A Probabilistic Algorithm for Efficient and Robust Data Propagation in Smart Dust Networks *

I. Chatzigiannakis † *T. Dimitriou* ‡ *S. Nikoletseas* † *P. Spirakis* †

Abstract

We study the problem of data propagation in sensor networks, comprised of a large number of very small and low-cost nodes, capable of sensing, communicating and computing. The distributed co-operation of such nodes may lead to the accomplishment of large sensing tasks, having useful applications in practice. We present a new protocol for data propagation towards a control center ("sink") that avoids flooding by probabilistically favoring certain ("close to optimal") data transmissions.

This protocol is very simple to implement in sensor devices and operates under total absence of co-ordination between sensors. We consider a network model of *randomly deployed sensors* of sufficient density. As shown by a geometry analysis, the protocol is *correct*, since it always propagates data to the sink, under ideal network conditions (no failures). Using stochastic processes, we show that the protocol is *very energy efficient*. Also, when part of the network is inoperative, the protocol manages to propagate data very close to the sink, thus in this sense it is *robust*. We finally present and discuss large-scale experimental findings validating the analytical results.

1 Introduction, our Results and Related Work

Recent dramatic developments in micro-electro-mechanical (MEMS) systems, wireless communications and digital electronics have led to the development of small in size, low-power, low-cost sensor devices. Such extremely small devices integrate sensing, data processing and communication capabilities. Examining each such device individually might appear to have small utility, however the effective *distributed co-ordination* of large numbers of such devices may lead to the efficient accomplishment of large sensing tasks.

Large numbers of sensors can be deployed in areas of interest (such as inaccessible terrains or disaster places) and use *self-organization and collaborative methods* to form a sensor network. Their wide range of applications is based on the use of various sensor types (i.e. thermal, visual, seismic, acoustic, radar, magnetic, etc.) to monitor a wide variety of conditions (e.g. temperature, object presence and movement, humidity, pressure, noise levels etc.). Thus, sensor networks can be used for continuous sensing, event detection, location sensing as well as micro-sensing. Hence, sensor networks have important applications, including (a) military (like forces and equipment monitoring, battlefield surveillance, targeting, nuclear, biological and chemical attack detection), (b) environmental applications (such as fire detection, flood detection, precision agriculture), (c) health applications (like telemonitoring of human physiological data) and (d) home applications (e.g. smart environments and home automation). For a survey of wireless sensor networks see [1] and also [5, 8].

Note however that the efficient and robust realization of such large, highly-dynamic, complex, non-conventional networks is *a challenging algorithmic and technological task*. Features including the huge number of sensors involved, the severe power, computational and memory limitations, their dense deployment and

*This work has been partially supported by the IST Programme of the European Union under contract numbers IST-1999-14186 (ALCOM-FT) and IST-2001-33116 (FLAGS).

†1. Computer Technology Institute (CTI) and Patras University, P. 0. Box 1122, 261 10 Patras, Greece. Email: {ichatz,nikole,spirakis}@cti.gr

‡2. Athens Information Technology (AIT), Athens, Greece. Email: tassos@ait.gr

frequent failures, pose *new design and implementation aspects* which are essentially different not only to distributed computing and systems approaches but also to ad-hoc networking techniques.

Problem Description: We focus on an important problem under a model of sensor networks that we present. More specifically, we study the problem of *local detection and propagation*, i.e. the local sensing of a crucial event and the energy and time efficient propagation of data reporting its realization to a control center. This center could be some human authorities responsible of taking action upon the realization of the crucial event. We use the term *"sink"* for this control center. We note that the protocol we present here can also be used for the more general problem of data propagation in sensor networks (see [6]).

Our Protocol: For the above problem we propose a new protocol which tries to minimize energy consumption by *probabilistically favoring certain paths of local data transmissions towards the sink.* Thus we call this protocol PFR (Probabilistic Forwarding Protocol).

The basic idea behind this protocol is to *avoid flooding* by favoring (in a probabilistic manner) data propagation along sensors which lie "close" to the (optimal) transmission line, ES, that connects the sensor node detecting the event, E, and the sink, S. This is implemented by locally calculating the angle $\phi = (\widehat{EPS})$, whose corner point P is the sensor currently running the local protocol, having received a transmission from a nearby sensor, previously possessing the event information. If ϕ is equal or greater to a predetermined threshold, then p will transmit (and thus propagate the event information further). Else, it decides whether to transmit with probability equal to $\frac{\phi}{\pi}$. The way to estimate ϕ is explained in Section 4. Because of the probabilistic nature of data propagation decisions and in order to prevent the data propagation process from early failing, we initially use (for a short time period which we evaluate) a flooding mechanism that leads to a sufficiently large *"front"* of sensors possessing the data under propagation. When such a "front" is created, we perform probabilistic forwarding.

The protocol, as shown by a geometric analysis, always propagates data to the sink, under ideal network conditions (no failures), thus it is provably correct. Using properties of stochastic processes, we show that the protocol is very energy efficient. Also, when part of the network is inoperative (which is more realistic, because sensors are prone to faults), the protocol manages to propagate data very close to the sink, thus it is robust. The above analytical results are validated by large scale experiments we have carried out after implementing the protocol.

Essentially, our protocol captures the intuitive, deterministic idea "if my distance from ES is small, then send, else do not send". We have chosen to enhance this idea by random decisions (above a threshold) to allow some local flooding to happen with small probability and thus to cope with local sensor failures.

Related Work: Our protocol is inspired by the relevant work of [6], where *"directed diffusion"*, an approach to attribute-based data communication for wireless sensor networks is proposed. The goal of directed diffusion is to establish communication between sources and sinks. Directed diffusion consists of several elements. Data is *named* using attribute-valued pairs. A sensing task is disseminated throughout the sensor network as an *interest* for named data. This dissemination sets up gradients within the network designed to "draw" events (i.e. data matching the interest). Events start flowing towards the originators of interests along multiple paths. The sensor network *reinforces* one, or a small number of these paths. In particular, we build upon the proposal of [6] for the data propagation part of their paradigm, where "multipath delivery with probabilistic forwarding" is mentioned. We note that, in contrast to [6], our protocol does not need a "build-up" phase where gradients are set up, a fact that may reduce the overhead of "control" messages. In our case, only transmissions of data reporting the event realization occur.

Furthermore, this work is related to previous research of [2, 3], where new local detection and propagation protocols are proposed, that are very energy and time efficient, as shown by a rigorous average case analysis performed in these works under certain simplifying assumptions.

2 The Model

Sensor networks are comprised of a vast number of ultra-small homogenous sensors, which we call *"grain" particles.* Each grain particle is a fully-autonomous computing and communication device, characterized mainly by its available power supply (battery) and the energy cost of computation and transmission of data. Such particles (in our model here) cannot move. Each particle is equipped with a set of monitors (sensors) for light, pressure, humidity, temperature etc. and has a *broadcast* (digital radio) *beacon mode.*

Certainly, because of the dense deployment of sensors close to each other, multi-hop communication consumes less power than a single hop propagation. Also, multi-hop communication can effectively overcome some of the signal propagation effects in long-distance wireless transmissions. Furthermore, short-range hop-by-hop transmissions may help to smoothly adjust propagation around obstacles. Finally, the low energy transmission in hop-by-hop propagation may enhance security, protecting from undesired discovery of the data propagation operation. Because of the above reasons, data propagation should be done in a multi-hop way.

We assume the case where particles are *randomly deployed* in a given area of interest. Such a placement may occur e.g. when throwing sensors from an airplane to the area of interest.

Remark: *As a special case*, we could consider the network being a lattice (or grid) deployment of sensors. This grid placement of grain particles is motivated by certain applications, where it is possible to have a pre-deployed sensor network, where sensors are put (possibly by a human or a robot) in a way that they form a *2-dimensional lattice.* Note indeed that such sensor networks, deployed in a structured way, might be useful in precise agriculture, where humans or robots may want to deploy the sensors in a lattice structure to monitor in a rather homogenous and uniform way certain conditions in the spatial area of interest. Certainly, exact terrain monitoring in military applications may also need some sort of a grid-like shaped sensor network. Note also that Akyildiz et al in a recent (2002) state of the art survey published in the Journal of Computer Networks ([1]) do not exclude the pre-deployment possibility. Also Intanagonwiwat, Govindan and Estrin, a group of pioneering researchers in smart dust, in a recent paper appeared in ACM MOBICOM 2002 ([6]) explicitly refer to the lattice case. Moreover, as they state in an extended version of their work ([7]), they consider, for reasons of “analytic tractability”, a square grid topology.

Let N be the number of deployed grain particles. There is a single point in the network area, which we call the sink S, and represents a control center where data should be propagated to (see Fig. 1).

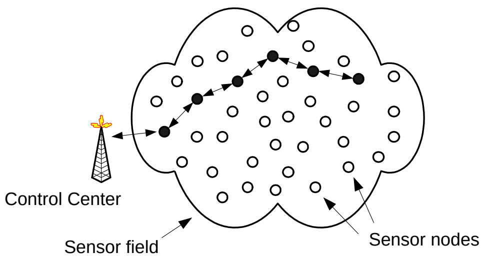

Figure 1: A Sensor Network

We assume that each grain particle has the following abilities:

(i) It can estimate the direction of a received transmission (e.g. via the technology of direction-sensing antennae).

(ii) It can estimate the distance from a nearby particle that did the transmission (e.g. via estimation of the attenuation of the received signal).

(iii) It knows the direction towards the sink S. This can be implemented during a set-up phase, where the (very powerful in energy) sink broadcasts the information about itself to all particles.

(iv) All particles have a common co-ordinates system.

Notice that GPS information is not needed for our protocol. Also, there is no need to know the global structure of the network.

3 The Problem

Assume that a single particle, P, senses the realization of a *local crucial event* $\mathcal{E}$. Then the *propagation problem* $\mathcal{P}$ is the following:

> "How can particle P, via cooperation with the rest of the grain particles, efficiently propagate information $info(\mathcal{E})$, reporting realization of event $\mathcal{E}$, to the sink S?"

To minimize the energy consumption in the sensor network we wish to avoid flooding and *to minimize the number of hops (directed transmissions)* performed in the data propagation process, while still managing to reach the sink.

Definition 3.1 Let E be the position of P which senses event $\mathcal{E}$. Let S be the position of the sink S.

Any protocol Π solving the propagation problem $\mathcal{P}$ must satisfy:

- **Correctness**. Π must guarantee that data arrives to the position S, given that the whole network exists and is operational.
- **Robustness**. Π must guarantee that data arrives at enough points in a small interval around S, in cases where part of the network has become inoperative.
- **Efficiency**. If Π activates k particles during its operation then Π should have a small ratio of the number of activated over the total number of particles $r = \frac{k}{N}$. Thus r is an energy efficiency measure of Π.

4 The Probabilistic Forwarding Protocol (PFR)

As already mentioned in Section 1, the basic idea of the protocol lies in probabilistically favoring transmissions towards the sink within a *thin zone* of particles around the line connecting the particle sensing the event $\mathcal{E}$ and the sink (see Fig. 2). Note that transmission along this line is energy optimal. However it is not always possible to achieve this optimality, basically because certain sensors on this direct line might be inactive, either permanently (because their energy has been exhausted) or temporarily (because these sensors might enter a sleeping mode to save energy). Further reasons include (a) *physical damage* of sensors, (b) *deliberate removal* of some of them (possibly by an adversary in military applications), (c) *changes in the position* of the sensors due to a variety of reasons (weather conditions, human interaction etc). and (d) *physical obstacles* blocking communication.

The protocol evolves in two phases:

Phase 1: The "Front" Creation Phase. Initially we build (by using a limited, in terms of rounds, flooding) a sufficiently large "front" of particles, in order to guarantee the survivability of the data propagation process. During this phase, each particle having received the data to be propagated, deterministically forwards them towards the sink. In particular, and for a sufficiently large number of steps $s = 180\sqrt{2}$, each particle broadcasts the information to all its neighbors, towards the sink. Remark that to implement this phase, and in particular to count the number of steps, we use a counter in each message. This counter needs at most $\lceil \log 180\sqrt{2} \rceil$ bits.

Phase 2: The Probabilistic Forwarding Phase. During this phase, each particle P possessing the information under propagation, calculates an angle ϕ by calling the subprotocol "ϕ-calculation" (see description below) and broadcasts $info(\mathcal{E})$ to all its neighbors with probability $\mathbb{P}_{fwd}$ (or it does not propagate any data with probability $1 - \mathbb{P}_{fwd}$) defined as follows:

$$\mathbb{P}_{fwd} = \begin{cases} 1 & \text{if } \phi \geq \phi_{threshold} \\ \frac{\phi}{\pi} & \text{otherwise} \end{cases}$$

where ϕ is the angle defined by the line EP and the line PS and $\phi_{threshold} = 134^o$ (the selection reasons of this $\phi_{threshold}$ will become more evident in Section 5.1).

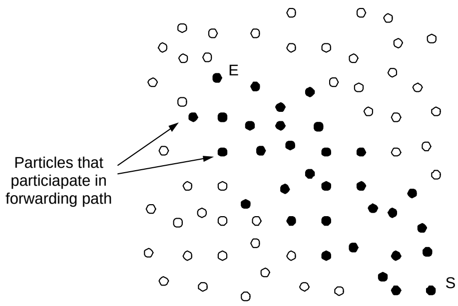

Figure 2: Thin Zone of particles

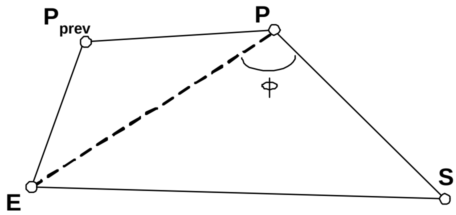

Figure 3: Angle ϕ calculation example

In both phases, if a particle has already broadcast $info(\mathcal{E})$ and receives it again, it ignores it. Also the PFR protocol is presented for a single event tracing. Thus no multiple paths arise and packet sizes do not increase with time.

Remark that when $\phi = \pi$ then P lies on the line ES and vice-versa (and always transmits).

If the density of particles is appropriately large, then for a line ES there is (with high probability) a sequence of points "closely surrounding ES" whose angles ϕ are larger than $\phi_{threshold}$ and so that successive points are within transmission range. All such points broadcast and thus essentially they follow the line ES (see Fig. 2).

The ϕ-calculation subprotocol (see Fig. 3)

Let P_{prev} the particle that transmitted $info(E)$ to P.

(1) When P_{prev} broadcasts $info(E)$, it also attaches the info $|EP_{prev}|$ and the direction $\overrightarrow{P_{prev}E}$.

(2) P estimates the direction and length of line segment $P_{prev}P$, as described in the model.

(3) P now computes angle $(\widehat{EP_{prev}P})$, and computes $|EP|$ and the direction of $\overrightarrow{PE}$ (this will be used in further transmission from P).

(4) P also computes angle $(\widehat{P_{prev}PE})$ and by subtracting it from $(\widehat{P_{prev}PS})$ it finds ϕ.

Notice the following:

(i) The direction and distance from activated sensors to E is inductively propagated (i.e. P becomes P_{prev} in the next phase).

(ii) Our protocol needs only messages of length bounded by $\log A$, where A is some measure of the size of the network area, since (because of (i) above) there is no cumulative effect on message lengths.

5 Properties of PFR

Consider a partition of the network area into small squares of a fictitious grid G (see Fig. 4). Let the length of the side of each square be l. Let the number of squares be q. The area covered is bounded by

ql^2. Assuming that we randomly throw in the area at least $\alpha q \log q = N$ particles (where $\alpha > 0$ a suitable constant), then the probability that a particular square is avoided is

$$\left(1 - \frac{1}{q}\right)^{\alpha q \log q} \leq e^{-\alpha \log q} = q^{-\alpha}$$

So the probability that all squares get particles is at least

$$1 - q \cdot q^{-\alpha} = 1 - q^{-(\alpha-1)} = 1 - \left(\Theta\left(\frac{N}{\log N}\right)\right)^{-(\alpha-1)}$$

We condition all the analysis on this event, call it F, of at least one particle in each square (see also [4]).

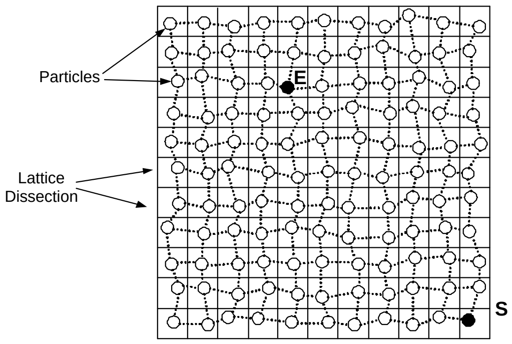

Figure 4: A Lattice Dissection G

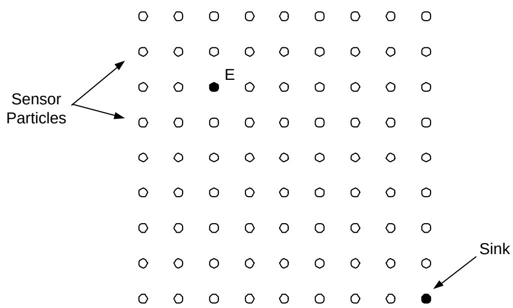

Figure 5: A Lattice Sensor Network

5.1 The Correctness of PFR

Without loss of generality, we assume each square of the fictitious lattice G to have side length 1.

Lemma 5.1 PFR *succeeds with probability 1* in sending the information from E to S given the event F.

Proof: In the (trivial) case where $|ES| \leq 180\sqrt{2}$, the protocol is clearly correct due to front creation phase.

Let Σ a unit square of G intersecting ES in some way (see Fig. 6). Since a particle always exists somewhere in Σ, we will only need to examine the worst case of it being in one of the corners of Σ. Consider vertex A. The line EA is always to the left of AB (since E is at end of ES). The same is true for AS (S is to the right of B').

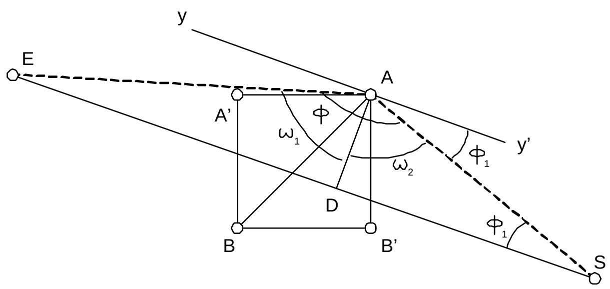

Figure 6: The Square Σ

Let AD the segment from A perpendicular to ES (and D its intersection point) and let yy' be the line from A parallel to ES. Then,

$$\begin{aligned}\phi &= (\widehat{EAS}) = 180^o - (\widehat{yAE}) - (\widehat{y'AS}) \\ &= 90^o - (\widehat{yAE}) + 90^o - (\widehat{y'AS}) = (\widehat{yAD}) - (\widehat{yAE}) + (\widehat{DAy'}) - (\widehat{y'AS})\end{aligned}$$

Let $\widehat{\omega_1} = (\widehat{yAD}) - (\widehat{yAE})$ and $\widehat{\omega_2} = (\widehat{DAy'}) - (\widehat{y'AS})$ and, without loss of generality, let $ED < DS$. Then always $\widehat{\omega_1} > 45^o$, since it includes half of the 90^o-angle of A in the unit square. Also

$$\sin(\widehat{y'AS}) \;=\; \sin(\widehat{ASD}) \;=\; \frac{AD}{AS} \;<\; \frac{AD}{DS}$$

$$\text{but} \quad AD \;\leq\; \sqrt{2} \quad (= AB) \quad \text{and} \quad DS \geq \frac{ES}{2} \geq 90\sqrt{2}$$

$$\Longrightarrow \sin(\widehat{y'AS}) \leq \frac{1}{90} \quad \Longleftrightarrow \quad (\widehat{y'AS}) < 1^o$$

(Note that here we use the elementary fact that when $x < 90^o$ then $sinx < \frac{1}{90} \Longleftrightarrow x < 1^o$)

Then $\widehat{\omega_2} > 89^o$ since $(\widehat{DAy'}) = 90^o$ by construction. Thus

$$\phi \;=\; \widehat{\omega_1} + \widehat{\omega_2} \;>\; 45^o + 89^o \;=\; 134^o$$

There are two other ways to place an intersecting to ES unit square Σ, whose analysis is similar.

But (a) the initial square (from E) always broadcasts due to the first protocol phase and (b) any intermediate intersecting square will be notified (by induction) and thus will broadcast. Hence, S is always notified, when the whole grid is operational.

■

5.2 The Energy Efficiency and the Robustness of PFR

Consider the fictitious lattice G of the network area and let the event F hold. We have (at least) one particle inside each square. Now join all nearby particles of each particle to it, thus by forming a new graph G' which is "lattice-shaped" but its elementary "boxes" may not be orthogonal and may have varied length. When G's squares become smaller and smaller, then G' will look like G. Thus, for reasons of analytic tractability, we will assume in the sequel that our particles form a lattice (see Fig. 5). We also assume length $l = 1$ in each square, for normalization purposes. Notice however that when $l \to 0$ then "$G' \to G$" and thus all our results in this Section hold for any random deployment "in the limit".

The analysis of the energy efficiency considers particles that are active but are as far as possible from ES. Thus the approximation we do are suitable for remote particles.

Here we estimate an upper bound on the number of particles in an $n \times n$ (i.e. $N = n \times n$) lattice. If k is this number then $r = \frac{k}{n^2}$ $(0 < r \leq 1)$ is the "energy efficiency ratio" of PFR.

We want r to be less than 1 and as small as possible (clearly, $r = 1$ leads to flooding). Since, by Lemma 5.1, ES is always "surrounded" by active particles, we will now assume without loss of generality that ES is part of a horizontal grid line (see Fig. 7) somewhere in the middle of the lattice.

Recall that $|ES| = n_0$ particles of all the active, via PFR, particles that continue to transmit, and let Q be such that its shortest distance from particles in ES be maximum. Then L_Q is the locus (curve) of such points Q. The number of particles included in L_Q (i.e. the *area* inside L_Q) is the number k.

Now, if the distance of such points Q of L_Q from ES is ω then $k \leq 2(n_0 + 2\omega)\omega$ (see Fig. 8) and thus $r \leq \frac{2\omega(n_0+2\omega)}{n^2}$. Notice however that ω is a random variable hence we will estimate its expected value $\mathbb{E}(\omega)$ and the moment $\mathbb{E}(\omega^2)$ to get

$$\mathbb{E}(r) \;\leq\; \frac{4\mathbb{E}(\omega^2)}{n^2} \;+\; \frac{2n_0}{n^2}\,\mathbb{E}(\omega) \tag{1}$$

Now, look at points $Q : (\widehat{EQS}) \;=\; \phi \;<\; 30^o$, i.e. $\phi < \frac{\pi}{6}$. We want to use the approximation $|ED| \simeq x$.

Let $\frac{|ED|}{x} = 1+\epsilon$. Then a bound on the approximation factor ϵ determines a bound on the "cut-off" angle ϕ_0 since $|ED|\cos\left(\frac{\phi_o}{2}\right) = \epsilon x$.

$$(1+\epsilon)\,x\,\cos\left(\frac{\phi_o}{2}\right) \;=\; x \;\Longrightarrow\; \cos\left(\frac{\phi_o}{2}\right) \;=\; \frac{1}{1+\epsilon}$$

We chose $\phi_0 = 30^o$ so $\epsilon = 0.035$. Also, we remark here that the energy spent increases with $x_0 = n_0\left(1 - \frac{\xi}{2}\right)$ in the area below x_0, where $\xi = \tan 75^o$, but decreases with x for $x > x_0$. This is a trade-off and one can carefully estimate the angle ϕ_0 (i.e. ϵ) to minimize the energy spent.

Let $\phi_1 = (\widehat{EQD})$, $\phi_2 = (\widehat{DQS})$ where $QD \perp ES$ and D in ES. In Fig. 9, $\phi = \phi_2 - \phi_1$.

We approximate ϕ by $\sin\phi_2 - \sin\phi_1$ (note: $\phi \simeq \sin\phi \geq -\sin\phi_1 + \sin\phi_2$).

$$\text{Note} \quad \sin\phi_1 \;\simeq\; \frac{|ED|}{|QE|} \;\simeq\; \frac{|ED|}{x} \qquad \text{and} \qquad \sin\phi_2 \;\simeq\; \frac{|DS|}{x}$$

$$\Longrightarrow \quad \sin\phi \;\simeq\; \sin\phi_2 \;-\; \sin\phi_1 \;\simeq\; \frac{|ES|}{x} \;=\; \frac{n_0}{x}$$

Thus $\frac{\phi}{\pi} \;\simeq\; \frac{n_0}{\pi x}$ for such points Q.

Let $\mathcal{M}$ be the stochastic process that represents the vertical (to ES) distance of active points from ES, and $\mathcal{W}$ be the random walk on the vertical line yy' (see fig. 10) such that, when the walk is at distance $x \geq x_0$ from ES then it (a) goes to $x+1$ with probability $\frac{n_0}{\pi x}$ or (b) goes to $x-1$ with probability $1 - \frac{n_0}{\pi x}$ and never goes below x_0, where x_0 is the "30^o-distance" i.e. $x_0 \;=\; \frac{n_0\,\xi}{2}$.

Clearly $\mathcal{W}$ dominates M i.e. $\mathbb{P}_{\mathcal{M}}\Big\{x \;\geq\; x_1\Big\} \;\leq\; \mathbb{P}_{\mathcal{W}}\Big\{x \;\geq\; x_1\Big\}, \forall\, x_1 \;>\; x_0$.

Furthermore, $\mathcal{W}$ is dominated by the continuous time "discouraged arrivals" birth-death process $\mathcal{W}'$ (for $x \geq x_0$) where the rate of going from x to $x+1$ in $\mathcal{W}'$ is $\frac{\alpha}{x} \;=\; \frac{n_0/\pi}{x}$ and the rate of returning to $x-1$ is $1 - \frac{n_0}{\pi x_0} \;=\; 1 - \frac{2}{\pi\xi} \;=\; \beta$.

We know from [9] that for $\Delta x \;=\; x - x_0$ $(x > x_0)$

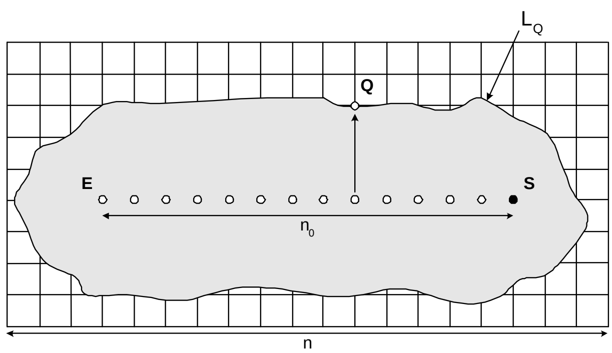

Figure 7: The L_Q Area

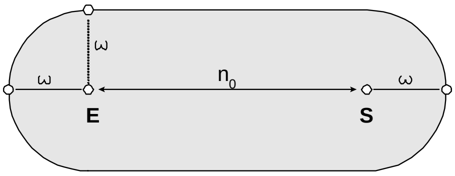

Figure 8: The particles inside the L_Q Area

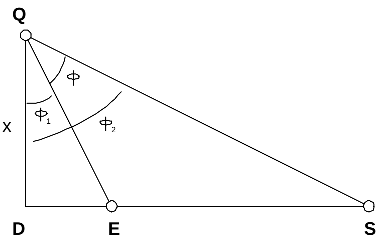

Figure 9: The QES Triangle

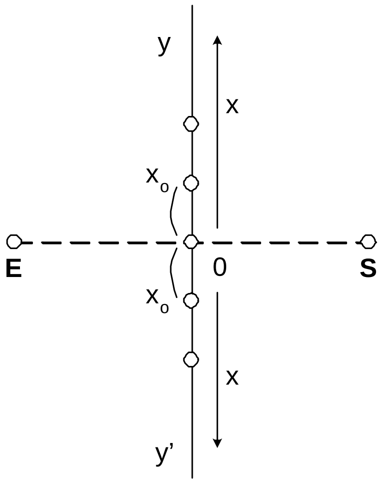

Figure 10: The Random Walk $\mathcal{W}$

$$\mathbb{E}_{\mathcal{W}'}(\Delta x) \;=\; \frac{\alpha}{\beta} \;=\; \frac{n_0}{\pi\left(1 - \frac{2}{\pi\xi}\right)}$$

Thus, $\mathbb{E}_{\mathcal{W}'}(x) \;=\; x_0 + \frac{\alpha}{\beta}$, hence by domination

$$\mathbb{E}(\omega) \;\leq\; \mathbb{E}_{\mathcal{M}}(x) \;\leq\; x_0 \,+\, \frac{\alpha}{\beta} \tag{2}$$

Also from [9] the process $\mathcal{W}'$ is a Poisson one and $\mathbb{P}\left\{\Delta x = k\right\} \;=\; \frac{(\alpha/\beta)^k}{k!}\, e^{-(\alpha/\beta)}$. From this, the variance of Δx is $\sigma^2 \;=\; \alpha/\beta$ (again) i.e. for $\omega \;=\; x_0 + \Delta x$

$$\begin{aligned}\mathbb{E}_{\mathcal{W}'}(\omega^2) \;&=\; \mathbb{E}_{\mathcal{W}'}\Big((x_0 \,+\, \Delta x)^2\Big) = x_0^2 \,+\, 2x_0\mathbb{E}_{\mathcal{W}'}(\Delta x) \,+\, \mathbb{E}_{\mathcal{W}'}(\Delta x^2) \\ &=\; x_0^2 \,+\, 2\frac{\alpha}{\beta}x_0 \,+\, \Big(\sigma^2 \,+ \mathbb{E}^2(\Delta x)\Big) = x_0^2 \,+\, 2\frac{\alpha}{\beta}x_0 \,+\, \frac{\alpha}{\beta} \,+\, \left(\frac{\alpha}{\beta}\right)^2\end{aligned}$$

$$\text{So} \qquad \mathbb{E}_{\mathcal{W}'}(\omega^2) \;=\; \frac{3n_0^2}{4} \;+\; 2\,\frac{\alpha}{\beta}\,\frac{n_0\,\xi}{2} \;+\; \frac{\alpha}{\beta} \;+\; \left(\frac{\alpha}{\beta}\right)^2$$

$$\text{where} \quad \frac{\alpha}{\beta} \;=\; \frac{n_0}{\pi\left(1-\frac{2}{\pi\,\xi}\right)} \;=\; \frac{n_0}{\tau} \qquad \text{and} \quad \tau \;=\; \pi\left(1-\frac{2}{\pi\,\xi}\right) \;=\; \pi-\frac{2}{\xi}$$

$$\text{thus} \qquad \mathbb{E}_{\mathcal{W}'}(\omega^2) \;=\; n_0^2\left(\frac{3}{4}+\frac{\xi}{\tau}+\frac{1}{\tau^2}\right) + \frac{n_0}{\tau} \tag{3}$$

and by domination $\mathbb{E}(\omega^2) \;\leq\; \mathbb{E}_{\mathcal{M}}(\omega^2) \;\leq\; \mathbb{E}_{\mathcal{W}'}(\omega^2)$. So, finally

$$\mathbb{E}(r) \;\leq\; \frac{4\left[n_0^2\left(\frac{3}{4}+\frac{\xi}{\tau}+\frac{1}{\tau^2}\right)+\frac{n_0}{\tau}\right]}{n^2} \;+\; \frac{n_0^2}{n^2}\left(\xi+\frac{2}{\tau}\right) \tag{4}$$

So, we get a very satisfactory result:

Theorem 5.2 The energy efficiency of the PFR protocol is $\Theta\left(\left(\frac{n_0}{n}\right)^2\right)$ where $n_0 = |ES|$ and $n = \sqrt{N}$. For $n_0 = |ES| = o(n)$, this is $o(1)$.

5.3 The Robustness of PFR

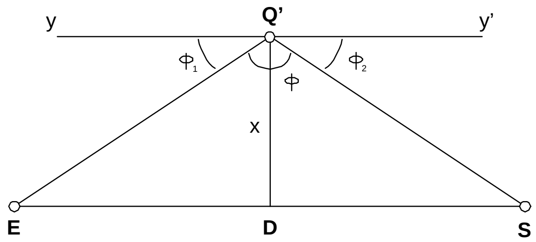

Figure 11: The $Q'ES$ Triangle

Here we consider particles very near the line ES. Thus the approximations that we do are suitable for nearby particles. Let Q' an active particle at vertical distance x from ES and $D \in ES$ such that $Q'D \perp ES$.

Let $yy' \parallel ES$, drawn from Q' (see fig. 11) and $\phi_1 = (\widehat{yQ'E})$, $\phi_2 = (\widehat{SQ'y'})$. Then $\phi = 180^o - (\phi_1+\phi_2)$, i.e. $\frac{\phi}{\pi} = 1 - \frac{\phi_1+\phi_2}{\pi}$.

Since ϕ_1, ϕ_2 are small (for small x) we use the approximation $\phi_1 \simeq \sin\phi_1$ and $\phi_2 \simeq \sin\phi_2$

$$\phi_1 + \phi_2 \;\simeq\; \sin\phi_1 \,+\, \sin\phi_2 \;=\; \frac{x}{EQ'} + \frac{x}{Q'S}$$

Since $\phi > 90^0$, ES is the biggest edge of triangle $(\widehat{EQ'S})$, thus $EQ' \leq n_0$ and $Q'S \leq n_0$. Hence,

$$\sin\phi_1 \,+\, \sin\phi_2 \;\leq\; \frac{2x}{n_0} \qquad \text{and} \qquad 1 - \frac{\sin\phi_1 \,+\, \sin\phi_2}{\pi} \;\geq\; 1 - \frac{2x}{\pi n_0} \qquad (\text{for small } x)$$

Without loss of generality, assume ES is part of a horizontal grid line. Here we study the case in which some of the particles on ES or very near ES (i.e. at angles $> 134^o$) are not operating.

Consider now a horizontal line ℓ in the grid, at distance x from ES. Clearly, some particles of ℓ near E will be activated during the initial broadcasting phase (see Phase 1).

Let $\mathcal{A}$ be the event that all the particles of ℓ (starting from p) will be activated, until a vertical line from S is reached. Then

$$\mathbb{P}(\mathcal{A}) \geq \left(1 - \frac{2x}{\pi n_0}\right)^{n_0} \simeq e^{-\frac{2x}{\pi}}$$

Lemma 5.3 PFR manages to propagate the crucial data across lines parallel to ES, and of constant distance, with *fixed* nonzero probability (not depending on n, $|ES|$).

6 Experimental Findings

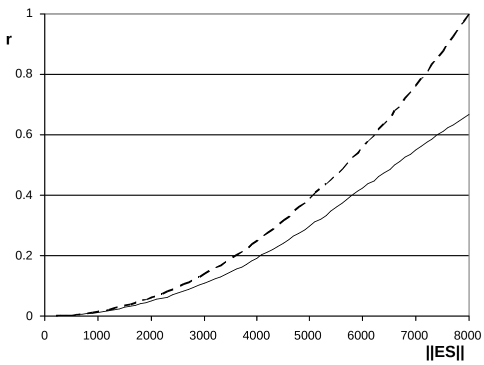

Figure 12: Energy efficiency measure r for different $||ES||$ when $n = 8000$.

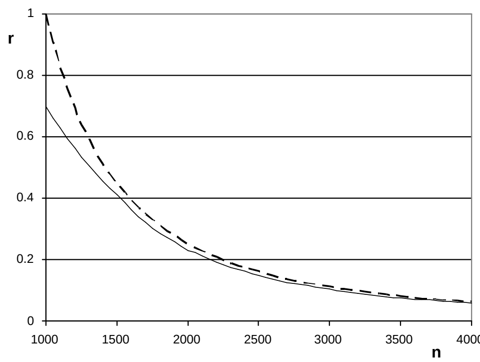

Figure 13: Energy efficiency measure r for different n when $||ES|| = 1000\mathcal{R}$.

We validate the theoretical results and investigate the asymptotic behavior of the PFR protocol by conducting a set of large scale experiments. Our implementation follows closely the protocol description of Section 4 and is based on C++ and the Library of Efficient Data types and Algorithms (LEDA) [10].

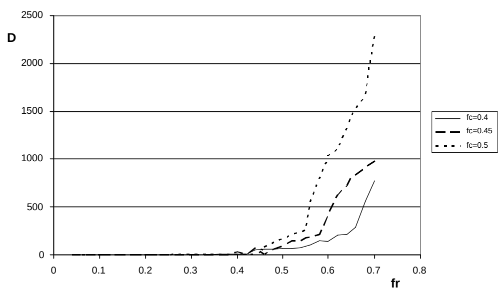

Figure 14: Robustness for $||ES|| = 1000\mathcal{R}$ and $n = 3000$

We start by examining the *energy efficiency* of PFR by measuring the ratio r of the number of activated particles over the total number of particles $r = \frac{k}{n^2}$. We used lattice shaped sensor fields comprised of *extremely large number* of sensors ($n \in [1000, 8000]$) that are put in (horizontal and vertical) distances $\delta = 5m$ apart and each particle's broadcast range $\mathcal{R} = \delta\sqrt{2}$. In each case, we repeated the experiments for at least 100 times to get good average results.

In the first set of experiments, we drop a fixed number of particles ($n = 8000$) and then position $\mathcal{E}$ and $\mathcal{S}$ in a way such that $||ES|| \in [1000\mathcal{R}, 4000\mathcal{R}]$. Figure 12 depicts the ratio r for various $||ES||$.

In the second set of experiments we work in a different way: we drop sensor fields comprised of different number of total particles ($n \in 1000, 4000$) and then position $\mathcal{E}$ to $\mathcal{S}$ so that their distance is always fixed to $||ES|| = 1000\mathcal{R}$. Fig. 13 depicts the ratio r for different n.

In all three figures the dashed line depicts the theoretical result of Theorem 5.2, $r' = \left(\frac{n_0}{n}\right)^2$, where $||ES|| = n_0$.

The above figures indeed validate our analytical results, since in each case the efficiency ratio r exhibits a quadratic behavior with respect to $||ES||$ (in the first case) and n (in the second). Furthermore, notice that our analysis is tight enough, since the curve of the experimental measurements is very close to the curve of the analytic result.

We then investigate the robustness of our protocol, in the case where some particles (permanently) fail. In particular, the failure probability is taken "large" for particles close to ES line (i.e. $f_c = \{0.4, 0.45, 0.5\}$). The particles that are located further away from ES line, fail with smaller failure probability (i.e. $f_r \in [0.01, 0.7]$). The different failure probabilities for "close" and "remote" (to ES line) particles captures the fact that the particles tend to consume high (or low) energy in each case respectively, due to the fact that "close" to ES particles transmit more frequently than "remote" ones. To evaluate the robustness of the protocol we measure in the case of failure to reach the sink, the distance D from it.

Figure 14 shows that our protocol is very robust. In particular, even for large failure probabilities as high as 0.5, the protocol successfully propagates data to the sink. As the failure probability increases (i.e. $f_r > 0.4$) the protocol manages to get close enough to the sink (i.e. in the worst case the final position of the propagation is $1400m$, while $||ES|| = 7000m$). We note that proximity of final propagation position to the sink also increases with failure probability of "close" particles (f_c).

Note that the proximity of the final position to the sink seems to exhibit a certain threshold behavior (in the case of values studied, this threshold is around $f_r = 0.5$). This threshold behavior is probably due to the stochastic process evolution and we intend to also evaluate it analytically.

References

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci: Wireless sensor networks: a survey. In the Journal of Computer Networks, Volume 38, pp. 393-422, 2002.

[2] I. Chatzigiannakis, S. Nikoletseas and P. Spirakis: Smart Dust Protocols for Local Detection and Propagation. In *Proc. 2nd ACM Workshop on Principles of Mobile Computing* – POMC'2002.

[3] I. Chatzigiannakis and S. Nikoletseas: A Sleep-Awake Protocol for Information Propagation in Smart Dust Networks. In Proc. of the 3rd Workshop on Mobile and Ad-hoc Networks (WMAN 2003), IPDPS Workshops 2003.

[4] J. Diaz, J. Petit and M. Serna: Evaluation of basic protocols for optical smart dust networks. In *Proc. 2nd International Workshop on Experimental and Efficient Algorithms* – WEA'2003.

[5] D. Estrin, R. Govindan, J. Heidemann and S. Kumar: Next Century Challenges: Scalable Coordination in Sensor Networks. In *Proc. 5th ACM/IEEE International Conference on Mobile Computing* – MOBICOM'1999.

[6] C. Intanagonwiwat, R. Govindan and D. Estrin: Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks. In *Proc. 6th ACM/IEEE International Conference on Mobile Computing* – MOBICOM'2000.

[7] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann and F. Silva: Directed Diffusion for Wireless Sensor Networking. Extended version of [6].

[8] J.M. Kahn, R.H. Katz and K.S.J. Pister: Next Century Challenges: Mobile Networking for Smart Dust. In *Proc. 5th ACM/IEEE International Conference on Mobile Computing*, pp. 271-278, September 1999.

[9] L. Kleinrock: Queueing Systems, Theory, Vol. I, pp. 100. *John Wiley & Sons*, 1975.

[10] K. Mehlhorn and S. Näher: LEDA: A Platform for Combinatorial and Geometric Computing. *Cambridge University Press*, 1999.