On the data delivery delay taken by random walks in wireless sensor networks

Issam Mabrouki, Gwillerm Froc, Xavier Lagrange

On the Data Delivery Delay taken by Random Walks in Wireless Sensor Networks

Issam MABROUKI
Réseaux, Sécurité et Multimédia
TELECOM Bretagne, Rennes, France
Email: issam.mabrouki@telecom-bretagne.eu

Gwillerm FROC
MITSUBISHI ELECTRIC
ITE-TCL, Rennes, France
Email: froc@tcl.ite.mee.fr

Xavier LAGRANGE
Réseaux, Sécurité et Multimédia
TELECOM Bretagne, Rennes, France
Email: xavier.lagrange@telecom-bretagne.eu

***Abstract*—In recent years, the use of random walk techniques in wireless sensor networks has attracted considerable interest among numerous research efforts. The popularity of this approach is attributed to the natural properties of random walks such as locality, simplicity, low-overhead and inherent robustness. However, throughout the variety of research works that assess the effectiveness of random walk techniques, most results are derived from a qualitative view or by means of simulations. Furthermore, when analytical tools are used, the obtained results often provide bounds on various performance metrics of interest, which may have little consequences for practical applications. Instead, our goal in this paper is to quantify the effectiveness of such techniques based on the derivation of closed-form expressions. In particular, we focus on the data delivery delay taken for the random walk to deliver messages from sensor to sink nodes and study its statistics through closed-form derivations.**

***Index Terms*—Wireless Sensor Networks, Random Walk Theory, Performance Evaluation**

I. Introduction

Wireless Sensor Networks (WSN) have been one of the most prosperous research areas in recent years thanks to its wide spectrum of potential applications, including environment and habitat monitoring, healthcare application, home or industrial automation and control, precision agriculture and inventory tracking [1]. Faced to this general trend of application diversification, large amounts of research being done in the WSN area are trying to provide useful tools and design methods for better architectures and protocols [17].

Most application scenarios for WSN involve small devices called sensor nodes, which are equipped with sensing capabilities, wireless communication and limited power supply, CPU and memory. On top of that, sensor nodes are often supposed to operate unattended and under strict energy constraints. Such adverse conditions make the design of robust, scalable and energy efficient systems a considerable challenge. The extensive research in this field, though, allowed to learn a few principles for the design of efficient WSN [17]. For example, topology-driven algorithms are at a disadvantage for such networks as they induce an excessive amount of communication, which is problematic for WSN.

In the search of an alternative solution, many earlier recent research efforts have investigated the use of randomization to build robust, scalable and energy efficient protocols in the context of WSN [7]. One example consists of the use of random walks to convey data from a source node to a destination one. The use of this technique is not new and has been extensively explored in many networking models providing a variety of algorithms including routing [25], self-stabilization [8], data gathering and query processing in wireless networks [2], [19], [23], peer-to-peer networks [14] and other distributed systems.

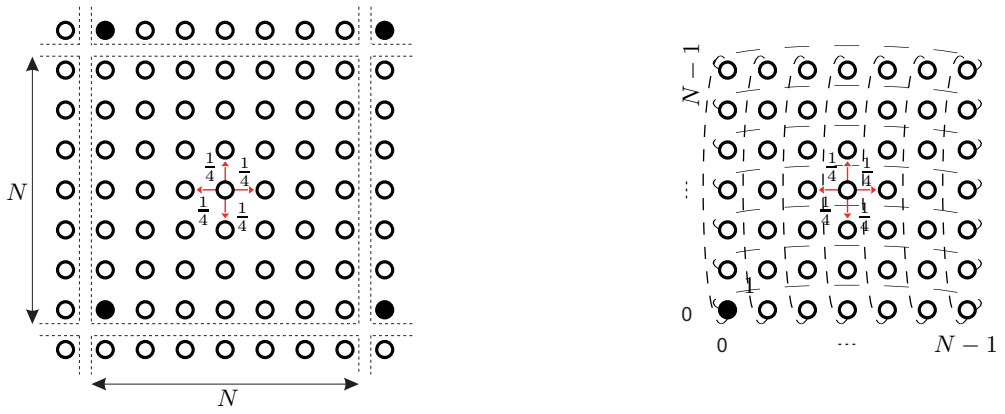
However, throughout the variety of research works that assess the effectiveness of random walk techniques, most results are derived from a qualitative view or by means of simulations [25]. Furthermore when analytical tools are used, the obtained results often provide bounds on various performance metrics of interest [3], [22]. For example, different authors are interested in the well studied concept of cover time, which is the expected time taken by a random walk to visit every node in a graph. This property is relevant to a wide range of algorithmic applications [2], [6], and various methods of bounding the cover time of graphs have been thoroughly investigated [10], [16]. Recently, it has been proven that for any size-n geometric graph with connectivity radius r, when $r = \Theta(r_{con})$[1] then *w.h.p.*[2] the optimal cover time behaves as $O\big(n \ln(n)\big)$ where r_{con} growing as $O(\sqrt{\ln(n)/(\pi n)})$ is the critical radius to guarantee connectivity *w.h.p.* [4].

There are other properties of the random walk also that need to be evaluated. One is the data delivery delay, which is the time required for the random walk to deliver a message from a given node to a destination. This property is of great interest in many WSN applications where the primary task of the network is to gather information from a particular location [9]. Such systems are generally composed of two kinds of nodes: a large number of sensor nodes with limited storage, processing and communication capabilities, and a smaller number of sink nodes with more complex capabilities to gather, process and control data. Each sensor node performs some sensing of a particular confined area, and sends messages to sink nodes through a random walk motion.

Many recent research efforts have studied this random walk property on different classes of graphs. However, most of the obtained results often provide upper bounds on the data

[1]We recall the following notation: (i) $f(n) = O\big(g(n)\big)$ means that exists a constant C and integer K such that $f(n) \leq Cg(n)$ for $n > K$. (ii) $f(n) = \Theta\big(g(n)\big)$ means that $f(n) = O\big(g(n)\big)$; $g(n) = O\big(f(n)\big)$.

[2]Event $\mathcal{E}_m$ occurs with high probability (*w.h.p.*) if probability $\mathbf{Pr}(\mathcal{E}_m)$ is such that $\lim_{m\to\infty} \mathbf{Pr}(\mathcal{E}_m) = 1$.



(a) Original deployment pattern

(b) Equivalent deployment pattern

Fig. 1. Dark-filled circles show sink nodes whereas sensor nodes are represented by light-filled circles.

delivery delay and not a closed form derivation [5], [12]. Clearly, results of this nature may have little consequences for practical applications because the constants hidden in the bounds notation can be very large. Moreover, most of the works concentrate on the evaluation of the mean value of the data delivery delay [13], which makes difficult to get statistically robust conclusions about the relevance of the random walk efficiency.

Instead, our take in this paper is to obtain a complete insight into the data delivery delay induced by a random walk by constructing an analytical model that owes much to the powerful analytic tools developed in the physics community [20], [21], [24]. In particular, this model enables us to study both the central tendency and the dispersion of the data delivery delay through close form derivations. Although our work mainly focuses on the data delivery delay, the level of abstraction of the proposed model is such that it can be applied to investigate other performance metrics of the random walk.

The remainder of the paper is organized as follows. A formal network model description followed by backgrounds and theoretical elements of random walk theory are given in section II. In section III, we focus on the data delivery delay with prior attention to two properties: the mean value and the dispersion. In section IV, we show that the data delivery delay can be approximated by a geometric distribution under some conditions. We provide some conclusions in section V and point out aspects that will be subject of future research.

II. Random Walk Modeling

A. Network Description and Assumptions

We consider a WSN made of two kinds of nodes: a large number of sensor nodes and a smaller number of sink nodes with more complex capabilities to gather, process and control data. Each sensor node performs some sensing of a particular confined area, and sends messages to sink nodes in a multihop fashion, using other sensor nodes as relays and without any specific mapping between sensor and sink nodes. We look more specifically at a common regular and periodic deployment topology where nodes are spread over an area of interest with a square pattern. As illustrated in Fig. 1(a), this pattern is formed by a periodically repeated square unit cell of size $N \times N$ containing N^2 nodes of which $N^2 - 1$ are sensor nodes and one is a sink node.

Based on these underlying assumptions, the data delivery process can be described as follows. When a message reaches a given node, the next hop occurs uniformly at random only to the nearest-neighbors. Thus, in the case of 4-connectivity all the 4 nearest-neighbors away from the current node are equally likely with transition probability $\frac{1}{4}$. A message generated at a given sensor node performs then a random walk until it reaches for the *first* time a sink node where it will be trapped. At this moment, we consider that the data delivery process has occurred with success and then, we no longer care about the outcome of the walk.

With sufficiently large number of unit cells, we can assume that the envisioned network is infinite, and hence the considered deployment pattern corresponds to embedding nodes into the space of 2-dimensional integers, $\mathbb{Z}^2$. Recalling the assumption that there is no specific mapping between sensor and sink nodes, and given the structural periodicity of the envisioned pattern, we can consider that the formed network is equivalent to a torus lattice $\mathfrak{T}$ of size $N \times N$ formed by connecting the opposite ends of a unit cell. The resulting structure, as it is shown in Fig. 1(b), contains a single sink node located at the origin whereas the other sites are sensor nodes. Based on this observation, the original network pattern and the torus lattice are used interchangeably. However, for the sake of simplicity, we actually investigate the data delivery process in reference to torus lattice $\mathfrak{T}$ where every node $\vec{\mathbf{r}}$ is labeled with (r_1, r_2): r_1 and r_2 are integers such that $0 \leq r_1, r_2 \leq N - 1$.

There are many motivations that prompted us to choose this network structure. First, many WSN applications are often desired to follow regular patterns for at least two reasons:

(i) convenience of deployment and (ii) to achieve a higher degree of connectivity and coverage. Second, this division of the network into unit cells suggests a very natural way of grouping nodes together (clustering). Such a clustering is often required by protocols in order to deal with a large number of nodes. Third, it is natural to start with regular patterns before addressing more complex ones of interest to us. Finally, this pattern is simple enough to allow a complete analytical treatment of the random walk problem while still being useful to incorporate specific key issues of WSN such as connectivity and coverage.

B. Problem Formulation and Characterization

We define, for $n \geq 1$, $P_n(\vec{\mathbf{r}}, \vec{\mathbf{s}})$ the probability of being at node $\vec{\mathbf{s}}$ after n hops, given that a message has been issued at node $\vec{\mathbf{r}}$. We also define $F_n(\vec{\mathbf{r}}, \vec{\mathbf{s}})$ the probability of arriving at node $\vec{\mathbf{s}}$ for the *first* time on the nth hop, given that the walk started at node $\vec{\mathbf{r}}$. We shall refer to these probabilities as the *node occupation probability* and the *first-passage probability* respectively. By convention we have $P_0(\vec{\mathbf{r}}, \vec{\mathbf{s}}) = \delta_{\vec{\mathbf{r}}\vec{\mathbf{s}}}$ and $F_0(\vec{\mathbf{r}}, \vec{\mathbf{s}}) = 0$. We denote also by $P(\vec{\mathbf{r}}, \vec{\mathbf{s}}|z)$ and $F(\vec{\mathbf{r}}, \vec{\mathbf{s}}|z)$ the generating functions [18] associated with $\{P_n(\vec{\mathbf{r}}, \vec{\mathbf{s}})\}_{n\in\mathbb{N}}$ and $\{F_n(\vec{\mathbf{r}}, \vec{\mathbf{s}})\}_{n\in\mathbb{N}}$ respectively. Hereafter we present a well-known classical relation extensively used in random walk theory, and upon which the theoretical results derived in this paper rely. For proof, refer to [15].

Lemma 1. *$F(\vec{\mathbf{r}}, \vec{\mathbf{s}} \,|\, z)$ and $P(\vec{\mathbf{r}}, \vec{\mathbf{s}} \,|\, z)$ are related to each other according to the relation*

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

A key issue in random walk problem is the resolution of the following question: how likely does the walk evolve in the future under some initial conditions? Answering this question consists in finding an explicit expression of $P(\vec{\mathbf{r}}, \vec{\mathbf{s}} \,|\, z)$, which completely determines the node occupation probability distribution. Globally, throughout the large number of interdisciplinary works in random walk theory, the exact closed-form solution was mostly carried out under restrictive conditions such as the periodicity of the network, the homogeneity of the system and the infiniteness of the structure on which the random walk takes place. Among these special cases, the problem of random walk on finite lattices with periodic boundary conditions (*i.e.,* toroidal lattices) has been extensively studied. Montroll and Weiss [24] originally proposed this special problem and solved it for k-dimension. In particular, in the case of a torus lattice, they established an explicit expression of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ as follows.

Lemma 2. *$P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ can be expressed as*

$$P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z) = \frac{1}{N^2} \sum_{m_1=0}^{N-1} \sum_{m_2=0}^{N-1} \frac{e^{i\frac{2\pi}{N}m_1 r_1}\ e^{i\frac{2\pi}{N}m_2 r_2}}{1 - \frac{z}{2}\big(\cos(\frac{2\pi}{N}m_1) + \cos(\frac{2\pi}{N}m_2)\big)}. \tag{2}$$

Even though relation (2) does not give a simple form of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$, it is of great importance and represents our basic relation upon which relies most of our contribution in this paper. In the following, we focus our attention on one interesting performance metric of the proposed random walk scheme and attempt to study its statistics through close form derivations. Generating function techniques and asymptotic expansion calculus are the key tools here and they are used in tandem to elaborate interesting results.

III. Data Delivery Delay Analysis

A primary objective of the proposed random walk scheme is to perform successful data delivery while achieving desired guarantees on delay times. It is worth noting that although there are different categories of delay times in WSN depending on application requirements, the time or the number of hops experienced by a message generated at a sensor node before being trapped at the sink node represents one of the significant metrics to measure the random walk performance. We refer to this delay time as the *data delivery delay*.

Formally, given a sensor node $\vec{\mathbf{r}} \in \mathfrak{T}$, let $\mathcal{D}_N(\vec{\mathbf{r}})$ be the data delivery delay defined at this sensor node. $\mathcal{D}_N(\vec{\mathbf{r}})$ is a discrete random variable taking nonnegative integer values. Thus, it can be characterized by its probability distribution. Once the probability distribution of $\mathcal{D}_N(\vec{\mathbf{r}})$ is determined, all the statistics of the data delivery delay can be theoretically derived. Unfortunately, it is not a simple matter to derive this probability distribution. This difficulty is well known in random walk theory while studying first-passage times [15], [26]. Indeed, most of the analytical works in this respect concentrate on the derivation of either the short- or the long-time behaviors of first-passage times and not on the exact expression of the probability distribution.

In spite of this, we show in this paper that many statistical properties of the data delivery delay can be deduced by calculating its first- or higher-order moments using generating function techniques. In particular, we are interested in this section in central tendency, dispersion and their dependencies on key model parameters. These two basic properties can be reflected by calculating for example the mean value and the standard deviation of random variable $\mathcal{D}_N(\vec{\mathbf{r}})$. To do this, we first determine the generating function associated with random variable $\mathcal{D}_N(\vec{\mathbf{r}})$ and then, we apply the generating function techniques.

Proposition 1. *Let $\mathcal{D}_N(\vec{\mathbf{r}} \,|\, z)$ be the generating function of random variable $\mathcal{D}_N(\vec{\mathbf{r}})$. $\mathcal{D}_N(\vec{\mathbf{r}} \,|\, z)$ can be then expressed as*

$$\mathcal{D}_N(\vec{\mathbf{r}} \,|\, z) = \frac{P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)}{P(\mathbf{0}, \mathbf{0} \,|\, z)}. \tag{3}$$

Proof: From the definition of $\mathcal{D}_N(\vec{\mathbf{r}})$, we can see that the probability of a message generated at sensor node $\vec{\mathbf{r}}$ to be gathered at the sink node on the n-th hop is equal to the probability of being at the origin for the first time on the n-th hop, given that the walk started at node $\vec{\mathbf{r}}$, that is, first-passage probability $F_n(\vec{\mathbf{r}}, \mathbf{0})$. Thus, in terms of probability notation, we have

$$\mathbf{Pr}\big\{\mathcal{D}_N(\vec{\mathbf{r}}) = n\big\} = F_n(\vec{\mathbf{r}}, \mathbf{0}), \qquad n \geq 0.$$

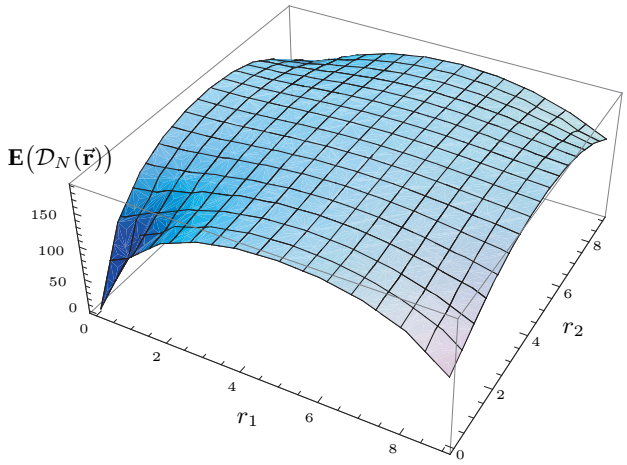


Fig. 2. Spatial distribution of the mean value of the data delivery delay.

By multiplying both sides of this relation by z^n and summing over all n, we obtain $\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z) = F(\vec{\mathbf{r}}, \mathbf{0}\,|\,z)$. Finally, from Lemma 1, we obtain (3). ■

A. Mean Value

1) Generating Function Analysis: Generally, the most frequently used measure for describing central tendency is the mean value or in probability parlance the expectation. We propose here to study this property for the data delivery delay by calculating its expectation denoted by $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$. Applying generating function techniques, mean value $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ can be derived by taking the limit of the first-derivative of generating function $\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z)$ as $z \to 1^-$, so that

$$\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big) = \lim_{z \to 1^-} \frac{\partial}{\partial z} \mathcal{D}_N(\vec{\mathbf{r}}\,|\,z).$$

Although this relation does not give an explicit solution of the mean value, we can fortunately extract a closed-form expression from the Taylor's series expansion of generating function $\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z)$ at point $z = 1$. Indeed, if $\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z)$ is holomorphic at $z = 1$, then the limit of its first derivative with respect to z as $z \to 1^-$ is nothing but the first-order term of its Taylor's series expansion at point $z = 1$. To evaluate this Taylor's series expansion, we proceed as follows. We first estimate the asymptotic expansion of $P(\vec{\mathbf{r}}, \mathbf{0}\,|\,z)$ as $z \to 1^-$, which is provided by (16) in Appendix A. Second, by setting $\vec{\mathbf{r}} = \mathbf{0}$ in (16), we derive the asymptotic expansion of $P(\mathbf{0}, \mathbf{0}\,|\,z)$ as $z \to 1^-$ and then, after elementary calculus, we prove that function $\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z) = P(\vec{\mathbf{r}}, \mathbf{0}\,|\,z)/P(\mathbf{0}, \mathbf{0}\,|\,z)$ is holomorphic at $z = 1$. So, it can be represented by its Taylor's series expansion at point $z = 1$, which is given by (17) in Appendix B. By differentiating (17) with respect to z and taking the limit as $z \to 1^-$, we obtain the following result:

Result 1. *Mean value $\boldsymbol{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ can be written in a closed-form expression as*

$$\boldsymbol{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big) = N^2\big(\varphi_N(\mathbf{0}, 1) - \varphi_N(\vec{\mathbf{r}}, 1)\big) + 2r_2(N - r_2), \quad (4)$$

where $\varphi_N(\vec{\mathbf{r}}, z)$ is defined by (13) in Appendix A.

2) Discussion: Some general remarks can be drawn from the previous analysis. First, from the Taylor's series expansion of generating function $\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z)$ at point $z = 1$ given by (17) in Appendix B, the value of $\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z)$ at point $z = 1$ is equal to unit, which represents the probability that a message issued from sensor node $\vec{\mathbf{r}}$ is ever gathered by the sink node. This means that the proposed data delivery process is certain. This result is at first glance surprising since we have assumed that the original network is infinite (a large number of unit cells), which implies that there would be potential infinite paths allowing messages to escape sink nodes and diffuse indefinitely around the entire network without being trapped. However, since the network is equivalent to a torus lattice, a *finite-size* effect arises precluding messages to continuously move away from sink nodes.

Second, referring to (4), the mean value of the data delivery delay is finite. This is not a trivial property because it means that only short paths between sensor and sink nodes are significant. The main reason is that, with a regular and periodic deployment of sink nodes, there are no network areas where a message can spend a lot of time without meeting a sink node. Third, the dependance of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ on parameter N and hence, indirectly on the concentration of sink nodes highlights the scalability property of the proposed scheme. That is, if additional sensor nodes are added, the mean value does not change provided that the concentration of sink nodes (or equivalently parameter N) remains constant. This is once again a simple manifestation of the regular and periodic deployment of sink nodes, which ensures that the maximum distance between sensor and sink nodes ($\sim N$) is very small compared to the effective network size.

Let us now study the spatial distribution of expectation $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ over torus lattice $\mathfrak{T}$. A first comment is that, for symmetric considerations, coordinates r_1 and r_2 play similar roles in the expression of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ although at first glance it is not obviously straightforward from (4). This property is observable in Fig. 2 where $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ is depicted as a function of position $\vec{\mathbf{r}} = (r_1, r_2)$ over a unit cell of size 10×10. Other values of parameter N are possible to plot, however, our goal is to show numerical examples of practical interest. It should also be emphasized that the observed properties are not limited to value $N = 10$ but are valid for all finite ones.

The first important feature of Fig. 2 is the fact that the mean number of hops required to reach the sink node increases initially with a high rate as we move away from the origin towards the middle region, where $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ saturates. The relatively low value of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ observed around the origin (or the sink node) means that messages generated in vicinity of the sink node do not spend a lot of time exploring the neighborhood they were created in and they are trapped quickly at early time. It is then found the closer a sensor node is to the sink node, the faster the data delivery process is performed. However, far away from the sink node, generated messages can escape their initial territory and visit more new sensor nodes before being trapped, which argues the higher values of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ with low slope in the middle of the unit cell. As we

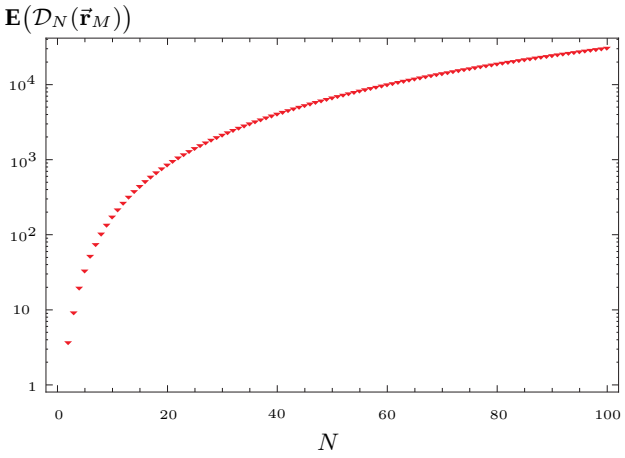


Fig. 3. Semi-logarithmic plot of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}}_M)\big)$ as a function of N.

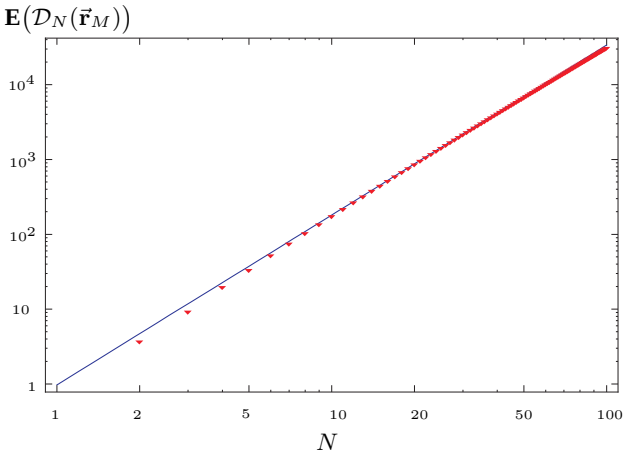


Fig. 4. Fitting of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}}_M)\big)$ to a power law.

can see in the plot, though, there is a crossover in this behavior. Beyond the middle region, the value of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ decreases again with low slope as we move towards the other corners. This crossover behavior is due to the periodicity property of the torus lattice, which means that the opposite corners of the unit cell are connected.

To get deeper insight into the mean value efficiency induced by the random walk scheme, we propose now to compare it with the analogue one if the shortest path routing (SPR) scheme is used. This implies that whenever a message is generated at a given sensor node, it will be rather trapped by the closest sink node. At a given sensor node $\vec{\mathbf{r}}$, let us define $\mathcal{D}_N^{SPR}(\vec{\mathbf{r}})$ as the data delivery delay induced by the SPR scheme. Therefore, we readily obtain $\mathcal{D}_N^{SPR}(\vec{\mathbf{r}}) = \min(r_1, N - r_1) + \min(r_2, N - r_2)$. As we should expect, the SPR scheme will achieve the optimal data delivery delay due to its deterministic nature. For example, the minimum/maximum mean values achieved by the random walk scheme for a unit cell of size 10×10 exceeds the one achieved by the SPR scheme about one hundred/19 times respectively. In fact, this efficiency is achieved at the expense of sophisticated capabilities required for explicit route discovery/repair computations, and maintaining explicit state information about available routes at the nodes. However, optimality is not generally a fundamental issue in many WSN applications. Instead, the focus is more on properties such as scalability, robustness and load balancing.

We investigate now the worst case of the mean value efficiency, which corresponds to the maximum value of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$. This maximum value corresponds to the maximum possible distance between an arbitrary sensor node and its nearest sink node. This maximum distance occurs in the middle of the unit cell. Thus, if N is even, $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ is maximum only at the single position $\vec{\mathbf{r}} = (\frac{N}{2}, \frac{N}{2})$. In contrast, if N is odd, the maximum value of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ is reached at four positions, which are $\vec{\mathbf{r}} = (\lfloor \frac{N}{2} \rfloor + i, \lfloor \frac{N}{2} \rfloor + j)$ where the $\lfloor . \rfloor$ symbol stands for the floor function and $i, j = 0, 1$. Merging the two cases into one, we can see that the maximum value of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ is always reached at position $\vec{\mathbf{r}}_M = (\lfloor \frac{N}{2} \rfloor, \lfloor \frac{N}{2} \rfloor)$.

By a linear scale plot, we can show that $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}}_M)\big)$ increases initially with a slow rate and then the increase is almost superlinear. However, the large range of produced values makes a semi-logarithmic scale more appropriate. This is illustrated in Fig. 3. The obtained curve suggests that $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}}_M)\big)$ may follow a power law. The log-log plot described in Fig. 4, where both horizontal and vertical axis are plotted on a logarithmic scale confirms our prediction of a power law growth of $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}}_M)\big)$. Indeed, by a linear regression model, we can estimate the slope of the best straight line fitted to this plot. Finally, the result of the fit is $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}}_M)\big) \simeq 0.97 N^{2.27}$ and the quality of the fit is very satisfactory with a coefficient of correlation equal to 0.99. Although this procedure is partially heuristic, the obtained fit formula could help practitioners to rapidly determine the concentration of sink nodes to be deployed in order to guarantee an upper bound of the mean value.

We turn now our attention to a simple example on how the previous analysis can be applied to a concrete scenario. We consider IEEE $802.15.4$ enabled nodes with supported data transmission rate of 20 kbps. One reason for choosing the IEEE $802.15.4$ standard stems from its suitability for the development and deployment of WSN. If we don't take into account the physical limitations posed by the radio medium, the time required for a message to perform a single hop is simply its size divided by the data transmission rate. Therefore, it is interesting to study the effect of varying the message size on the maximum mean value. This is illustrated in Fig. 5 where $\mathbf{E}\big(\mathcal{D}_N(\vec{\mathbf{r}}_M)\big)$ (in seconds) is plotted as a function of parameter N for different message sizes.

As indicated in Fig. 5, the increase in message size at fixed values of parameter N leads to performance degradation. On the other hand, at a given message size, it is always possible to enhance the mean value efficiency by reducing parameter N, or say equivalently by increasing the concentration of sink nodes. However, this enhancement is lower-bounded by the maximum mean value achievable at the smallest unit cell size,

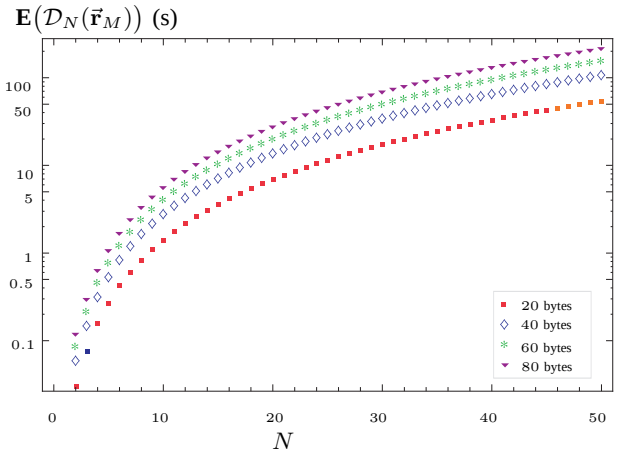


Fig. 5. Effect of message size on the maximum mean value.

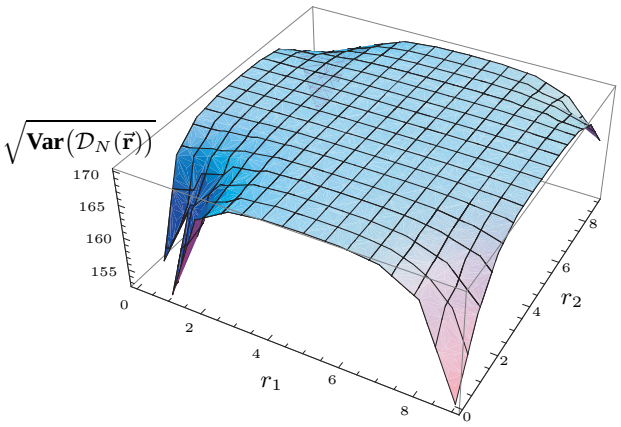


Fig. 6. Spatial distribution of the standard deviation.

which corresponds to $N = 2$. To illustrate this assertion, we can see from Fig. 5 that this lower bound is about 32 ms for a message size of 20 bytes, whereas it is equal to 128 ms for a message size of 80 bytes.

Since in many applications of WSN sensor nodes often send only beep-like small messages to sink nodes to report their status, we find that routing messages based on random walk can achieve acceptable delay provided that the concentration of sink nodes is carefully tuned. For illustration purposes, we can see from Fig. 5 that in order to guarantee a delay threshold of 1 s, the unit cell size should not exceed $N^2 = 8^2, 6^2, 5^2, 4^2$ nodes for a message size of $20, 40, 60, 80$ bytes respectively. Thus, we can conclude that the random walk scheme can be viable in the scenarios where (i) applications do not require too stringent delay, (ii) the sensed data to be transmitted are of small sizes (tens bytes) and (iii) the concentration of sink nodes to be deployed is carefully managed.

B. Dispersion

1) Generating Function Analysis: The dispersion of the data delivery delay is another basic property that indicates the degree of spread around the mean value. A commonly-used measure of dispersion is the standard deviation, which is the square root of the variance. In the context of the proposed random walk scheme, by knowing the standard deviation of the data delivery delay, we should be able to make statistically robust conclusions about the relevance of the mean value that was previously calculated. Indeed, a small standard deviation indicates that the values actually taken by the data delivery delay are clustered closely around the mean value thereby its probability distribution can be reasonably summarized by the mean value. However, a large standard deviation indicates that they are far from the mean value and hence the latter is a bad representative measure for the data delivery delay.

We propose here to derive the standard deviation of the data delivery delay and study its dependence on key model parameters. This study will provide us a somewhat different picture of the data delivery delay efficiency than we have previously obtained by considering the mean value alone. Denoting the variance by $\mathbf{Var}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$, the standard deviation is found by computing first the variance using generating function techniques and then taking the square root. So that,

$$\mathbf{Var}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big) = \lim_{z\to 1^-} \frac{\partial^2}{\partial z^2}\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z) + \lim_{z\to 1^-} \frac{\partial}{\partial z}\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z) - \left\{ \lim_{z\to 1^-} \frac{\partial}{\partial z}\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z) \right\}^2. \tag{5}$$

We can see from this relation that only the first and second partial derivatives of generating function $\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z)$ with respect to z at $z = 1^-$ are involved while defining the variance of the data delivery delay. In a similar way as that is used for the calculation of the mean value, these partial derivatives can be expressed in an explicit form using the second order Taylor's series expansion of $\mathcal{D}_N(\vec{\mathbf{r}}\,|\,z)$ at point $z = 1$, which is provided by (17) in Appendix B. Plugging (17) into (5) leads to the following result:

Result 2. *The variance of the data delivery delay can be written in exact closed-form expression as*

$$\begin{aligned}\mathbf{Var}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big) &= 4N^2 r_2(N-r_2)\,\varphi_N(\vec{\mathbf{r}},1) \\ &+ \frac{1}{3}N^2(1+2N^2)\big(\varphi_N(\mathbf{0},1) - \varphi_N(\vec{\mathbf{r}},1)\big) \\ &+ 2N^2\big(\varphi_N^{(1)}(\mathbf{0},1) - \varphi_N^{(1)}(\vec{\mathbf{r}},1)\big) + N^4\big(\varphi_N^2(\mathbf{0},1) - \varphi_N^2(\vec{\mathbf{r}},1)\big) \\ &+ \frac{2}{3}r_2(N-r_2)\big(2N^2 - 4r_2(N-r_2) - 1\big)\end{aligned} \tag{6}$$

where $\varphi_N(\vec{\mathbf{r}},z)$ is defined by (13) in Appendix A.

2) Discussion: Using this result, it is now straightforward to derive the standard deviation by taking the square root of both sides of (6). In the following, we propose to illustrate numerically the basic properties of the standard deviation and to revisit the main question, previously raised, as to whether the mean value already investigated would be a good summary of the data delivery delay. Fig. 6 depicts the typical form of how the standard deviation varies over the network. We set here $N = 10$. An important property to note from Fig. 6 is

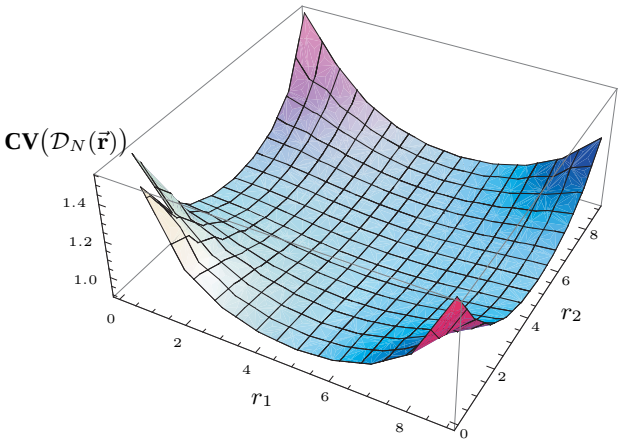


Fig. 7. Spatial distribution of the coefficient of variation.

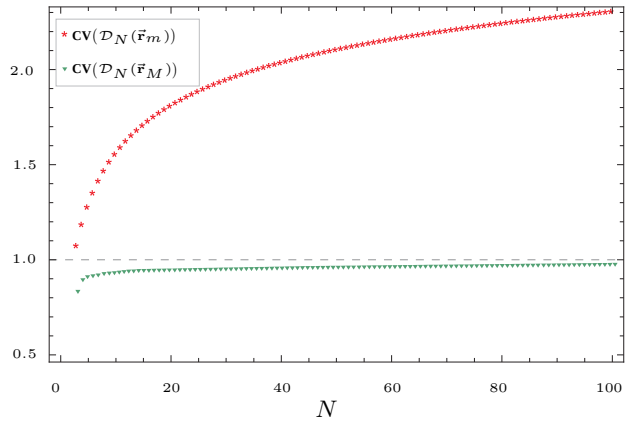


Fig. 8. Effect of N on the maximum and minimum coefficient of variation.

that the standard deviation exhibits a closely similar behavior to that found while studying the mean value. Indeed, the basic pattern is that, except near the corners or equivalently near any sink node, the standard deviation is reasonably similar throughout the network and achieves its maximum value in the middle of the unit cell.

At first glance, the relatively small values of the standard deviation near the corners may mislead us to believe that the mean value can be considered as a good indicator for the data delivery delay. With the benefit of hindsight, we rather find that near the corners the standard deviation always exceeds the mean value regardless of the unit cell size. Furthermore, as we move away from the corners of the unit cell towards the middle region, the standard deviation increases rapidly but becomes nearly constant over most of the unit cell. This high dispersion results in a significant probability that the effective data delivery delay will be greatly different from the mean value. Therefore, the mean value provides us little information about what the data delivery delay actually will be, and it should not be interpreted as a representative value. It is largely for this reason that a treatment of the data delivery delay based only on the mean value is inadequate.

To have a deeper analysis of this question, it is now convenient to define $\mathbf{CV}\big(\mathcal{D}_N(\vec{\mathbf{r}})\big)$ as the coefficient of variation of the data delivery delay, which is defined as the ratio of the standard deviation to the mean value. Fig. 7 shows the spatial distribution of the coefficient of variation over a unit cell for $N = 10$. A key observation to note from this plot is that the coefficient of variation roughly changes around the corners but becomes nearly constant as we move far away. Note also that values greater than 1 are achieved around the corners where a maximum value is reached at node $\vec{\mathbf{r}}_m = (0, 1)$, but a slight decrease lying over a broad and flat plateau occurs towards the middle region in which a minimum value lower than 1 is achieved at $\vec{\mathbf{r}}_M = (\lfloor \frac{N}{2} \rfloor, \lfloor \frac{N}{2} \rfloor)$. The expanse of the plateau can be characterized by more than 79% and 86% of the achieved values lying in the range $[0.9, 1.1]$ for $N = 10$ and 20 respectively.

This peculiar behavior can be attributed to the *proximity-sink* effect. Indeed, consider a message that is generated in the vicinity of a given sink node. As already demonstrated while studying the mean value, this message does not spend a lot of time exploring the neighborhood it was created in and it is trapped quickly at early time on the average. If such a message happens to avoid the sink node, it is likely to have done so by reducing its proximity to the sink node, thereby significantly decreasing its likelihood of being trapped soon and tending to occupy as many nodes as possible on its way. This leads to very high variability, and so the standard deviation will greatly exceed the mean value. In contrast, when generated in the middle region, the message does not initially find itself in a preferential position, so any escape from its initial territory will likely transport it towards one of the four nearest sink nodes around it, thereby significantly increasing its likelihood of being trapped soon. This yields again high variability but the difference between the standard deviation and the mean value becomes no longer pronounced.

Let us now concentrate more specifically on the maximum and minimum values of the coefficient of variation achieved near the corners (at sensor node $\vec{\mathbf{r}}_m = (0, 1)$) and at the middle of the unit cell (at sensor node $\vec{\mathbf{r}}_M = (\lfloor \frac{N}{2} \rfloor, \lfloor \frac{N}{2} \rfloor)$) respectively. We are interested here in how these typical values depend on parameter N or equivalently on the concentration of sink nodes. Our interest stems from the fact that the mean value, as previously demonstrated, achieves its minimum and maximum at sensor nodes $\vec{\mathbf{r}}_m = (0, 1)$ and $\vec{\mathbf{r}}_M = (\lfloor \frac{N}{2} \rfloor, \lfloor \frac{N}{2} \rfloor)$ respectively, which makes particularly the following analysis a good benchmark for the data delivery delay efficiency.

Fig. 8 shows that the maximum value of the coefficient of variation, almost above 1, monotonically grows with N whereas the minimum value, markedly less sensitive to parameter N, varies a little but rapidly equilibrates to the constant value 1 as N increases. The latter behavior suggests a good evidence for nearly memorylessness property which characterizes both the exponential distribution and the geometric distribution with very small probability of success [11].

A heuristic explanation of such a property can be given as follows. Consider a message that is initially generated at

sensor node $\vec{\mathbf{r}}_M = (\lfloor \frac{N}{2} \rfloor, \lfloor \frac{N}{2} \rfloor)$. For sufficiently large values of N, it is always possible to neglect the distance traveled by the message during a given hop number n_0 in comparison with the distance required for the message to be trapped at the sink node. Thus, if the message has not been trapped by hop n_0, the probability that it will be trapped after an additional n hops can be approximated closely by the original probability that it will be trapped after the very first n-th hop. This statement is nothing but one of the characterization of memorylessness property. Formally, this can be expressed in the form of $\mathbf{Pr}\{\mathcal{D}_N(\vec{\mathbf{r}}_M) > n_0 + n \mid \mathcal{D}_N(\vec{\mathbf{r}}_M) > n_0\} = \mathbf{Pr}\{\mathcal{D}_N(\vec{\mathbf{r}}_M) > n\}$. In passing, note that this property is not restricted to sensor node $\vec{\mathbf{r}} = (\lfloor \frac{N}{2} \rfloor, \lfloor \frac{N}{2} \rfloor)$, which corresponds to the maximum possible distance between an arbitrary sensor node and its nearest sink node, but can be extended to the surrounding region provided that N is large enough.

IV. Geometric Distribution Approximation

The analysis of dispersion has shown a high spread of the data delivery delay around the mean value regardless of sensor node positions. As a consequence, the mean value is a poor indicator for the data delivery delay efficiency. It is therefore interesting to look at upper bounds for the data delivery delay rather than the mean value. The cumulative distribution function of the data delivery delay would give the probability that this delay is lower than a certain threshold. Unfortunately, the intractability of the exact probability distribution function makes this task difficult to be carried out. However, exploiting memorylessness property observed at the middle of the unit cell, and since it is also discrete, the data delivery delay at sensor node $\vec{\mathbf{r}}_M = (\lfloor \frac{N}{2} \rfloor, \lfloor \frac{N}{2} \rfloor)$ can be reasonably approximated by a geometric random variable provided that N is sufficiently large (small concentration of sink nodes).

The geometric distribution by definition deals with a series of independent, identical trials (here, moves across nodes) each of which has only two possible outcomes, success or failure (here, trapping or move on) with probabilities that are constant from trial to trial. The geometric random variable is the number of trials needed to get one success (here, trapping), supported on the set $\{1, 2, 3, \cdots\}$. Formally, we can therefore express approximately the probability that a message initially released at sensor node $\vec{\mathbf{r}}_M = (\lfloor \frac{N}{2} \rfloor, \lfloor \frac{N}{2} \rfloor)$ will be trapped by the n-th hop in the form of $\mathbf{Pr}\{\mathcal{D}_N(\vec{\mathbf{r}}_M) = n\} \simeq a(1-a)^{n-1}$ for all $n \geq 1$, where $0 < a < 1$ is the geometric parameter also referred to as the success probability [11]. In our case, this parameter corresponds to the probability that the message will be trapped while moving from one node to another.

In this simplest version of the model, only one parameter, namely a, determines the geometric distribution approximation. An estimate of a can be quickly obtained by remarking that expectation $\mathbf{E}(\mathcal{D}_N(\vec{\mathbf{r}}_M))$ can be related to a as $\mathbf{E}(\mathcal{D}_N(\vec{\mathbf{r}}_M)) = 1/a$ if random variable $\mathcal{D}_N(\vec{\mathbf{r}}_M)$ is geometrically distributed. Under such a condition, the variance of $\mathcal{D}_N(\vec{\mathbf{r}}_M)$ can be also written as $\mathbf{Var}(\mathcal{D}_N(\vec{\mathbf{r}}_M)) = (1-a)/a^2$. In passing, note that the given estimate of a has obviously the necessary condition that $\mathbf{E}(\mathcal{D}_N(\vec{\mathbf{r}}_M)) > 1$, which is consistent with the results obtained while studying the central tendency of the data delivery delay. The cumulative distribution function of the data delivery delay at sensor node $\vec{\mathbf{r}}_M$ can be then approximately written as $\mathbf{Pr}\{\mathcal{D}_N(\vec{\mathbf{r}}_M) \leq n\} \simeq 1 - \{1 - 1/E(\mathcal{D}_N(\vec{\mathbf{r}}_M))\}^n$. Knowing the exact value of $E(\mathcal{D}_N(\vec{\mathbf{r}}_M))$, which is given by Result 1, it is then possible to infer this cumulative distribution function as a function of parameter N. More interestingly enough, this enables us to find the value of N for a given percentile, or equivalently the concentration of sink nodes to be deployed while achieving a threshold delay with high probability.

It is now possible to assess the quality of the mean value achieved at sensor node $\vec{\mathbf{r}}_M$ by answering the question how likely the data delivery delay is effectively below this mean value. In probability parlance, this consists in calculating the probability

$$\mathbf{Pr}\{\mathcal{D}_N(\vec{\mathbf{r}}_M) \leq \mathbf{E}(\mathcal{D}_N(\vec{\mathbf{r}}_M))\} = 1 - \{1 - \frac{1}{E(\mathcal{D}_N(\vec{\mathbf{r}}_M))}\}^{E(\mathcal{D}_N(\vec{\mathbf{r}}_M))}.$$

Remarking that $1/E(\mathcal{D}_N(\vec{\mathbf{r}}_M))$ decreases rapidly to zero as N increases (recall that $\mathbf{E}(\mathcal{D}_N(\vec{\mathbf{r}}_M)) \simeq 0.97N^{2.27}$), the above probability can be reasonably approximated by the constant value $1 - 1/e \simeq 0.63$. Concretely, this value indicates that nearly two-fifth of messages issued at the middle of the unit cell would experience a delay time larger than the mean value. This high variability makes it interesting to look at upper bounds for the data delivery delay. For example, applications with strict requirements on delay would benefit from estimating the data delivery delay with a high percentile. For instance, we find that the 95-percentile is about three times as much as the mean value regardless of the unit cell size. Equivalently, this means that only 5% of the messages issued at the middle of the unit cell would spend at least three times as long as the mean value before being trapped.

V. Conclusion

In this paper we addressed the problem of random walk to model data delivery in wireless sensor networks. In this approach, a packet generated from a given sensor node performs a random motion until reaching a sink node where it is collected. The primary objective of this work was to study the data delivery delay induced by a random walk taking place on a regular and periodic topology.

Three different studies were reported. First, we studied central tendency of the data delivery delay and derived a closed-form expression for its mean value as a function of sink node concentration. This study showed that the random walk can achieve acceptable performance in terms of data delivery delay in the scenarios where (i) applications do not require too stringent delay, (ii) the sensed data to be transmitted are of small sizes (tens bytes) and (iii) the concentration of sink nodes to be deployed is carefully tuned. Second, we derived a closed-form expression for the variance and showed how it varies with the concentration of sink nodes. However,

the study of dispersion revealed a high spread of the data delivery delay around the mean value. So, a treatment of the data delivery delay based only on the mean value is inadequate. Third, the analysis of the coefficient of variation for small concentration of sink nodes argued for a reasonably memorylessness property and suggested a geometric distribution approximation of the data delivery delay. Exploiting this approximation, we assessed the effective quality of the mean value as a representative indicator for the data delivery delay. Indeed, we found that nearly two-fifth of messages issued far away from sink nodes would experience a data delivery delay larger than the mean value, and hence, applications with severe requirements on delay would rather benefit from estimating the data delivery delay with a high percentile.

To improve the data delivery delay efficiency, it is therefore necessary to avoid the "undesirable" aspects of random walks like visiting a node more than once or moving in the wrong direction. Several interesting ideas in this direction can be investigated in further work. For example, we can consider biased random walks that give priority to unvisited neighbors instead of choosing uniformly at random or issuing multiple copies of a message in parallel.

APPENDIX A
ASYMPTOTIC EXPANSION OF $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$

Before calculating the asymptotic expansion of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ as $z \to 1^-$, we propose first to simplify the expression of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ given by (2) and to study its singularity at $z = 1$. By factorizing the denominator of the summand and using the addition theorems of trigonometric functions, we obtain

$$\begin{aligned} P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z) &= \frac{1}{N^2} \sum_{m_1=0}^{N-1} \sum_{m_2=0}^{N-1} \left\{ \frac{e^{i\frac{2\pi}{N} m_1 r_1}}{1 - \frac{z}{2}\cos(\frac{2\pi}{N} m_1)} \right. \\ &\qquad \left. \times \frac{e^{i\frac{2\pi}{N} m_2 r_2}}{1 - c_{m_1}(z)\cos(\frac{2\pi}{N} m_2)} \right\} \\ &= \frac{1}{N^2} \sum_{m_1=0}^{N-1} \left\{ \frac{e^{i\frac{2\pi}{N} m_1 r_1}}{1 - \frac{z}{2}\cos(\frac{2\pi}{N} m_1)} \times \mathcal{S}_{m_1}(z) \right\} \end{aligned} \tag{7}$$

where functions $c_{m_1}(z)$ and $\mathcal{S}_{m_1}(z)$ are defined as

$$c_{m_1}(z) = \frac{z}{2 - z\cos(\frac{2\pi}{N} m_1)} \tag{8a}$$

$$\mathcal{S}_{m_1}(z) = \sum_{m_2=0}^{N-1} \frac{e^{i\frac{2\pi}{N} m_2 r_2}}{1 - c_{m_1}(z)\cos(\frac{2\pi}{N} m_2)}. \tag{8b}$$

Before studying the singularity of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ at $z = 1$, we propose to simplify $\mathcal{S}_{m_1}(z)$. The first step is to see from (8a) that $0 < |c_{m_1}(z)| < 1$ for $0 < z < 1$. Using the exponential representation of trigonometric functions, $\mathcal{S}_{m_1}(z)$ can be written as

$$\begin{aligned} \mathcal{S}_{m_1}(z) = -\frac{2}{c_{m_1}(z)} \sum_{m_2=0}^{N-1} \left\{ \frac{e^{i\frac{2\pi}{N} m_2 (1+r_2)}}{e^{i\frac{2\pi}{N} m_2} - \alpha_{m_1}(z)} \right. \\ \left. \times \frac{1}{e^{i\frac{2\pi}{N} m_2} - \alpha_{m_1}^{-1}(z)} \right\} \end{aligned}$$

where $\alpha_{m_1}(z)$ is the smaller root of the equation

$$X^2 - \frac{2}{c_{m_1}(z)} X + 1 = 0$$

whose discriminant is positive for $0 < z < 1$. Thus, we find

$$\alpha_{m_1}(z) = \frac{1 - \sqrt{1 - c_{m_1}^2(z)}}{c_{m_1}(z)}. \tag{9}$$

Now, using partial fraction decomposition, we can write $\mathcal{S}_{m_1}(z)$ as

$$\begin{aligned} \mathcal{S}_{m_1}(z) = \left(1 - c_{m_1}^2(z)\right)^{-\frac{1}{2}} \sum_{m_2=0}^{N-1} \left\{ \frac{e^{i\frac{2\pi}{N} m_2 r_2}}{1 - \alpha_{m_1}(z)\, e^{-i\frac{2\pi}{N} m_2}} \right. \\ \left. + \frac{\alpha_{m_1}(z)\, e^{i\frac{2\pi}{N} m_2 (1+r_2)}}{1 - \alpha_{m_1}(z)\, e^{i\frac{2\pi}{N} m_2}} \right\}. \end{aligned}$$

Noting that $|\alpha_{m_1}(z)| < 1$, it is then possible to expand each sum involved in $\mathcal{S}_{m_1}(z)$ by using successively the expansion $1/(1-x) = \sum_{k=0}^{\infty} x^k$ and the identity

$$\sum_{m=0}^{N-1} e^{i\frac{2\pi}{N} mn} = \begin{cases} N & \text{for} \quad n = 0, \pm N, \pm 2N, \cdots \\ 0 & \text{otherwise} \end{cases}$$

which can be derived by remarking that the vectors $e^{i\frac{2\pi}{N} mn}$ form an orthogonal basis over the set of N-dimensional complex vectors. Therefore, we obtain

$$\mathcal{S}_{m_1}(z) = \frac{N}{\left(1 - c_{m_1}^2(z)\right)^{\frac{1}{2}}} \times \frac{\alpha_{m_1}^{r_2}(z) + \alpha_{m_1}^{N-r_2}(z)}{1 - \alpha_{m_1}^{N}(z)}. \tag{10}$$

By substituting (10) into (7), we find

$$\begin{aligned} P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z) = \frac{1}{N} \sum_{m_1=0}^{N-1} \left\{ \frac{e^{i\frac{2\pi}{N} m_1 r_1}}{\left(1 - \frac{z}{2}\cos(\frac{2\pi}{N} m_1)\right)\left(1 - c_{m_1}^2(z)\right)^{\frac{1}{2}}} \right. \\ \left. \times \frac{\alpha_{m_1}^{r_2}(z) + \alpha_{m_1}^{N-r_2}(z)}{1 - \alpha_{m_1}^{N}(z)} \right\}. \end{aligned} \tag{11}$$

Note that the summand involved in (11) is holomorphic over $0 < z < 1$ for all $0 \le m_1 \le N-1$. However, it diverges at $z = 1$ if and only if $m_1 = 0$. Thus, the singularity of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ at $z = 1$ comes only from the first term of the sum given by (11). It is convenient therefore to separate out the singular and non-singular parts of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ as follows

$$P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z) = \frac{1}{N(1-z)^{\frac{1}{2}}} \times \frac{\alpha_0^{r_2}(z) + \alpha_0^{N-r_2}(z)}{1 - \alpha_0^{N}(z)} + \varphi_N(\vec{\mathbf{r}}, z) \tag{12}$$

where

$$\begin{aligned} \varphi_N(\vec{\mathbf{r}}, z) = \frac{1}{N} \sum_{m_1=1}^{N-1} \left\{ \frac{e^{i\frac{2\pi}{N} m_1 r_1}}{\left(1 - \frac{z}{2}\cos(\frac{2\pi}{N} m_1)\right)\left(1 - c_{m_1}^2(z)\right)^{\frac{1}{2}}} \right. \\ \left. \times \frac{\alpha_{m_1}^{r_2}(z) + \alpha_{m_1}^{N-r_2}(z)}{1 - \alpha_{m_1}^{N}(z)} \right\} \end{aligned} \tag{13}$$

is holomorphic at $z = 1$. The first term involved in (12) corresponds to the term $m_1 = 0$ in (11), and the second term, that is $\varphi_N(\vec{\mathbf{s}}, z)$, corresponds to the sum over the range $1 \leq m_1 \leq N-1$.

To obtain the first-order asymptotic expansion of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ as $z \to 1^-$, let us successively expand the first term of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ involved in Eq. (12) and then function $\varphi_N(\vec{\mathbf{s}}, z)$ close to $z = 1$. Hence, the first term of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ can be expanded as

$$\begin{aligned}
&\frac{1}{N(1-z)^{\frac{1}{2}}} \times \frac{\alpha_0^{r_2}(z) + \alpha_0^{N-r_2}(z)}{1 - \alpha_0^N(z)} = -\frac{1}{N^2(z-1)} \\
&+ \frac{N^2 - 6Nr_2 + 6r_2^2 - 1}{3N^2} + \frac{z-1}{45N^2} \times \Big\{ N^4 - 5N^2(1 + 6r_2^2) \\
&+ 30Nr_2(1 + 2r_2^2) - 30r_2^2(1 + r_2^2) + 4 \Big\} + o(z-1).
\end{aligned} \tag{14}$$

Being holomorphic at $z = 1$, $\varphi_N(\vec{\mathbf{r}}, z)$ can be represented by its first-order Taylor's series expansion at point $z = 1$, so that

$$\varphi_N(\vec{\mathbf{r}}, z) = \varphi_N(\vec{\mathbf{r}}, 1) + \varphi_N^{(1)}(\vec{\mathbf{r}}, 1)(z-1) + o(z-1). \tag{15}$$

Finally, combining (14) with (15), the asymptotic expansion of $P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z)$ as $z \to 1^-$ can be written as

$$\begin{aligned}
&P(\vec{\mathbf{r}}, \mathbf{0} \,|\, z) = \\
&- \frac{1}{N^2(z-1)} + \frac{N^2\big(1 + 3\varphi_N(\vec{\mathbf{r}}, 1)\big) - 6Nr_2 + 6r_2^2 - 1}{3N^2} \\
&+ \frac{z-1}{45N^2} \times \Big\{ N^4 - 5N^2\big(1 + 6r_2^2 - 9\varphi_N^{(1)}(\vec{\mathbf{r}}, 1)\big) \\
&+ 30Nr_2(1 + 2r_2^2) - 30r_2^2(1 + r_2^2) + 4 \Big\} + o(z-1).
\end{aligned} \tag{16}$$

Appendix B
Asymptotic Expansion of $\mathcal{D}_N(\vec{\mathbf{r}} \,|\, z)$

Using (16), the Taylor's series expansion of generating function $\mathcal{D}_N(\vec{\mathbf{r}} \,|\, z)$ as $z \to 1^-$ can be expressed as

$$\begin{aligned}
&\mathcal{D}_N(\vec{\mathbf{r}} \,|\, z) = \\
&1 + \Big\{ N^2\big(\varphi_N(\mathbf{0}, 1) - \varphi_N(\vec{\mathbf{r}}, 1)\big) + 2r_2(N - r_2) \Big\}(z-1) \\
&+ \frac{1}{3}\Big\{ N^4\big(\varphi_N(\mathbf{0}, 1) - \varphi_N(\vec{\mathbf{r}}, 1)\big)\big(1 + 3\varphi_N(\mathbf{0}, 1)\big) \\
&\quad + 2N^3 r_2\big(1 + 3\varphi_N(\mathbf{0}, 1)\big) \\
&\quad + N^2\big(\varphi_N(\vec{\mathbf{r}}, 1) - \varphi_N(\mathbf{0}, 1) - 3\big(\varphi_N^{(1)}(\vec{\mathbf{r}}, 1) - \varphi_N^{(1)}(\mathbf{0}, 1)\big) \\
&\quad - 6r_2^2\varphi_N(\mathbf{0}, 1)\big) - 4Nr_2(1 + r_2^2) + 2r_2^2(2 + r_2^2) \Big\}(z-1)^2 \\
&+ o(z-1)^2.
\end{aligned} \tag{17}$$

References

[1] I. F. Akyildiz, S. Weilian, Y. Sankarasubramaniam, and E. E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, pp. 102–114, 2002.

[2] C. Avin and C. Brito, "Efficient and robust query processing in dynamic environments using random walk techniques," in *IPSN '04: Proceedings of the third international symposium on Information processing in sensor networks*. ACM Press, 2004, pp. 277–286.

[3] C. Avin and B. Krishnamachari, "The power of choice in random walks: an empirical study," in *Proc. of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems*. ACM Press, 2006, pp. 219–228.

[4] C. Avin and G. Ercal, "On the cover time and mixing time of random geometric graphs," *Theor. Comput. Sci.*, vol. 380, no. 1-2, pp. 2–22, 2007.

[5] Z. Bar-Yossef, R. Friedman, and G. Kliot, "RaWMS: random walk based lightweight membership service for wireless ad hoc network," in *MobiHoc '06: Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing*. ACM Press, 2006, pp. 238–249.

[6] N. Bisnik and A. Abouzeid, "Optimizing random walk search algorithms in P2P networks," *Comput. Netw.*, vol. 51, no. 6, pp. 1499–1514, 2007.

[7] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: efficient aggregation for sensor networks," in *IPSN '06: Proceedings of the fifth international conference on Information processing in sensor networks*. ACM Press, 2006, pp. 69–76.

[8] S. Dolev, E. Schiller, and J. Welch, "Random walk for self-stabilizing group communication in ad hoc networks," in *PODC '02: Proceedings of the twenty-first annual symposium on Principles of distributed computing*. ACM Press, 2002, pp. 259–259.

[9] E. J. Duarte-Melo and M. Liu, "Data-gathering wireless sensor networks: organization and capacity," *Computer Networks, Special Issue on Wireless Sensor Networks*, vol. 43, no. 4, pp. 519–537, 2003.

[10] U. Feige and Y. Rabinovich, "Deterministic approximation of the cover time," *Random Struct. Algorithms*, vol. 23, no. 1, pp. 1–22, 2003.

[11] W. Feller, *An introduction to probability theory and its applications*, 2nd ed. New York: Wiley, 1970.

[12] H. Fukś, A. T. Lawniczak, and S. Volkov, "Packet delay in models of data networks," *ACM Trans. Model. Comput. Simul.*, vol. 11, no. 3, pp. 233–250, 2001.

[13] E. Gelenbe, "A diffusion model for packet travel time in a random multihop medium," *ACM Trans. Sen. Netw.*, vol. 3, no. 2, p. 10, 2007.

[14] C. Gkantsidis, M. Mihail, and A. Saberi, "Random walks in peer-to-peer networks," in *Proc. of IEEE INFOCOM*, 2004.

[15] B. D. Hugues, *Random Walks and Random Environments*. New York: Oxford University Press, 1995.

[16] J. Jonasson and O. Schramm, "On the cover time of planar graphs," *ELECTRON.COMM.PROBAB.*, vol. 5, p. 85, 2000.

[17] H. Karl and A. Willig, *Protocols and Architectures for Wireless Sensor Networks*, 2nd ed. John Wiley, 2003, pp. 1–14, 68–71.

[18] S. K. Lando, *Lectures on Generating Functions*. American Mathematical Society, 2003.

[19] L. Lima and J. Barros, "Random walks on sensor networks," in *Proc. of the 5th International Syposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks*, April 2007.

[20] K. Lindenberg, V. Seshadri, K. E. Shuler, and G. H. Weiss, "Lattice random walks for sets of random walkers. first passage times," *Journal of Statistical Physics*, vol. 23, no. 1, pp. 11–25, July 1980.

[21] L. H. Liyanage, C. M. Gulati, and J. M. Hill, "A bibliography on applications of random-walks in theoretical chemistry and physics," *Advances in Molecular Relaxation and Interaction Processes*, vol. 22, pp. 53–72.

[22] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and replication in unstructured peer-to-peer networks," in *ICS '02: Proceedings of the 16th international conference on Supercomputing*. ACM Press, 2002, pp. 84–95.

[23] I. Mabrouki, X. Lagrange, and G. Froc, "Random walk based routing protocol for wireless sensor networks," in *ValueTools '07: Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*. ICST, 2007, pp. 1–10.

[24] E. W. Montroll and G. H. Weiss, "Random walks on lattices. II," *Journal of Mathematical PLIYANAGE1980hysics*, vol. 6, no. 2, pp. 167–181, February 1965.

[25] S. D. Servetto and G. Barrenechea, "Constrained random walks on random graphs: routing algorithms for large scale wireless sensor networks," in *Proc. of the 1st ACM international workshop on Wireless sensor networks and applications*. ACM Press, 2002, pp. 12–21.

[26] F. Spitzer, *Principles of Random Walk*, 2nd ed. New York: Springer, 1976.