaries. *Journal of Statistical Physics*, 9(2):101–135, 1973.

[SB02] S. Servetto and G. Barrenechea. Constrained random walks on random graphs: Routing algorithms for large scale wireless sensor networks. In *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, Atlanta, GA, April 2002.