

# Random Walk Based Routing Protocol for Wireless Sensor Networks

Issam Mabrouki  
MITSUBISHI ELECTRIC  
ITE-TCL  
Rennes, France  
mabrouki@tcl.ite.mee.com

Xavier Lagrange  
GET/ENST Bretagne  
Rennes, France  
xavier.lagrange@enst-  
bretagne.fr

Gwillerm Froc  
MITSUBISHI ELECTRIC  
ITE-TCL  
Rennes, France  
froc@tcl.ite.mee.com

## ABSTRACT

In recent years, design of wireless sensor networks using methodologies and mechanisms from other disciplines has gained popularity for addressing many networking aspects and providing more flexible and robust algorithms. We address in this paper the problem of random walk to model routing for data gathering in wireless sensor networks. While at first glance, this approach may seem to be overly simplistic and highly inefficient, many encouraging results that prove its comparability with other approaches have been obtained over the years. 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 objective of this paper is to give an analytical model to evaluate the performance of the envisioned routing scheme with special attention to two metrics: the mean system data gathering delay and the induced spatial distribution of energy consumption. The main result shows that this approach achieves acceptable performance for applications without too stringent QoS requirements provided that the ratio of sink nodes over the total number of sensor nodes is carefully tuned.

## Keywords

Sensor networks, Routing, Random Walk, Data Gathering

## 1. INTRODUCTION

Random walks have been studied extensively in a large number of interdisciplinary works, to which mathematicians, physicists, chemists, engineers, and others have toiled and all made significant contributions. More recently, connection with networking research area have been established providing a variety of algorithms including routing, self-stabilization, data gathering and query processing in wireless networks, peer-to-peer networks and other distributed systems [14, 2, 7, 3, 9, 19]. In wireless sensor networks (WSNs) research area, this approach is gaining popularity because random walk techniques present locality, simplicity,

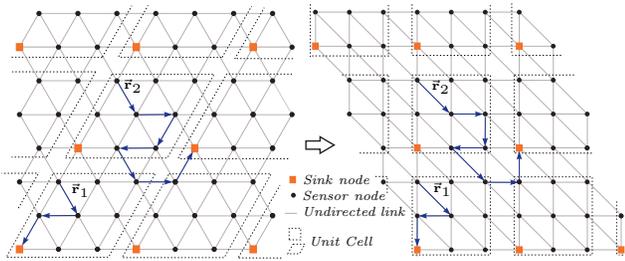
low overhead and inherent robustness to structural changes. Such networks, formed by a large number of small, simple, battery-operated and resource-constrained nodes, are usually subject to dramatic structural changes created by sleep modes, channel fluctuations, mobility, frequent device failures, and other factors. Therefore, *topology-driven* algorithms are at a disadvantage for such networks as they induce high overhead to maintain up-to-date topology and routing information, and also have to provide recovery mechanisms for critical points of failure (*e.g.*, cluster heads, nodes close to the root in a spanning tree). By contrast, *stateless algorithms*, such as the random walk, require no knowledge of network topology to make decisions, and thus, fit better the WSN constraints but often at the expense of QoS support.

Many earlier recent research efforts have raised this vision by focusing primarily on basic properties of random walks. For example, in [14] the authors addressed the problem of data gathering in large-scale WSNs with static sensor nodes and one mobile collector node that performs a random walk on a square lattice. Whenever the collector node enters the transmission range of a sensor node, the data are collected. In this context, the authors derived analytical bounds for the expected number of distinct visited sensor nodes within a given time frame. To improve this performance metric, they proposed a practical algorithm that constrains the random walk and validated it by simulations. Constrained random walk techniques, already suggested in [19] for multipath routing, have the advantage to achieve load balancing property in uncontrolled dynamics characterized by random ON-OFF transitions to save energy. Besides the load balancing property, which is difficult to achieve for other routing protocols, it is also proven in [20] that a random walk based routing in regular patterned WSNs consumes the same amount of energy as the shortest path routing provided that messages are of small size, which characterizes many WSN applications.

Throughout the variety of random walk based algorithms in WSN area, we realize that most of the results are derived from a qualitative view or by means of simulations and that little analytical studies use the powerful techniques of random walk theory. In addition, these results often provide lower and upper bounds on system performance and not a closed form analytical evaluation [14, 3, 19]. Instead, our take in this paper is to obtain a fundamental insight into random walk performance by constructing an analytical model that owes much to the powerful analytic tools developed in the physics community [15, 17, 16].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Inter-Perf '07, October 26, 2007, Nantes, France  
Copyright 2007 ICST 978-963-9799-00-4.



**Figure 1: Topological equivalence between the hexagon lattice and the square lattice with additional diagonal bond inserted in each fundamental square.**

In particular, we address here the problem of data gathering in WSNs comprised of a large number of sensor nodes and a smaller number of sink nodes to gather, process and control data. A packet generated at any sensor node performs a random walk until it reaches a sink node for the first time. At this moment, we consider that the data gathering process occurs with success. Our motivation for considering random walks in this context comes from the following observations. First, the many important papers of Scher and his collaborators [18] for modeling *stochastic transport* in the physics community strongly influenced us to use this approach but from a networking view. Second, to cope with the resource constraint, the multihop strategy is better, since the transmission energy consumption dominates the total power consumption induced by the circuitry of active nodes [5, 6, 8]. Third, 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. Thus, without any state information, sensor nodes would *blindly* forward data. This leads to the Unicast Random Walk based Routing (URWR).

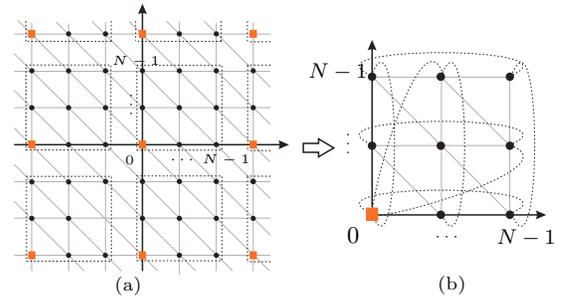
The remainder of this paper is organized as follows. A formal network model description followed by backgrounds and theoretical elements of random walk theory are given in Section II. In section III, which forms our original contribution, we focus on the URWR performance with prior attention to two performance metrics: the mean system data gathering delay and the induced spatial distribution of energy consumption. Finally, we conclude this paper in section IV.

## 2. RANDOM WALK MODEL

### 2.1 Network Description

We consider a graph  $G(\Omega, E)$  where  $\Omega$  is a countable set of nodes wirelessly connected pairwise by a set  $E$  of undirected arcs or edges to represent communication links between nodes. At a given node  $\vec{r}$ , let  $Z(\vec{r}) = \{\vec{s}_1, \dots, \vec{s}_l\}$  be the set of neighbors of  $\vec{r}$ . Let  $\pi_{\vec{r}} = \{p_1, \dots, p_l\}$  be nonnegative reals such that  $\sum_i p_i = 1$ . This defines a transition probability distribution over the neighbors of  $\vec{r}$ . When a packet reaches node  $\vec{r}$ , the next hop is chosen by tossing a die whose  $i$ -th face occurs with probability  $p_i$ , and the packet is then forwarded over the link  $(\vec{r}, \vec{s}_i)$ . The random sequence of nodes selected this way is a *random walk* on graph  $G(\Omega, E)$ .

By making different assumptions on the topology of the underlying network and on constraints imposed on  $\pi_{\vec{r}}$ , we are



**Figure 2: Equivalence between the infinite periodic square lattice (a) and the torus lattice (b). For the sake of clarity, we have deliberately missed some transitions except for the origin.**

able to explore a large space of possible routing schemes. In particular, we consider in this paper a regular deployment topology where nodes are spread over an area of interest with a hexagon lattice pattern (or equivalently an equilateral triangular lattice). This pattern is of fourfold interest. 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. Second, under full sensing coverage requirement, the hexagon lattice has the lowest node density [12], which is important for economic reasons. Third, sensor placement can be viewed as spatial sampling from signal processing perspective. The literature there also suggests the superiority of hexagonal sampling lattice over the square lattice when the spatial spectrum of 2-D signal being measured (such as a temperature field) is bandlimited with a circular support. Finally, with a communication range  $r_c \geq \sqrt{3}r_s$ , where  $r_s$  stands for the sensing range, the hexagon lattice pattern provides both coverage and connectivity [4].

In this paper, we consider the case where  $r_c = \sqrt{3}r_s$  which corresponds to a 6-connectivity. Moreover, it is assumed that there is unbiased random motion, so that all of the 6 possible next steps away from the current position are equally likely, each having probability  $\frac{1}{6}$ . There exist two kinds of network nodes: sensor nodes and a smaller number of sink nodes distributed in a regular way, one sink node for every  $N^2$  vertices without any specific mapping between sensor and sink nodes. A packet generated at a given sensor node performs then a random walk until it reaches a sink node for the *first* time. At this moment, we consider that the data gathering process occurs with success.

As illustrated in Figure 1, the induced network has a periodic structure, in which each unit cell of size  $N \times N$  and in the form of a rhombus contains  $N^2 - 1$  sensor nodes and 1 sink node. Since we are interested only in which nodes are connected to which, not in the visual appearance of the unit cell, we may transform this unit cell to obtain a new cell in the form of a square lattice of size  $N \times N$  with one additional diagonal bond inserted in each fundamental square. The obtained cell has the advantage that we can assign to nodes integer coordinates while maintaining the original network topology. With a large number of unit cells and recalling the assumption that there is no specific mapping between sensor nodes and sink nodes, we can assume that the formed network is infinite and hence, it is equivalent to a torus lattice  $T$  formed by connecting the opposite ends of the square unit

cell (Figure 2). Therefore, in the remainder of this paper we investigate the URWR in reference to torus lattice  $T$ . Every node  $\vec{s} \in T$  is labeled with  $(s_1, s_2)$ , where  $s_1$  and  $s_2$  are integers such that  $0 \leq s_1, s_2 \leq N - 1$ . The obtained results will be valid for the original network. However, before investigating the performance of the URWR, we introduce in the following backgrounds and theoretical elements of random walk theory that will be used to evaluate the URWR. Deeper discussions of random walk results can be found in [11].

## 2.2 Elements of Random Theory

### 2.2.1 Node Occupation and First-Passage Probabilities

We define, for  $n \geq 1$ ,  $P_n(\vec{r}, \vec{s})$  the probability of being at node  $\vec{s}$  after  $n$  hops, given that the packet has been issued at node  $\vec{r}$ . We also define  $F_n(\vec{r}, \vec{s})$  the probability of arriving at node  $\vec{s}$  for the *first* time on the  $n$ th hop, given that the walk started at node  $\vec{r}$ . We shall refer to these probabilities as the *node occupation probability* and the *first-passage probability* respectively. By convention we have  $P_0(\vec{r}, \vec{s}) = \delta_{\vec{r}\vec{s}}$  and  $F_0(\vec{r}, \vec{s}) = 0$ .

We make use in this paper of the *generating function formalism* [13] 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  $\{c_n\}_{n \in \mathbb{N}}$ . The interest of this formalism arises especially in probability theory to completely characterize a discrete random variable by associating a generating function to its probability distribution. Once the generating function is determined, a lot of information about the random variable, such as the expectation or the variance, can be then derived.

In the following, we denote the generating functions associated with  $\{P_n(\vec{r}, \vec{s})\}_{n \in \mathbb{N}}$  and  $\{F_n(\vec{r}, \vec{s})\}_{n \in \mathbb{N}}$  as  $P(\vec{r}, \vec{s}|z)$  and  $F(\vec{r}, \vec{s}|z)$  respectively. Hereafter we present two well-known classical relations extensively used in random walk theory, and upon which the theoretical results derived in this paper rely.

LEMMA 1. *At fixed  $\vec{r}$ ,  $P(\vec{r}, \vec{s}|z)$  satisfies the normalization condition*

$$\sum_{\vec{s} \in T} P(\vec{r}, \vec{s}|z) = \frac{1}{1-z}. \quad (1)$$

PROOF. As long as there is no possibility that the packet is removed from  $T$  at any time  $n$  (this is so for the walks considered here),  $P_n(\vec{r}, \vec{s})$  is a distribution over the nodes of  $T$  at fixed  $n$  and  $\vec{r}$ , which leads to  $\sum_{\vec{s} \in T} P_n(\vec{r}, \vec{s}) = 1$ , for all  $n \geq 0$ . By taking the generating functions of both sides of this relation, (1) follows.  $\square$

LEMMA 2.  *$F(\vec{r}, \vec{s}|z)$  and  $P(\vec{r}, \vec{s}|z)$  are related to each other according to the relation*

$$F(\vec{r}, \vec{s}|z) = \frac{P(\vec{r}, \vec{s}|z) - \delta_{\vec{r}\vec{s}}}{P(\vec{s}, \vec{s}|z)}, \quad \vec{r}, \vec{s} \in T. \quad (2)$$

PROOF. The proof is presented in Appendix A.  $\square$

### 2.2.2 Random Walk Characterization

A key issue in the problem of random walk 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  $P(\vec{r}, \vec{s}|z)$ , 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 lattice with periodic boundary conditions (*i.e.*, toroidal lattices) has been extensively studied. The investigation of this type of system was motivated by many situations and process encountered in statistical mechanics, solid state theory (*e.g.*, diffusion of electrons, excitons, energy transfer, conductivity, dislocations). Montroll and Weiss [17] originally proposed this special problem and solved it for  $k$ -dimension. In particular, in case of square lattice with diagonal bond, they established an explicit expression of  $P(\mathbf{0}, \vec{s}|z)$ , that is

$$P(\mathbf{0}, \vec{s}|z) = \frac{1}{N^2} \sum_{\vec{m} \in T} \frac{e^{-i\frac{2\pi}{N}\vec{m} \cdot \vec{s}}}{1 - z\lambda(\frac{2\pi}{N}\vec{m})} \quad (3)$$

where function  $\lambda(\vec{\theta})$  is defined on  $\mathbb{R}^2$  as

$$\lambda(\vec{\theta}) = \frac{1}{3} (\cos(\theta_1) + \cos(\theta_2) + \cos(\theta_1 - \theta_2)).$$

Even though relation (3) does not give a simple form of  $P(\mathbf{0}, \vec{s}|z)$ , it is of great importance and represents our basic relation upon which relies most of our contribution in this paper.

## 3. PERFORMANCE ANALYSIS

In the present section, we investigate the performance of the URWR. We shall focus our attention primarily on two performance metrics: (i) how long does it take for the data gathering process and (ii) how much energy does it require? These two metrics, which characterize the global system behavior, are analytically studied by using the general formalism introduced in subsection 2.2. We first focus on (i) the system data gathering delay and then (ii) the spatial distribution of the energy consumption.

### 3.1 System Data Gathering Delay

We define the *data gathering delay* of a packet at a given sensor node  $\vec{s}$  as the time or the number of hops it takes to reach the sink node for the first time after it leaves sensor node  $\vec{s}$ . The *system data gathering delay* is then obtained by averaging the packet delay over all sensor nodes contained in  $T$ . This time is a random variable denoted by  $D_N$ . We consider here a uniform traffic distribution, that is, a packet has the probability  $\frac{1}{N^2-1}$  of being generated at any sensor node in  $T$ . For studying the *mean system data gathering delay*, we determine first the generating function associated with the probability distribution of random variable  $D_N$ .

#### 3.1.1 Generating Function Analysis of $D_N$

Suppose that at time  $n = 0$  a packet has the probability  $\frac{1}{N^2-1}$  of being generated at any sensor node in  $T$ . If  $G_n(N)$  denotes the probability that the packet will reach the sink node for the *first* time on the  $n$ th hop, then in terms of probability notation we have

$$\Pr\{D_N = n\} = G_n(N), \quad n \geq 0.$$

Using now the law of total probability, it is possible to decompose the event that a packet generated anywhere will reach the sink node for the *first* time on the  $n$ th hop, which has the probability  $G_n(N)$ , into the  $N^2 - 1$  mutually exclusive events that the packet is *initially* released at sensor node  $\vec{s}$  with probability  $\frac{1}{N^2-1}$ , and then arrives at the sink node for the *first* time after  $n$  hops, which has the probability  $F_n(\vec{s}, \mathbf{0})$ . Thus,  $G_n(N)$  can be expressed as

$$G_n(N) = \frac{1}{N^2 - 1} \sum_{\vec{s} \neq \mathbf{0}} F_n(\vec{s}, \mathbf{0}), \quad n \geq 0.$$

Multiplying both sides of this relation by  $z^n$  and summing over all  $n$  yield

$$G(N|z) = \frac{1}{N^2 - 1} \sum_{\vec{s} \neq \mathbf{0}} F(\vec{s}, \mathbf{0} | z),$$

where  $G(N|z)$  stands for the generating function associated with sequence  $\{G_n(N)\}_{n \in \mathbb{N}}$ . Then, from (2) we obtain

$$\begin{aligned} G(N|z) &= \frac{1}{N^2 - 1} \sum_{\vec{s} \neq \mathbf{0}} \frac{P(\vec{s}, \mathbf{0} | z)}{P(\mathbf{0}, \mathbf{0} | z)} \\ &= \frac{1}{(N^2 - 1)P(\mathbf{0}, \mathbf{0} | z)} \left\{ \sum_{\vec{s} \in T} P(\vec{s}, \mathbf{0} | z) - P(\mathbf{0}, \mathbf{0} | z) \right\}. \end{aligned}$$

However, since it is assumed that there is no directional bias on any hop, the considered random walk is symmetric and therefore we have  $P(\vec{s}, \mathbf{0} | z) = P(\mathbf{0}, \vec{s} | z)$ . Further, using the normalization condition stated in (1), we find

$$G(N|z) = \frac{1}{N^2 - 1} \left\{ \frac{1}{(1 - z)P(\mathbf{0}, \mathbf{0} | z)} - 1 \right\}. \quad (4)$$

Various statistics of the system data gathering delay can be now extracted from this key relation using the general formalism of generating function. In particular, we propose next to derive the mean system data gathering delay, *i.e.*, the expectation of  $D_N$  denoted by  $\mathbf{E}(D_N)$ .

### 3.1.2 Mean System Data Gathering Delay

The mean system data gathering delay can be written as

$$\begin{aligned} \mathbf{E}(D_N) &= \sum_{n=0}^{\infty} n G_n(N) \\ &= \lim_{z \rightarrow 1^-} \frac{\partial}{\partial z} G(N|z). \end{aligned}$$

Although this formula does not give an explicit expression of the mean system data gathering delay, we can fortunately extract a closed-form by a *Taylor's* series expansion of  $G(N|z)$  at point  $z = 1$ . Indeed, the limit of the first derivative of  $G(N|z)$  as  $z \rightarrow 1^-$  is nothing but the first-order term of its *Taylor's* series. To obtain the first order *Taylor's* series expansion of  $G(N|z)$  at point  $z = 1$ , we proceed as follows. We first calculate the asymptotic expansion of  $P(\mathbf{0}, \mathbf{0} | z)$  as  $z \rightarrow 1$ , which is obtained by setting  $\vec{s} = \mathbf{0}$  in the asymptotic expansion of  $P(\mathbf{0}, \vec{s} | z)$  as  $z \rightarrow 1$  given by (19) in Appendix B. Second, by plugging the asymptotic expansion of  $P(\mathbf{0}, \mathbf{0} | z)$  as  $z \rightarrow 1$  into (4), we obtain

$$G(N|z) = 1 + \frac{N^4(1 + 4\varphi_N(\mathbf{0}, 1)) - N^2}{4(N^2 - 1)} (z - 1) + o(z - 1)$$

where  $\varphi_N(\mathbf{0}, 1)$  is an  $N$ -dependent series obtained by setting  $\vec{s} = \mathbf{0}$  and  $z = 1$  in  $\varphi_N(\vec{s}, z)$  defined by (18) in Appendix B.

By differentiating this *Taylor's* series expansion with respect to  $z$  and then taking the limit as  $z \rightarrow 1^-$ , we obtain the following closed-form expression of the mean system data gathering delay

$$\mathbf{E}(D_N) = \frac{1}{4}N^2 + \frac{N^4}{N^2 - 1} \varphi_N(\mathbf{0}, 1). \quad (5)$$

Some general remarks can be drawn from the previous results. First, from the first-order *Taylor's* series expansion, the value of  $G(N|z)$  at point  $z = 1$  is equal to unit, which represents the probability that the sink node is *ever* reached by a packet generated anywhere in the network. This means that the data gathering process ensured by the URWR is certain. Second, referring to (5), the mean system data gathering delay is finite and depends only on the size of a unit cell via parameter  $N$ . To study this dependence, one has to estimate series  $\varphi_N(\mathbf{0}, 1)$ . Two ways are possible. First, the mean system data gathering delay is estimated by a numerical calculation of series  $\varphi_N(\mathbf{0}, 1)$ . A second way consists in approximating  $\mathbf{E}(D_N)$  by an asymptotic expansion of series  $\varphi_N(\mathbf{0}, 1)$  as  $N \rightarrow \infty$ , which is provided by (21) in Appendix C. Plugging (21) into (5), we obtain our first analytical result

RESULT 1.

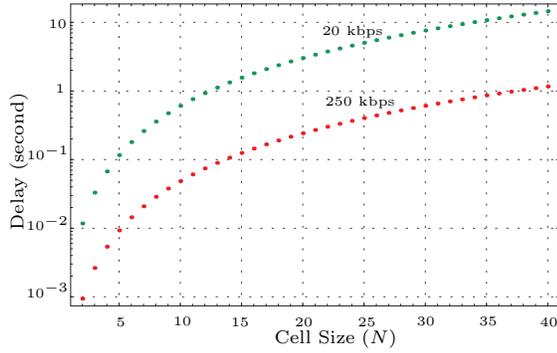
$$\begin{aligned} \mathbf{E}(D_N) &= \frac{\sqrt{3}}{\pi} N^2 \ln(N) + N^2 \left( \frac{\sqrt{3}}{\pi} \gamma + \frac{\sqrt{3}}{\pi} \ln\left(\frac{\sqrt{3}}{\pi}\right) + \ell + \frac{1}{4} \right) \\ &\quad + \frac{\sqrt{3}}{\pi} \ln(N) + \frac{\sqrt{3}}{\pi} \left( \gamma + \ln\left(\frac{\sqrt{3}}{\pi}\right) + \frac{\sqrt{3}}{3} \pi \ell \right) \\ &\quad + \frac{\sqrt{3}}{\pi} \frac{\ln(N)}{N^2} + O\left(\frac{1}{N^2}\right). \end{aligned} \quad (6)$$

where “ $O$ ” stands for the “*Big-O*” Landau symbol,  $\gamma$  is the Euler's constant ( $\gamma = 0.5772156649$ ) and  $\ell = -0.0047473394$ .

### 3.1.3 Discussion

As mentioned earlier, the main question we want to ask in this paper is to what extent the URWR can be efficient while being simple and light. In this section, we have focused on the mean system data gathering delay as a performance metric of the URWR. As shown by *Result 1*, this performance metric depends only on cell size or equivalently on the number of sensor nodes to be spread *per* sink node. This result is of threefold interest. First, the dependence of the achievable mean system data gathering delay on cell size alone and not on the overall network size shows that the URWR can scale up with the network size. Second, it is not necessary that  $N$  be very large for approximation formula (6) to be numerically useful. For  $N = 3$ , (6) gives  $\mathbf{E}(D_3) = 8.226$  hops, which is only about 0.29 per cent different from the exact result from (5) that  $\mathbf{E}(D_3) = 8.25$  hops. Third, it is always possible to guarantee an acceptable mean system data gathering delay provided that the number of sensor node per sink node is carefully tuned. Thus, approximation formula (6) can be put into practical use for WSN dimensioning with respect to crucial parameters such as the minimum ratio of sink nodes to be deployed over the total number of sensor nodes while ensuring a required threshold mean system data gathering delay.

Concretely, consider now IEEE 802.15.4 enabled network nodes with maximum and minimum supported data rates of 250 kbps and 20 kbps respectively. Since in many applications of WSNs, sensor nodes often send only beep-like



**Figure 3: Mean system data gathering delay  $\mathbf{E}(D_N)$  vs. cell size  $N$  for IEEE 802.15.4 enabled network nodes, and with an average packet size of 10 bytes.**

small packets to the sink node to report their status, we can assume here an average packet size of about 10 bytes. Therefore, the minimum achievable mean system data gathering delay for the maximum data rate is about  $938 \mu\text{s}$  while for the minimum data rate, it is about 11.7 ms. This corresponds to 2.93 hops for a unit cell of size  $N^2 = 4$  nodes. However, as illustrated in Figure 3, to guarantee a mean system data gathering delay threshold of 1 s, the unit cell size should not exceed  $N^2 = 37^2$  nodes for the maximum data rate, and  $N^2 = 12^2$  for the minimum data rate. As these values approach upper bounds of practical interest, we can conclude that for applications without too stringent required delay, which often characterize a large range of WSN applications, the URWR can perform acceptable performance in terms of delay provided that the number of sensor nodes to be deployed *per* sink node is well managed.

## 3.2 Energy Consumption

Besides the mean system data gathering delay, another important performance metric of the URWR is the energy consumption: how much energy does it require for a packet to be gathered by a sink node? We do not take here into account all possible sources of energy consumption, but only the contribution of the URWR induced by successive transmissions/receptions at relay sensor nodes. A straightforward metric that measures the successive transmissions/receptions cost is the number of visits of a packet to a relay sensor node *before* reaching the sink node. At a given relay sensor node  $\vec{s}$ , we denote by  $V_N(\vec{s})$  the number of visits during the random walk before visiting the sink node, and we assume again a uniform traffic distribution. In the following, we first put this physical picture into a mathematical form and then evaluate at relay sensor node  $\vec{s}$  the *mean number of visits* denoted by  $\mathbf{E}(V_N(\vec{s}))$ .

### 3.2.1 Sink-Avoiding Node Occupation Probability

Let us define  $P_n^\dagger(\vec{r}, \vec{s})$  as the probability of being at relay sensor node  $\vec{s}$  on the  $n$ th hop before reaching the sink node, given that the packet has been initially issued at sensor node  $\vec{r}$ . We refer to this probability as the *sink-avoiding node occupation probability*. We show next how  $P_n^\dagger(\vec{r}, \vec{s})$  and  $P_n(\vec{r}, \vec{s})$  are related to each other. Indeed, using the law of total probability, it is possible to decompose the event that a packet generated at sensor node  $\vec{r}$  will be at relay sensor node  $\vec{s}$  on the  $n$ th hop, which has the probability  $P_n(\vec{r}, \vec{s})$ ,

into the two exclusive events: (i) the packet visits sensor node  $\vec{s}$  on the  $n$ th hop before visiting the sink node, which has the probability  $P_n^\dagger(\vec{r}, \vec{s})$ , (ii) the packet *first* arrived at the sink node after  $j$  hops and subsequently performed a walk of  $n - j$  hops arriving at sensor node  $\vec{s}$ , which has the probability  $F_j(\vec{r}, \mathbf{0})P_{n-j}(\mathbf{0}, \vec{s})$ . Thus,

$$P_n(\vec{r}, \vec{s}) = P_n^\dagger(\vec{r}, \vec{s}) + \sum_{j=0}^{n-1} F_j(\vec{r}, \mathbf{0})P_{n-j}(\mathbf{0}, \vec{s}), \quad n \geq 0.$$

Multiplying both sides by  $z^n$ , summing over all  $n$  and observing that the convolution of two sequences corresponds to the multiplication of their generating functions, we obtain

$$P(\vec{r}, \vec{s} | z) = P^\dagger(\vec{r}, \vec{s} | z) + F(\vec{r}, \mathbf{0} | z)P(\mathbf{0}, \vec{s} | z) \quad (7)$$

where  $P^\dagger(\vec{r}, \vec{s} | z)$  stands for the generating function associated with  $\{P_n^\dagger(\vec{r}, \vec{s})\}_{n \in \mathbb{N}}$ . Then, substituting the expression of  $F(\vec{r}, \mathbf{0} | z)$  given by (2) into (7) yields this key relation

$$P^\dagger(\vec{r}, \vec{s} | z) = P(\vec{r}, \vec{s} | z) - P(\mathbf{0}, \vec{s} | z) \frac{P(\vec{r}, \mathbf{0} | z)}{P(\mathbf{0}, \mathbf{0} | z)}. \quad (8)$$

Next, we evaluate the mean number of visits  $\mathbf{E}(V_N(\vec{s}))$  based on this relation.

### 3.2.2 Mean Number of Visits

Let us suppose that at time  $n = 0$  a packet has the probability  $\frac{1}{N^2 - 1}$  of being generated at any sensor node in  $T$ . At a given relay sensor node  $\vec{s}$ , we define an indicator random variable  $I_n(\vec{s})$ , which takes the value 1 if relay sensor node  $\vec{s}$  is visited by the packet on the  $n$ th hop before reaching the sink node, and is zero otherwise. Let us also define  $H_n(\vec{s})$  and  $H(\vec{s} | z)$  the probability that relay sensor node  $\vec{s}$  is visited on the  $n$ th hop by the packet before reaching the sink node and its associated generating function respectively, so that, in terms of probability notation we have

$$\Pr\{I_n(\vec{s}) = 1\} = H_n(\vec{s}).$$

Therefore, it can be deduced that the number of visits of the packet to relay sensor node  $\vec{s}$  before reaching the sink node during the walk is simply

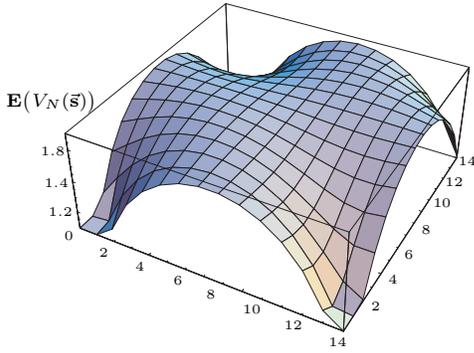
$$V_N(\vec{s}) = \sum_{n=0}^{\infty} I_n(\vec{s}),$$

and hence,

$$\begin{aligned} \mathbf{E}(V_N(\vec{s})) &= \mathbf{E}\left(\sum_{n=0}^{\infty} I_n(\vec{s})\right) = \sum_{n=0}^{\infty} \mathbf{E}(I_n(\vec{s})) \\ &= \sum_{n=0}^{\infty} H_n(\vec{s}) = \lim_{z \rightarrow 1^-} H(\vec{s} | z). \end{aligned}$$

Now, it remains to make explicit  $H(\vec{s} | z)$ . Using the law of total probability, it is possible to decompose the event that a packet generated at any sensor node will visit relay sensor node  $\vec{s}$  on the  $n$ th hop before visiting the sink node, which has the probability  $H_n(\vec{s})$ , into the  $N^2 - 1$  mutually exclusive events that the packet is initially generated at sensor node  $\vec{r}$  with probability  $\frac{1}{N^2 - 1}$ , and then visits relay sensor node  $\vec{s}$  after  $n$  hops before visiting the sink node, which occurs with probability  $P_n^\dagger(\vec{r}, \vec{s})$ . Thus, we obtain

$$H_n(\vec{s}) = \frac{1}{N^2 - 1} \sum_{\vec{r} \neq \mathbf{0}} P_n^\dagger(\vec{r}, \vec{s}), \quad n \geq 0.$$



**Figure 4: Spatial distribution of the energy consumption based on the mean number of visits over a unit cell of size  $15 \times 15$ .**

By multiplying both sides by  $z^n$  and summing over all  $n$ , we obtain

$$H(\vec{s}|z) = \frac{1}{N^2 - 1} \sum_{\vec{r} \neq \mathbf{0}} P^\dagger(\vec{r}, \vec{s}|z). \quad (9)$$

Finally, by plugging (8) into (9), using successively the symmetric property of  $P(\vec{r}, \mathbf{0}|z)$  and the normalization condition stated in (1), we find

$$H(\vec{s}|z) = \frac{1}{(N^2 - 1)(1 - z)} \left\{ 1 - \frac{P(\mathbf{0}, \vec{s}|z)}{P(\mathbf{0}, \mathbf{0}|z)} \right\}. \quad (10)$$

Since the mean number of visits is defined as the limit of  $H(\vec{s}|z)$  as  $z \rightarrow 1^-$ , a zero-order *Taylor's* series expansion of  $H(\vec{s}|z)$  at point  $z = 1$  is sufficient to derive  $\mathbf{E}(V_N(\vec{s}))$ . Likewise the calculation of the mean system data gathering delay, we proceed first to evaluate the asymptotic expansion of  $\frac{P(\mathbf{0}, \vec{s}|z)}{P(\mathbf{0}, \mathbf{0}|z)}$  as  $z \rightarrow 1^-$ , which can be derived from (19) in Appendix B. Second, after elementary calculus, we obtain the limit of  $H(\vec{s}|z)$  as  $z \rightarrow 1^-$ , *i.e.*,  $\mathbf{E}(V_N(\vec{s}))$  as follows

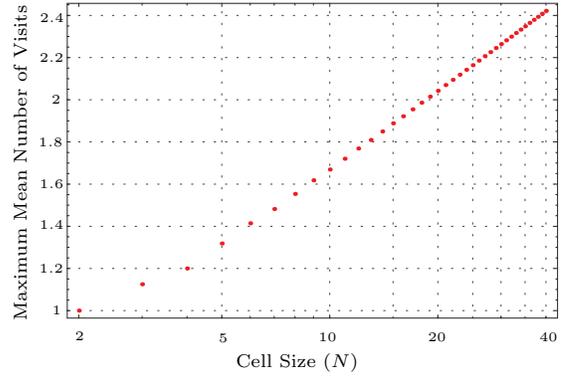
RESULT 2.

$$\mathbf{E}(V_N(\vec{s})) = \frac{N^2}{N^2 - 1} (\varphi_N(\mathbf{0}, 1) - \varphi_N(\vec{s}, 1)) + \frac{3s_2(N - s_2)}{2(N^2 - 1)}. \quad (11)$$

where  $\varphi_N(\vec{s}, 1)$  is a series obtained by setting  $z = 1$  in  $\varphi_N(\vec{s}, z)$  defined by (18) in Appendix B.

### 3.2.3 Discussion

Some general remarks can be drawn from *Result 2*. First, note that the mean number of visits depends on both cell size via parameter  $N$ , and the position of relay sensor node  $\vec{s}$ . This is predictable since the nodes of  $T$  are not equivalent: sensor nodes in the vicinity of the sink node are not equivalent to those in the middle of the cell. Second, as shown in Figure 4, the mean number of visits achieved far away from the sink node is much closer to uniform and with higher values than the one in the vicinity of the sink node. Then we can conclude that the URWR achieves a load balancing property. Intuitively, this can be explained by the fact that a packet visiting a relay sensor node located in the vicinity of the sink node is more likely to reach it during the next fewer hops than when visiting a relay sensor node far away from the sink node, and thus, it is more likely to return once again to the latter relay sensor node. In other words, the



**Figure 5: Maximum mean number of visits vs. cell size  $N$ .**

proximity to the sink node decreases the likelihood of the return to the initial position. In this way, relay sensor nodes in the middle of the cell are expected to be more visited than the others. This is consistent with Figure 4. Interestingly enough, and at first glance surprising to us –although rather obviously with the benefit of hindsight– is the fact that the energy consumption distribution achieved along the diagonal between the upper left corner and the lower right corner is lower than the one along the opposite diagonal. This can be explained by the fact that only hops along the former diagonal are allowed and hence, accelerating random walks along this direction.

Although there is no unified definition of *network lifetime*, as this concept depends on the objective of an application, we adopt here a common definition which characterizes the network lifetime as the time until the first node in the network depletes its energy. In case of the URWR, the first node that depletes its energy corresponds to the most visited node during the data gathering process. As illustrated in Figure 5, the maximum mean number of visits grows logarithmically with large values of  $N$ . Therefore, the network lifetime does not degrade quickly when the number of sensor nodes per sink node increases.

Let us now consider a practical case where nodes are equipped with the MC13192, an IEEE 802.15.4 compliant radio transceiver from *Freescale Semiconductor* [1], which offers a data rate of 250 kbps and operates in transmit/receive mode at (30 mA, 2.7 V) and (37 mA, 2.7 V) respectively. Since a visit to a sensor node counts for one reception followed by one transmission, the energy required for a 10 byte message per visit is about  $57.88 \mu\text{J}$ . Therefore, for a given energy source and a frequency for measure updates, it is possible to derive the network lifetime for a given cell size based on Figure 5.

## 4. CONCLUSION

In this paper we addressed the problem of random walk to model data gathering in large-scale WSNs with regular structure. Our motivation for this approach comes from the inherent properties of random walks such as the locality, the simplicity and the robustness to structural changes. We presented a mathematical formulation of the problem of random walk that largely owes to the powerful theoretical tools developed in the physics community. Using the generating function formalism, we established closed-form expressions

for the mean system data gathering delay and the energy consumption based on the mean number of visits.

Our main results could be summarized as follows. First, the dependence of the achievable mean system data gathering delay on only the ratio of sink nodes over the total number of sensor nodes shows that the random walk can scale with the network size. Second, it is always possible to guarantee an acceptable mean system data gathering delay for delay-tolerant applications provided that this ratio is carefully tuned and that the packets are of small size. Third, as regards to energy consumption, the random walk achieves a load balancing property.

There are several interesting directions that extend the model presented in this paper. This consists for instance in deriving other significant statistics of the random walk such as the dispersion of the system data gathering delay, studying *biased* random walks to improve performance and considering other network topologies such as the square lattice, which will be soon published.

## APPENDIX

### A. PROOF OF LEMMA 2

The event “the packet is at node  $\vec{s}$  after  $n$  hops, given that the walk started at node  $\vec{r}$ ”, which has the probability  $P_n(\vec{r}, \vec{s})$ , can be decomposed into  $n$  mutually exclusive events “the packet *first* arrived at node  $\vec{s}$  after  $j$  hops and subsequently performed a walk of  $n - j$  hops returning to node  $\vec{s}$ ”, with probabilities  $F_j(\vec{r}, \vec{s})P_{n-j}(\vec{s}, \vec{s})$ . Thus

$$P_n(\vec{r}, \vec{s}) = \sum_{j=1}^n F_j(\vec{r}, \vec{s})P_{n-j}(\vec{s}, \vec{s}), \quad n \geq 1.$$

If this sum is interpreted as void when  $n = 0$ , recalling the convention that  $P_0(\vec{r}, \vec{s}) = \delta_{\vec{r}\vec{s}}$  and  $F_0(\vec{r}, \vec{s}) = 0$ , we may rewrite this relation as

$$P_n(\vec{r}, \vec{s}) = \delta_{\vec{r}\vec{s}} \delta_{0n} + \sum_{j=0}^n F_j(\vec{r}, \vec{s})P_{n-j}(\vec{s}, \vec{s}), \quad n \geq 0.$$

Taking the generating functions of both sides yields (2).

### B. ASYMPTOTIC ANALYSIS OF $P(0, \vec{s} | Z)$

#### B.1 Singularity of $P(0, \vec{s} | z)$ at $z = 1$

We propose here to simplify the expression of  $P(0, \vec{s} | z)$  given by (3) and to study its singularity at point  $z = 1$ . By factorizing the denominator of the summand and using the addition theorems of trigonometric functions, we obtain

$$\begin{aligned} P(0, \vec{s} | z) &= \frac{1}{N^2} \sum_{m_1=0}^{N-1} \sum_{m_2=0}^{N-1} \left\{ \frac{e^{-i\frac{2\pi}{N}m_1s_1}}{1 - \frac{z}{3} \cos(\frac{2\pi}{N}m_1)} \right. \\ &\quad \left. \times \frac{e^{-i\frac{2\pi}{N}m_2s_2}}{1 - c_{m_1}(z) \cos(\frac{\pi}{N}(m_1 - 2m_2))} \right\} \\ &= \frac{1}{N^2} \sum_{m_1=0}^{N-1} \frac{e^{-i\frac{2\pi}{N}m_1s_1}}{1 - \frac{z}{3} \cos(\frac{2\pi}{N}m_1)} \times S_{m_1}(z) \end{aligned} \quad (12)$$

where functions  $c_{m_1}(z)$  and  $S_{m_1}(z)$  are defined as

$$c_{m_1}(z) = \frac{2z \cos(\frac{\pi}{N}m_1)}{3 - z \cos(\frac{2\pi}{N}m_1)} \quad (13a)$$

$$S_{m_1}(z) = \sum_{m_2=0}^{N-1} \frac{e^{-i\frac{2\pi}{N}m_2s_2}}{1 - c_{m_1}(z) \cos(\frac{\pi}{N}(m_1 - 2m_2))} \quad (13b)$$

Next, we propose to simplify  $S_{m_1}(z)$ . The first step is to see from (13a) that  $0 < |c_{m_1}(z)| < 1$  for  $0 < z < 1$  unless  $m_1 = \frac{N}{2}$ , which occurs only when  $N$  is even. In this case,  $c_{\frac{N}{2}}(z) = 0$  and we find  $S_{\frac{N}{2}}(z) = N\delta_{0, s_2}$  where the delta symbol stands for the *Kronecker* symbol. Let us now assume that  $m_1 \neq \frac{N}{2}$ , so that, by using the exponential representation of trigonometric functions,  $S_{m_1}(z)$  can be written as follows

$$S_{m_1}(z) = -\frac{2}{c_{m_1}(z)} \sum_{m_2=0}^{N-1} \left\{ \frac{e^{i\frac{\pi}{N}(m_1 - 2m_2(1+s_2))}}{e^{i\frac{\pi}{N}(m_1 - 2m_2)} - \alpha_{m_1}(z)} \right. \\ \left. \times \frac{1}{e^{i\frac{\pi}{N}(m_1 - 2m_2)} - \alpha_{m_1}^{-1}(z)} \right\}$$

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 nonnegative. Thus, we find

$$\alpha_{m_1}(z) = \frac{1 - \sqrt{1 - c_{m_1}^2(z)}}{c_{m_1}(z)}. \quad (14)$$

Using partial fraction decomposition,  $S_{m_1}(z)$  becomes

$$S_{m_1}(z) = (1 - c_{m_1}^2(z))^{-\frac{1}{2}} \sum_{m_2=0}^{N-1} \left\{ \frac{e^{-i\frac{2\pi}{N}m_2s_2}}{1 - \alpha_{m_1}(z) e^{-i\frac{\pi}{N}(m_1 - 2m_2)}} \right. \\ \left. + \frac{\alpha_{m_1}(z) e^{i\frac{\pi}{N}(m_1 - 2m_2(1+s_2))}}{1 - \alpha_{m_1}(z) e^{i\frac{\pi}{N}(m_1 - 2m_2)}} \right\}.$$

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

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

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

$$S_{m_1}(z) = \frac{N e^{-i\frac{\pi}{N}m_1s_2}}{(1 - c_{m_1}^2(z))^{\frac{1}{2}}} \times \frac{\alpha_{m_1}^{s_2}(z) + \alpha_{m_1}^{N-s_2}(z) \cos(\pi m_1)}{1 - \alpha_{m_1}^N(z) \cos(\pi m_1)} \quad (15)$$

which is established for  $m_1 \neq \frac{N}{2}$  and  $0 < z < 1$ . However, when  $c_{m_1}(z) \rightarrow 0$ , which occurs at  $m_1 \rightarrow \frac{N}{2}$ ,  $\alpha_{m_1}(z)$  goes to zero. Thus, if we extend the expression (15) to include  $m_1 = \frac{N}{2}$ , we obtain  $S_{\frac{N}{2}}(z) = N\delta_{0, s_2}$ , which is consistent with (13b). Therefore (15) holds for all  $0 \leq m_1 \leq N - 1$ .

Finally, by substituting (15) into (12), we obtain

$$P(\mathbf{0}, \vec{s} | z) = \frac{1}{N} \sum_{m_1=0}^{N-1} \left\{ \frac{e^{-i\frac{\pi}{N}m_1(2s_1+s_2)}}{\left(1 - \frac{z}{3} \cos\left(\frac{2\pi}{N}m_1\right)\right)\left(1 - c_{m_1}^2(z)\right)^{\frac{1}{2}}} \times \frac{\alpha_{m_1}^{s_2}(z) + \alpha_{m_1}^{N-s_2}(z) \cos(\pi m_1)}{1 - \alpha_{m_1}^N(z) \cos(\pi m_1)} \right\}. \quad (16)$$

Note that the summand involved in (16) is holomorphic over  $0 < z < 1$  for all  $0 \leq m_1 \leq N-1$ . However, it diverges at  $z = 1$  when  $m_1 = 0$ . Thus, the singularity of  $P(\mathbf{0}, \vec{s} | z)$  at  $z = 1$  comes only from the first term of the sum given by (16). It is convenient therefore to separate out the singular and non-singular parts of  $P(\mathbf{0}, \vec{s} | z)$  as follows

$$P(\mathbf{0}, \vec{s} | z) = \frac{\sqrt{3}}{N(1-z)^{\frac{1}{2}}(3+z)^{\frac{1}{2}}} \times \frac{\alpha_0^{s_2}(z) + \alpha_0^{N-s_2}(z)}{1 - \alpha_0^N(z)} + \varphi_N(\vec{s}, z) \quad (17)$$

where

$$\varphi_N(\vec{s}, z) = \frac{1}{N} \sum_{m_1=1}^{N-1} \left\{ \frac{e^{-i\frac{\pi}{N}m_1(2s_1+s_2)}}{\left(1 - \frac{z}{3} \cos\left(\frac{2\pi}{N}m_1\right)\right)\left(1 - c_{m_1}^2(z)\right)^{\frac{1}{2}}} \times \frac{\alpha_{m_1}^{s_2}(z) + \alpha_{m_1}^{N-s_2}(z) \cos(\pi m_1)}{1 - \alpha_{m_1}^N(z) \cos(\pi m_1)} \right\} \quad (18)$$

is holomorphic at  $z = 1$ . The first term involved in (17) corresponds to the term  $m_1 = 0$  in (16), and the second term,  $\varphi_N(\vec{s}, z)$ , corresponds to the sum  $1 \leq m_1 \leq N-1$ .

## B.2 Asymptotic expansion of $P(\mathbf{0}, \vec{s} | z)$ as $z \rightarrow 1$

To obtain the zero-order asymptotic expansion of  $P(\mathbf{0}, \vec{s} | z)$  as  $z \rightarrow 1$ , let us successively expand the first term of  $P(\mathbf{0}, \vec{s} | z)$  involved in (17) and then function  $\varphi_N(\vec{s}, z)$  close to  $z = 1$ . After expanding function  $\alpha_0(z)$ , it can be deduced that

$$\frac{\sqrt{3}}{N(1-z)^{\frac{1}{2}}(3+z)^{\frac{1}{2}}} \times \frac{\alpha_0^{s_2}(z) + \alpha_0^{N-s_2}(z)}{1 - \alpha_0^N(z)} = \frac{1}{N^2(1-z)} + \frac{N^2 - 6Ns_2 + 6s_2^2 - 1}{4N^2} + o((1-z)^{\frac{1}{2}}).$$

Using now the *Taylor's Theorem*,  $\varphi_N(\vec{s}, z)$  can be represented by its zero-order *Taylor series expansion* at  $z = 1$

$$\varphi_N(\vec{s}, z) = \varphi_N(\vec{s}, 1) + o((1-z)^{\frac{1}{2}}).$$

Finally, combining this asymptotic expansion with the one of the first term of  $P(\mathbf{0}, \vec{s} | z)$ , we obtain

$$P(\mathbf{0}, \vec{s} | z) = \frac{1}{N^2(1-z)} + \frac{N^2(1 + 4\varphi_N(\vec{s}, 1)) - 6Ns_2 + 6s_2^2 - 1}{4N^2} + o((1-z)^{\frac{1}{2}}). \quad (19)$$

## C. ASYMPTOTIC EXPANSION OF $\varphi_N(\mathbf{0}, 1)$

Setting  $\vec{s} = \mathbf{0}$  and  $z = 1$  in (18), we obtain

$$\varphi_N(\mathbf{0}, 1) = \frac{\sqrt{3}}{2N} \sum_{m_1=1}^{N-1} \left\{ \frac{1}{\sin\left(\frac{\pi}{N}m_1\right)\left(1 + \frac{1}{3} \sin^2\left(\frac{\pi}{N}m_1\right)\right)^{\frac{1}{2}}} \times \frac{1 + \alpha_{m_1}^N(1) \cos(\pi m_1)}{1 - \alpha_{m_1}^N(1) \cos(\pi m_1)} \right\}. \quad (20)$$

In this section, we show that series  $\varphi_N(\mathbf{0}, 1)$  has the following asymptotic expansion as  $N \rightarrow \infty$

$$\varphi_N(\mathbf{0}, 1) = \frac{\sqrt{3}}{\pi} \ln(N) + \frac{\sqrt{3}}{\pi} \left( \gamma + \ln\left(\frac{\sqrt{3}}{\pi}\right) + \frac{\sqrt{3}\pi}{3} \ell \right) + O\left(\frac{1}{N^4}\right) \quad (21)$$

where  $\gamma$  is the *Euler's constant* and  $\ell$  is a constant. This asymptotic expansion is obtained by writing  $\varphi_N(\mathbf{0}, 1)$  as

$$\varphi_N(\mathbf{0}, 1) = Q_1(N) + Q_2(N) + Q_3(N),$$

where

$$\begin{aligned} Q_1(N) &= \frac{\sqrt{3}}{2N} \sum_{m_1=1}^{N-1} \frac{1}{\sin\left(\frac{\pi}{N}m_1\right)}, \\ Q_2(N) &= \frac{\sqrt{3}}{2N} \sum_{m_1=1}^{N-1} \frac{\left(1 + \frac{1}{3} \sin^2\left(\frac{\pi}{N}m_1\right)\right)^{-\frac{1}{2}} - 1}{\sin\left(\frac{\pi}{N}m_1\right)}, \\ Q_3(N) &= \frac{\sqrt{3}}{N} \sum_{m_1=1}^{N-1} \frac{\left(1 + \frac{1}{3} \sin^2\left(\frac{\pi}{N}m_1\right)\right)^{-\frac{1}{2}} \alpha_{m_1}^N(1) \cos(\pi m_1)}{\sin\left(\frac{\pi}{N}m_1\right)\left(1 - \alpha_{m_1}^N(1) \cos(\pi m_1)\right)} \end{aligned}$$

and then by calculating separately the asymptotic expansion of each sum as  $N \rightarrow \infty$ .

### C.1 Asymptotic expansion of $Q_1(N)$ as $N \rightarrow \infty$

Let  $f(x)$  be the function defined for all real numbers  $x$  between 0 and  $\pi$  as follows

$$f(x) = \begin{cases} \frac{1}{\sin(x)} - \frac{1}{x} - \frac{1}{\pi - x}, & x \in ]0, \pi[ \\ -\frac{1}{\pi}, & x = 0, \pi. \end{cases}$$

We can show that  $f$  is indefinitely differentiable, in particular

$$f^{(1)}(0) = \frac{1}{6} - \frac{1}{\pi^2} \quad \text{and} \quad f^{(1)}(\pi) = -\frac{1}{6} + \frac{1}{\pi^2}.$$

Remark also that  $Q_1(N)$  can be expressed as

$$Q_1(N) = \frac{\sqrt{3}}{2N} \sum_{m_1=1}^{N-1} f\left(\frac{\pi}{N}m_1\right) + \frac{\sqrt{3}}{\pi} \sum_{m_1=1}^{N-1} \frac{1}{m_1}. \quad (22)$$

Therefore, using the *Euler-Maclaurin summation formula* [10]

$$\begin{aligned} \frac{1}{N} \sum_{m_1=1}^{N-1} h\left(\frac{\pi}{N}m_1\right) &= \frac{1}{\pi} \int_0^\pi h(x) dx - \frac{1}{2N} (h(\pi) + h(0)) \\ &\quad + \frac{\pi}{12N^2} (h^{(1)}(\pi) - h^{(1)}(0)) + O\left(\frac{1}{N^4}\right) \end{aligned} \quad (23)$$

where  $h$  is an indefinitely differentiable function, and since  $\int_0^\pi f(x) dx = 2 \ln(\frac{2}{\pi})$ , we obtain

$$\frac{1}{N} \sum_{m_1=1}^{N-1} f\left(\frac{\pi}{N} m_1\right) = \frac{2}{\pi} \ln\left(\frac{2}{\pi}\right) + \frac{1}{N\pi} - \frac{\pi}{6N^2} \left(\frac{1}{6} - \frac{1}{\pi^2}\right) + O\left(\frac{1}{N^4}\right). \quad (24)$$

It remains now to expand the second sum involved in (22), which can be recognized as the *Harmonic* series, that is  $H_N = \sum_{m_1=1}^{N-1} \frac{1}{m_1}$ . According to [10],  $H_N$  has the following asymptotic expansion

$$H_N = \ln(N) + \gamma - \frac{1}{2N} - \frac{1}{12N^2} + O\left(\frac{1}{N^4}\right), \quad (25)$$

where  $\gamma$  is the *Euler's* constant ( $\gamma = 0.5772156649$ ). By plugging (24) and (25) into (22), we find

$$Q_1(N) = \frac{\sqrt{3}}{\pi} \ln(N) + \frac{\sqrt{3}}{\pi} \left(\gamma + \ln\left(\frac{2}{\pi}\right)\right) - \frac{\sqrt{3}\pi}{72N^2} + O\left(\frac{1}{N^4}\right). \quad (26)$$

## C.2 Asymptotic expansion of $Q_2(N)$ as $N \rightarrow \infty$

Let  $g(x)$  be the function defined for all real numbers  $x$  between 0 and  $\pi$  as follows

$$g(x) = \begin{cases} \frac{(1 + \frac{1}{3} \sin^2(x))^{-\frac{1}{2}} - 1}{\sin(x)}, & x \in ]0, \pi[ \\ 0, & x = 0, \pi. \end{cases}$$

We can show that  $g$  is indefinitely differentiable, in particular

$$g^{(1)}(0) = -\frac{1}{6} \quad \text{and} \quad g^{(1)}(\pi) = \frac{1}{6}.$$

Note that  $Q_2(N)$  can be expressed as

$$Q_2(N) = \frac{\sqrt{3}}{2N} \sum_{m_1=1}^{N-1} g\left(\frac{\pi}{N} m_1\right).$$

Using the *Euler-Maclaurin* summation formula, we obtain

$$\begin{aligned} \frac{1}{N} \sum_{m_1=1}^{N-1} g\left(\frac{\pi}{N} m_1\right) &= \frac{1}{\pi} \int_0^\pi g(x) dx - \frac{1}{2N} (g(\pi) + g(0)) \\ &\quad + \frac{\pi}{12N^2} (g^{(1)}(\pi) - g^{(1)}(0)) + O\left(\frac{1}{N^4}\right). \end{aligned}$$

The integral has the value

$$\int_0^\pi g(x) dx = \ln\left(\frac{3}{4}\right)$$

so that

$$Q_2(N) = \frac{\sqrt{3}}{2\pi} \ln\left(\frac{3}{4}\right) + \frac{\sqrt{3}\pi}{72N^2} + O\left(\frac{1}{N^4}\right). \quad (27)$$

## C.3 Asymptotic expansion of $Q_3(N)$ as $N \rightarrow \infty$

Before evaluating  $Q_3(N)$ , note that the terms corresponding to the values of  $m_1$  near 1 or  $N-1$  contributes to  $Q_3(N)$  more significantly than the ones corresponding to the values of  $m_1$  near  $[\frac{N}{2}]$  when  $N$  is large. Indeed, from (14) we see that when  $m_1 \simeq 1$  or  $N-1$ ,  $\sin(\frac{\pi}{N} m_1)$  is close to 0 and  $|\alpha_{m_1}(1)|$  close to 1. In this range, the summand of  $Q_3(N)$  diverges as  $N \rightarrow \infty$ . However, when  $m_1$  is close to  $[\frac{N}{2}]$ ,  $\sin(\frac{\pi}{N} m_1)$  is close to 1 and  $\alpha_{m_1}(1)$  close to 0, and hence,  $\alpha_{m_1}^N(1)$  vanishes exponentially to zero as  $N \rightarrow \infty$ . Therefore, the summand of  $Q_3(N)$  decreases to zero faster than any power of  $\frac{1}{N}$  and consequently can be neglected.

We can see also that the summand of  $Q_3(N)$  is symmetric about the  $\frac{N}{2}$ -axis, that is, the  $(N-m_1)$ -term equals to the  $m_1$ -term. Therefore, by neglecting the  $\frac{N}{2}$ -term only when  $N$  is even,  $Q_3(N)$  can be rewritten as

$$Q_3(N) = \frac{2\sqrt{3}}{N} \sum_{m_1=1}^{\lfloor \frac{N-1}{2} \rfloor} \frac{(1 + \frac{1}{3} \sin^2(\frac{\pi}{N} m_1))^{-\frac{1}{2}} \alpha_{m_1}^N(1) \cos(\pi m_1)}{\sin(\frac{\pi}{N} m_1) (1 - \alpha_{m_1}^N(1) \cos(\pi m_1))}. \quad (28)$$

In order to approximate  $Q_3(N)$  when  $N$  is very large, we use a method pioneered by *Laplace* and described in [10]. It consists in breaking the sum into two disjoint ranges  $D_N$  and  $T_N$ . The summation over  $D_N$  should be the *dominant* part, in the sense that it includes enough terms to determine the significant digits of the sum when  $N$  is very large. The summation over the other range  $T_N$  should be just the *tail* end, which contributes little to the overall total. As the big contributions to  $Q_3(N)$  occur when  $m_1$  is small, we can consider  $D_N$  as the range of small values of  $m_1$ , and  $T_N$  the range of large values of  $m_1$ . In the following we construct the asymptotic expansion of  $Q_3(N)$  by separating out the dominant and the tail ranges, and then bounding the tail range contributions.

Let  $a_{m_1}(N)$  denote the summand involved in (28). By separating out the dominant and the tail ranges,  $Q_3(N)$  can be then expressed as

$$Q_3(N) = \sum_{m_1 \in D_N} a_{m_1}(N) + \sum_{m_1 \in T_N} a_{m_1}(N). \quad (29)$$

As long as  $m_1$  is held in the dominant range,  $\frac{\pi m_1}{N}$  goes to zero as  $N \rightarrow \infty$ . Therefore,  $a_{m_1}(N)$  can be expanded as

$$a_{m_1}(N) = b_{m_1}(N) + O(c_{m_1}(N)) \quad (30)$$

where

$$\begin{aligned} b_{m_1}(N) &= \frac{2\sqrt{3} e^{-\sqrt{3}\pi m_1} \cos(\pi m_1)}{\pi m_1 (1 - e^{-\sqrt{3}\pi m_1} \cos(\pi m_1))} \\ c_{m_1}(N) &= \frac{m_1^3 e^{-\sqrt{3}\pi m_1} \cos(\pi m_1)}{N^4 (1 - e^{-\sqrt{3}\pi m_1} \cos(\pi m_1))}. \end{aligned}$$

This asymptotic expansion is valid as long as  $m_1 \in D_N$  and thus, it is allowed to take the summation of both sides of (30). Hence

$$\sum_{m_1 \in D_N} a_{m_1}(N) = \sum_{m_1 \in D_N} b_{m_1}(N) + O\left(\sum_{m_1 \in D_N} |c_{m_1}(N)|\right). \quad (31)$$

However,

$$\sum_{m_1 \in D_N} |c_{m_1}(N)| \leq \frac{1}{N^4} \sum_{m_1=1}^{\infty} \frac{m_1^3 e^{-\sqrt{3}\pi m_1}}{1 - e^{-\sqrt{3}\pi m_1} \cos(\pi m_1)}$$

where the series  $\sum \frac{m_1^3 e^{-\sqrt{3}\pi m_1}}{1 - e^{-\sqrt{3}\pi m_1} \cos(\pi m_1)}$  converges. Thus,

$$\sum_{m_1 \in D_N} |c_{m_1}(N)| = O\left(\frac{1}{N^4}\right), \text{ so that,}$$

$$\sum_{m_1 \in D_N} a_{m_1}(N) = \sum_{m_1 \in D_N} b_{m_1}(N) + O\left(\frac{1}{N^4}\right).$$

Therefore, (29) can be rewritten as

$$\begin{aligned}
Q_3(N) &= \sum_{m_1 \in D_N} b_{m_1}(N) + \sum_{m_1 \in T_N} a_{m_1}(N) + O\left(\frac{1}{N^4}\right) \\
&= \sum_{m_1=1}^{\lfloor \frac{N-1}{2} \rfloor} b_{m_1}(N) - \sum_{m_1 \in T_N} b_{m_1}(N) + \sum_{m_1 \in T_N} a_{m_1}(N) \\
&\quad + O\left(\frac{1}{N^4}\right).
\end{aligned} \tag{32}$$

We evaluate now the tail range contributions. It suffices to find good bounds on sums  $\sum_{m_1 \in T_N} a_{m_1}(N)$  and  $\sum_{m_1 \in T_N} b_{m_1}(N)$ . Indeed, as already shown, when  $m_1$  is held in  $T_N$ ,  $\alpha_{m_1}(1)$  goes to zero and hence,  $\alpha_{m_1}^N(1)$  decreases faster than any power of  $\frac{1}{N}$  as  $N \rightarrow \infty$ . It can be then deduced after a certain amount of elementary calculus that  $\sum_{m_1 \in T_N} a_{m_1}(N) = O\left(\frac{1}{N^4}\right)$ . Similarly, since  $e^{-\sqrt{3}\pi m_1}$  vanishes faster than any power of  $\frac{1}{N}$ , we can show that  $\sum_{m_1 \in T_N} b_{m_1}(N) = O\left(\frac{1}{N^4}\right)$ . One finds that

$$Q_3(N) = \sum_{m_1=1}^{\lfloor \frac{N-1}{2} \rfloor} b_{m_1}(N) + O\left(\frac{1}{N^4}\right). \tag{33}$$

However, we can show that series  $\sum_{m_1=1}^{\frac{N-1}{2}} b_{m_1}(N)$  converges to a constant  $\ell$  as  $N \rightarrow \infty$ . Since the exponential function grows faster than any power of  $N$ , we can then write

$$\sum_{m_1=1}^{\lfloor \frac{N-1}{2} \rfloor} b_{m_1}(N) = \ell + O\left(\frac{1}{N^4}\right), \text{ and finally}$$

$$Q_3(N) = \ell + O\left(\frac{1}{N^4}\right). \tag{34}$$

## D. REFERENCES

- [1] MC13192, 2.4 GHz low power transceiver for the IEEE 802.15.4 standard, may 2007.
- [2] M. Alanyali, V. Saligrama, and O. Sava. A random-walk model for distributed computation in energy-limited network. In *Proc. of 1st Workshop on Information Theory and its Application*, 2006.
- [3] C. Avin and C. Brito. Efficient and robust query processing in dynamic environments using random walk techniques. In *Proc. of the third international symposium on Information processing in sensor networks*, 2004.
- [4] X. Bai, S. Kumar, Z. Yun, D. Xuan, and T. H. Lai. Deploying wireless sensors to achieve both coverage and connectivity. In *Proceedings of the Seventh International Symposium on Mobile Ad Hoc Networking and Computing (ACM MobiHoc)*, Florence, Italy, 2006.
- [5] A. Chakrabarti, A. Sabharwal, and B. Aazhang. Multi-hop communication is order-optimal for homogeneous sensor networks. In *Proc. of the third international symposium on Information processing in sensor networks*, pages 178–185, NY, USA, April 2004. ACM Press.
- [6] P. Chen, B. O’Dea, and E. Callaway. Energy efficient system design with optimum transmission range for wireless ad hoc networks. In *Proc. of the IEEE International Conference on Communications*, volume 1, pages 945–952, April 2002.
- [7] S. Dolev, E. Schiller, and J. Welch. Random walk for self-stabilizing group communication in ad-hoc networks. In *Proc. of the 21st IEEE Symposium on Reliable Distributed Systems (SRDS’02)*. IEEE Computer Society, 2002.
- [8] S. Ergen and P. Varaiya. On multi-hop routing for energy efficiency. *IEEE Communications Letters*, 9(10):880–881, October 2005.
- [9] C. Gkantsidis, M. Mihail, and A. Saberi. Random walks in peer-to-peer networks. In *Proc. of IEEE INFOCOM*, 2004.
- [10] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley Professional, 2nd edition, 1994.
- [11] B. D. Hughes. *Random Walks and Random Environments*. Oxford University Press, New York, 1995.
- [12] R. Kershner. The number of circles covering a set. *American Journal of Mathematics*, 61(3):665–671, July 1939.
- [13] S. K. Lando. *Lectures on Generating Functions*. American Mathematical Society, 2003.
- [14] L. Lima and J. Barros. Random walks on sensor networks. In *Proc. of the 5th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks*, April 2007.
- [15] 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*, 23(1):11–25, July 1980.
- [16] 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*, 22:53–72, 1982.
- [17] E. W. Montroll and G. H. Weiss. Random walks on lattices. II. *Journal of Mathematical Physics*, 6(2):167–181, February 1965.
- [18] H. Scher and M. Lax. Stochastic transport in a disordered solid. I. theory. *Phys. Rev. B*, 7(10):4491–4502, May 1973.
- [19] 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*, pages 12–21, NY, USA, 2002. ACM Press.
- [20] H. Tian, H. Shen, and T. Matsuzawa. Random walk routing for wireless sensor networks. In *Proc. of the Sixth International Conference on Parallel and Distributed Computing Applications and Technologies*, pages 196–200, Washington, DC, USA, 2005. IEEE Computer Society.