Random Sensory Networks: A Delay Analysis*

Cedric Florens, Masoud Sharif and Robert J. McEliece

Dept. of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125

{florens, masoud, rjm}@systems.caltech.edu

Abstract

A fundamental function performed by a sensory network is the retrieval of data gathered collectively by sensor nodes. The metrics that measure the efficiency of this data collection process are time and energy. In this paper we study via simple discrete mathematical models the statistics of the data collection time in sensory networks. Specifically we analyze the average minimum delay in collecting randomly located/distributed sensors data for networks of various topologies when the number of nodes becomes large. Furthermore we analyze the impact of various parameters such as size of packet, transmission range, and channel erasure probability on the optimal time performance. Our analysis applies to directional antenna systems as well as omnidirectional ones. This paper focuses on directional antenna systems and briefly presents results on omnidirectional antenna systems. Finally a simple comparative analysis shows the respective advantages of the two systems.

1 Introduction

Recent technological advances in the Very Large Scale Integration (VLSI) field have contributed much to the development of Micro sensor systems. These combine various sensors, signal processing, data storage, wireless communication capabilities and energy sources on a single chip. Such

*This work was performed at Caltech and supported by Caltech's Lee Center for Advanced Networking

computational devices are referred to as sensor nodes and a collection of sensor nodes, possibly distributed over a wide area, connected through the wireless medium, form a sensory network. Applications for such networks are numerous and include environmental monitoring (seismic, meteorological) and military surveillance [1]. Sensory networks belong to the family of wireless ad-hoc networks and as such lack an infrastructure present in traditional wireless networks such as cellular networks. In the very near future it is expected that these sensor networks autonomously extract information about their surroundings, performing basic collective processing and transmit the collected data to the end user for further processing and analysis. It should be noted that in a sensory network while each node may be mobile, it is typically the case that once the target site of the particular sensing application is reached a semi-permanent stationary configuration is adopted for the purpose of gathering information.

In the field of general ad-hoc networks and particularly sensory networks, research efforts focusing on design issues of the network communication architecture have been widespread. A detailed investigation of current protocol and algorithm proposals in the physical, data link, network, transport and application layers are discussed for example in [2]. Technical issues and applications requirements to be dealt with by these protocols are multiple and often specific to the class of sensory networks. Among those, efficient management of energy budget is of paramount importance to the lifetime of the networks. Furthermore, depending on the application under consideration a trade-off between data collection delay and energy consumption has to be achieved. Finally the throughput of a sensory network is an important characteristic measure which is closely related to the delay of the data collection process. Theoretical results regarding capacity of general static ad-hoc networks has appeared in [3], [4], [5]. Also relevant to our research is the so called packet routing problem which consists in moving packets of data from one location to another as quickly as possible in a network and has been studied in conjunction with wireline and wireless network models (see for example [6], [7], [8] and [9]). In [10] the authors studied the problem in sensory networks of collecting sensors data at the network base station. They describe optimal strategies to perform data collection under various assumptions and derive corresponding time performances with respect to a simple discrete mathematical model for a sensor network. In this model the amount of data accumulated at each sensor node (characterized by a number of unit data packets) after some given observation period is assumed finite and determined. In typical scenarios however the exact amount of data accumulated at each sensor node is unknown.

In this paper we model the number of data packets as a random variable and analyze the delay (which is now a random variable) in collecting sensor data at the base station. More specifically, we derive the distribution and the expected value of the delay for a line network using the optimal scheduling. Furthermore, we look into the effect of various parameters including size of packet, transmission range, and channel erasure probability. We also propose a simple scheduling and anlayse its delay performance. Finally we extend our result to more general topologies such as multi-line networks.

This paper is organized as follows: In section 2 we present our sensory network model and recap results from [10] that will be used in the remainder of the paper. We present results on a line network in section 3. In section 4, we expose our derivation regarding multiline networks. We present general results on trees, and general graphs in section 5. Finally we give comparative results between directional and omnidirectional antenna systems in section 6. We conclude the paper with some simulations in section 7.

A node (BS included) cannot receive and transmit at the same time

2 Model and Problem Statement

In this section, we describe the sensor network model on which the subsequent analysis is based and formulate our problem within the framework of this model. Furthermore we briefly review results in [10] that are relevant to this study. As noted in the introduction, in most sensing applications sensor nodes adopt a stationary configuration while information is being gathered. Correspondingly, our models will be static. In stationary state, after the nodes have organized themselves into a network, we distinguish between two phases of operation. In the first phase or observation phase, area monitoring results in an accumulation of data at each sensor node. In the second phase or data transfer, the collected data is transmitted to some processing center located within the sensor network (we refer to this node as the base station (BS) of the sensor network). In this paper we investigate the efficiency limits with respect to time of such data transfers.

We define a sensor network as a collection of n identical nodes $\{N_1, ..., N_n\}$. Each node N_i is associated with an integer ν_i that represents the number of data packets collected by this node during the observation phase. N_0 denotes the BS which is located within the network. Nodes (BS included) have limited wireless communications capabilities and cannot receive and transmit at the same time. All the nodes including the base station have a common transmission range r and interference range r'. The interference model as defined in [3] for omnidirectional antenna systems is adopted here. That is, a transmission from node N_i to node N_j where $i, j \geq 0$ is successful if for every other node N_k, $k \geq 0$ simultaneously transmitting:

$$|N_i - N_j| \leq r,\ |N_k - N_j| \geq (1 + \delta)r,\ \delta > 0 \tag{1}$$

The second inequality specifies that node N_j must be outside the interference range of node N_k and defines the interference region of node N_k as the disc of radius $r(1+\delta)$ centered at N_k. In directional antenna systems, on the other hand, the interference region of node N_k is only a portion of that disc, the sector formed by some angle θ. Fig. 1 illustrates the characteristic parameters of the model: sensor nodes $N_1, ..., N_6$, the transmission range r and the interference range $r' = r(1+\delta)$. In directional antenna systems a transmission from N_1 to N_2 creates interference at node N_6 (inside the sector formed by θ). However the same transmission creates interference at nodes N_6, N_3 and is received by node N_4 (which is interference from the point of view of N_4) in omnidirectional antenna systems.

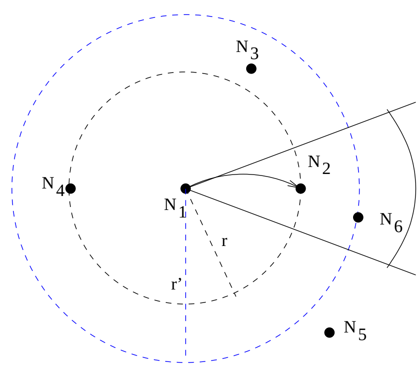

Fig. 1: 6-sensor network. Interference model parameters.

We assume in our model that time is slotted and a one-hop transmission consumes one time slot (TS). The network is further assumed to be synchronous. A node can only transmit/receive one data packet per time slot. Multiple transmissions may occur within the network in one TS under this interference model by virtue of spatial separation. Such a network may be represented by a weighted rooted graph $\{V, E, \boldsymbol{\nu_n}\}$ where $V = \{N_0, ..., N_n\}$, E denotes the set of links and $\boldsymbol{\nu_n} = (\nu_1, ..., \nu_n)$. In this graph model the root represents the BS (N_0) and an edge represents an existing wireless connection between two sensor nodes, or a sensor node and the BS. The *data collection problem* in a given sensory network is defined as the problem of routing all the data collected by the sensor nodes to the BS as efficiently as possible with respect to time and energy. The *data distribution problem*, on the other hand, is the problem of routing data to sensor nodes in a timely and energy efficient manner. In the following work we shall focus on the time efficiency alone of the data collection and distribution tasks.

In [11], an optimal strategy is proposed to minimize the data collection time when the transmission range is a single hop. Moreover, it is proved that for a one-sided line network of length[1] n in which the i'th node has ν_i packets and is equipped with directional antennas, the minimum collection time of the packets at the BS, achieved by the proposed optimal strategy and denoted by $T_{min}(\boldsymbol{\nu_n})$, is:

$$T_{min}(\boldsymbol{\nu_n}) = \max_{1 \leq i \leq n-1} (i - 1 + \nu_i + 2 \sum_{j \geq i+1}^{n} \nu_j) \tag{2}$$

where $\boldsymbol{\nu_n} = (\nu_1, \ldots, \nu_n)$. Furthermore it is proved that the distribution and collection problems are essentially the same and that the minimum data distribution time is the same as the minimum data collection time. We illustrate the optimal schedules on the following example where $V = \{0, 1, 2, 3, 4, 5, 6, 7\}$, $E = \{(i, i + 1), 0 \leq i \leq 6\}$, $\boldsymbol{\nu} = (2, 0, 0, 0, 3, 0, 1)$, $d < r < 2d$, $(1 + \delta)r < 2d$, where it is assumed that nodes are distance d apart from each other. In this example there are 6 packets to be collected (or distributed). The network is shown in Fig. 2. The schedules of transmissions are drawn below the network for the distribution and collection tasks respectively. Arrows represent a single data packet transmission from a node to its neighbor. Either way it is performed in 11 TS. In the distribution case the BS strategy is as follows: send first data packets destined for the furthest node, then data packets for the second furthest one and so on, as fast as possible while respecting the channel reuse constraints. Nodes between the BS and its destinations are required to forward packets as soon as they arrive (that is in the TS following their arrival). For example at TS 1, the packet destined for node 7 is transmitted by the BS to node 1, at TS 2 from node 1 to node 2 and so on and arrives at TS 7. Note that the optimum collection schedule is obtained from the distribution schedule by simple symmetry as shown in the figure and vice versa.

[1]By a line network, we mean one-sided line network. A line network where the base station is in the middle of the line, can be seen as a two-line network. Results for the case where we have more than one line are discussed in Section 4.

This is always true and therefore time performance results obtained in the rest of this paper will apply to both problems (i.e. collection/distribution), unless otherwise specified.

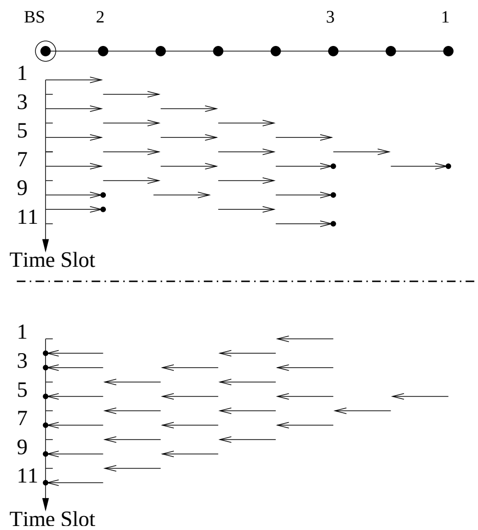

Fig. 2: 8-node line network ($\nu_1 = 2$, $\nu_2 = \nu_3 = \nu_4 = \nu_6 = 0$, $\nu_5 = 3$, $\nu_7 = 1$) followed by optimal transmission schedules for the distribution (upper schedule) and collection (lower schedule) problems. They are symmetric of one another. The job is performed in 11 TS.

3 Random Line Networks

In this section, we characterize the delay in collecting random amount of data randomly spread over a sensor network after the observation phase. More specifically, for a one-sided line network, we first derive a recursion to compute the probability distribution function of $T_{min}(\boldsymbol{\nu_n})$ and also we asymptotically analyze the average of $T_{min}(\boldsymbol{\nu_n})$ when n is sufficiently large.

We further look into the delay when each node is allowed to transmit over $h > 1$ hops and also the effect of packet splitting on the delay in subsections 3. 3 and 3. 4. In section 3. 5, we propose a simple scheme that does not use the knowledge of the number of packets at other nodes and achieves the same scaling law for the average delay. Finally in the last subsection, we consider the effect of error in the channel on the delay.

3.1 The Distribution of the Delay

In this section we derive, by means of a recursion, the cumulative distribution function (CDF) of $T(\boldsymbol{\nu_n})$ for a line network. Let's assume that ν_i corresponds to the number of packets at node i for $i = 1, \ldots, n$ and also ν_i's are i.i.d. random variables chosen from the set $S_m = \{0, 1, \ldots, m-1\}$.

Theorem 1. *Let $F_n(t)$ be the CDF of the minimum delay $T_{min}(\boldsymbol{\nu_n})$, i.e. $F_n(t) = \Pr\{T_{min}(\boldsymbol{\nu_n}) \leq t\}$. Then $F_n(t)$ satisfies the following recursion*

$$F_n(t) = \sum_{i=0}^{m-1} \Pr(\nu_n = i) F_{n-1}(t - 2i) \mathbf{1}_{t \geq n+2(i-1)} + \Pr(\nu_n = 0) F_{n-1}(t) \quad \textit{for } n \geq 2 \tag{3}$$

$$\textit{where } \mathbf{1}_{t \geq t_0} = \begin{cases} 1 & \textit{if} \quad t \geq t_0 \\ 0 & \textit{otherwise.} \end{cases} \quad \textit{and } F_1(t) = \begin{cases} \sum_{i=0}^{t} \Pr(\nu_1 = i) & \textit{if } t < m-1 \\ 1 & \textit{otherwise} \end{cases}$$

Proof. We may write $F_n(t)$ by conditioning on $\nu_n = i$ for $i = 0, \ldots, m-1$ as

$$F_n(t) = \sum_{i=0}^{m-1} \Pr\{T_{min}(\boldsymbol{\nu_n}) \leq t | \nu_n = i\} \Pr(\nu_n = i) \tag{4}$$

To compute the conditional probability in (4), we use (2) and the fact that for all $k = 1, \ldots, n-1$, and $i \geq 1$, $T_{min}(\boldsymbol{\nu_n}) \geq k - 1 + \nu_k + 2\sum_{j=k+1}^{n} \nu_j$. Therefore replacing $k = n-1$ and assuming $\nu_n = i$, we get

$$T_{min}(\boldsymbol{\nu_n}) \geq n - 2 + \nu_{n-1} + 2\nu_n \geq n + 2(i-1) \tag{5}$$

Thus if $t < n + 2(i-1)$, then $\Pr\{T_{min}(\boldsymbol{\nu_n}) \leq t | \nu_n = i\} = 0$. Using the definition of the function $\mathbf{1}_{t \geq t_0}$, for any $i \geq 1$ we may then write the conditional probability as

$$\Pr\{T_{min}(\boldsymbol{\nu_n}) \leq t | \nu_n = i\} = \Pr\{T_{min}(\boldsymbol{\nu_{n-1}}) \leq t - 2i\} \mathbf{1}_{t \geq n+2(i-1)} \tag{6}$$

Replacing (6) in (4), we get

$$F_n(t) = F_{n-1}(t)\Pr(\nu_n = 0) + \sum_{i\geq 1}^{m-1} \Pr\{T_{min}(\boldsymbol{\nu_{n-1}}) \leq t - 2i)\}\mathbf{1}_{t\geq n+2(i-1)} \Pr(\nu_n = i)$$

which leads to (3). □

We can use the result of Theorem 1 to compute the CDF of $T_{min}(\boldsymbol{\nu_n})$. This is illustrated in Fig. 3 and Fig. 4. Fig. 3 shows the distribution of the delay $T_{min}(\boldsymbol{\nu_n})$ for 40-sensor node line networks in which each node carries either 0 or 1 packet with probability 1/2. Fig. 4 shows the distribution of the delay $T_{min}(\boldsymbol{\nu_n})$ for 40-sensor node line networks in which each node carries either 0 or 1 packet with probability 0.8 and 0.2 respectively.

It is also worth noting that the result of Theorem 1 holds for any distribution of the data packets. In particular the ν_i's need not be i.i.d., however, in this paper we deal with the case that ν_i's are independent and identically distributed.

Interestingly, if we plot the expected value of T_{min} as in Fig. 6, we observe that the average delay scales linearly with the number of nodes n and the linear factor depends on the average number of packets per node μ. In the next section, we analyze the average delay and prove the observation rigorously.

3.2 Asymptotic Analysis of the Average Delay

In this subsection, we study the asymptotic behavior of the minimum average delay in collecting data from a line network as the number of nodes becomes large.

Theorem 2. *Let ν_i's be i.i.d. random variables $\nu_i \in S_m$ with mean μ, variance σ^2 where μ, σ^2, m*

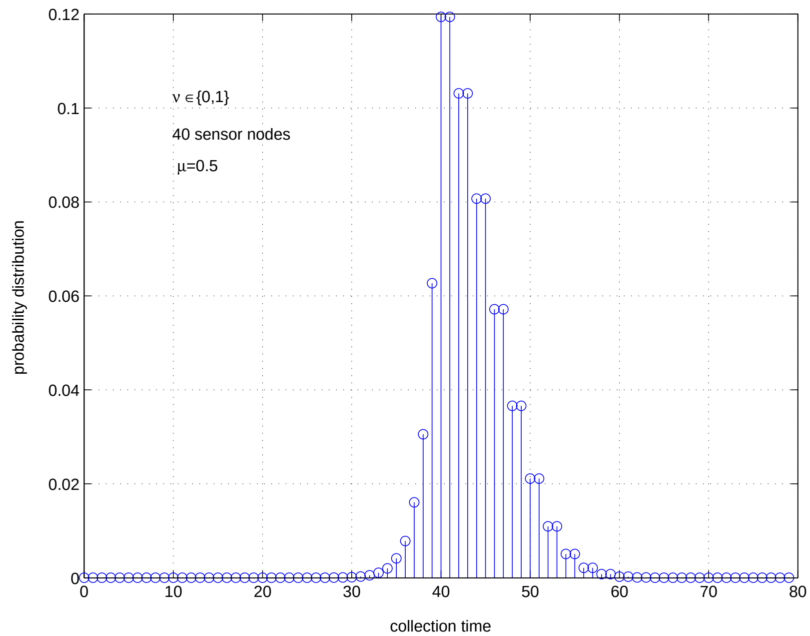

Fig. 3: Distribution of data collection time in 40-node line network. Each node in the particular type of network considered carries 0 or 1 data packet with probability 1/2.

are all constants independent of n. We have

$$\lim_{n\to\infty} \frac{\mathbb{E}\{T_u\}}{n} = \begin{cases} 2\mu & \textit{if } \mu \geq 1/2 \\ \\ 1 & \textit{if } \mu \leq 1/2 \end{cases} \tag{7}$$

Proof. We consider the case $\mu \geq 1/2$ first: Let's define $\nu_i' = \nu_i - \mu$. Using (2), we get,

$$\begin{aligned} \mathbb{E}\{T(\nu_{\mathbf{n}})\} &= 2\mu n + \mathbb{E}\left\{ \max_{1\leq i\leq n-1} \left(i(1-2\mu) + \nu_i' + 2\sum_{i+1}^{n} \nu_j' \right) \right\} \\ &\leq 2\mu n + 2\mu - 1 + 2\mathbb{E}\left\{ \max_{1\leq i\leq n} \sum_{j\geq i}^{n} \nu_j' \right\} \\ &= 2\mu n + 2\mu - 1 + 2\mathbb{E}\left\{ \max_{1\leq i\leq n} \sum_{j=1}^{n+1-i} \nu_{n-j+1}' \right\} \end{aligned} \tag{8}$$

where the inequality follows from the fact that ν_i' satisfies $\nu_i' + \mu \geq 0$, $1 \leq i \leq n-1$. In order to

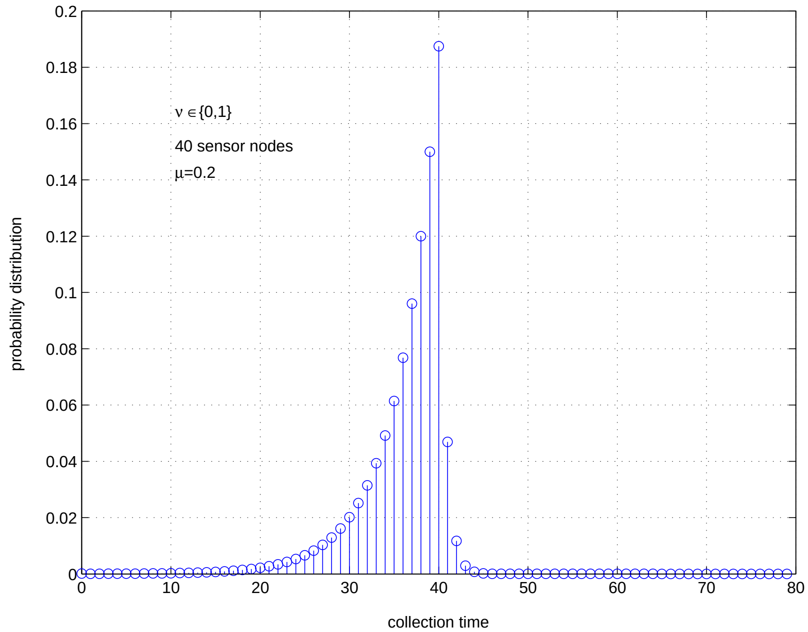

Fig. 4: Distribution of data collection time in 40-node line network. Each node in the considered network carries 0 or 1 data packet with probability 0.8 and 0.2 respectively.

find a bound for $\mathbb{E}(\max\limits_{1\leq i\leq n} \sum_{j\geq i}^{n} \nu'_j)$, we first state the following lemma which is based on a result of Erdos and Kac [12] where a convergence theorem for the distribution of the maximum of partial sums was proven.

Lemma 3. *For any* λ *and* $a > 1$*,*

$$\Pr\left\{\max_{1\leq i\leq n} \geq \lambda\sigma\sqrt{n}\right\} \leq \frac{a-1}{a}\Pr\left\{\sum_{j=1}^{n} \nu'_j \geq (\lambda - \sqrt{a})\sigma\sqrt{n}\right\}. \tag{9}$$

where $\nu'_i = \nu_i - \mu$ *and* ν_i *is as defined in Theorem 2.*

Proof. We first define $S_i = \sum_{j\geq i} \nu'_j$ and the events E_i as,

$$E_i = \left\{\max_{0\leq j<i} S_j \leq \lambda\sigma\sqrt{n} \leq S_i\right\} \quad i = 1, \ldots, n. \tag{10}$$

which is inspired by [12]. We can then state the following inequality by the union bound,

$$\Pr\left\{\max_{1\le i\le n} S_i \ge \lambda\sigma\sqrt{n}\right\} \le \Pr\left\{S_n > (\lambda-\sqrt{a})\sigma\sqrt{n}\right\} + \sum_{i=1}^{n} \Pr\left\{E_i \cap \left(S_n \le (\lambda-\sqrt{a})\sigma\sqrt{n}\right)\right\} \quad (11)$$

To evaluate the second term in the right hand side of (11), we note that $S_i \ge \lambda\sigma\sqrt{n}$ and $S_n \le (\lambda-\sqrt{a})\sigma\sqrt{n}$ imply $S_i - S_n \ge \sqrt{a}\sigma\sqrt{n}$. Then using the fact that $S_i - S_n$ is independent of S_j for $j \le i$, we may write,

$$\begin{aligned}
\sum_{i=1}^{n} \Pr\left\{E_i \cap \left(S_n \le (\lambda-\sqrt{a})\sigma\sqrt{n}\right)\right\} &\le \sum_{i=1}^{n} \Pr(E_i)\Pr\left(S_i - S_n \ge \sqrt{a}\sigma\sqrt{n}\right) \\
&\le \sum_{i=1}^{n} \Pr(E_i)\frac{E\left\{(S_i-S_n)^2\right\}}{a\sigma^2 n} \\
&= \sum_{i=1}^{n} \Pr(E_i)\frac{(n-i)\sigma^2}{a\sigma^2 n} \\
&\le \frac{1}{a}\sum_{i=1}^{n} \Pr(E_i) \\
&\le \frac{1}{a}\Pr\left(\max_{1\le i\le n} S_i \ge \lambda\sigma\sqrt{n}\right) \qquad (12)
\end{aligned}$$

where in the second inequality we used Chebychev's inequality and the last inequality follows from the definition of the events E_i and noting that

$$\sum_{i=1}^{n} \Pr(E_i) = \Pr\left(\cup_{i=1}^{n} E_i\right) = \Pr\left(\max_{1\le i\le n} S_i \ge \lambda\sigma\sqrt{n}\right)$$

since the events E_i are disjoint events. Therefore Lemma 3 follows from (12) and (11). □

Now we can use Chebychev's inequality to evaluate the right hand side of Lemma 3 as follows,

$$\Pr\left\{S_n = \sum_{i=1}^{n} \nu_i' \ge (\lambda-\sqrt{a})\sigma\sqrt{n}\right\} \le \frac{n\sigma^2}{(\lambda-\sqrt{a})^2\sigma^2 n} \le \frac{1}{(\lambda-\sqrt{a})^2}.$$

Therefore, substituting $\lambda = \log n$ we get,

$$\Pr\left(\max_{1\le i\le n}\sum_{j=i}^{n}\nu_j^{'} \ge \sigma\log n\sqrt{n}\right) = O\left(\frac{1}{\log^2 n}\right). \tag{13}$$

Eq. (13) implies that, with high probability $\max\limits_{1\le i\le n}\sum_{j\ge i}\nu_j'$ is less than $\sigma\log n\sqrt{n}$. Therefore, we may write:

$$\begin{aligned}\mathbb{E}\left\{\max_{1\le i\le n-1}\sum_{j=i}^{n}\nu_j'\right\} &\le \sigma\log n\sqrt{n}\Pr\left\{\max_{1\le i\le n-1}\sum_{j=i}^{n}\nu_j' < \sigma\log n\sqrt{n}\right\} + \\ &\quad (m-1-\mu)n\Pr\left\{\max_{1\le i\le n-1}\sum_{j\ge i}^{n}\nu_j' > \sigma\log n\sqrt{n}\right\} \\ &= \sigma\log n\sqrt{n} + O\left(\frac{n}{\log^2 n}\right)\end{aligned} \tag{14}$$

which follows from the fact that $\nu_i' \le m-1-\mu$.

We now derive a lower bound on $\mathbb{E}(T_{min}(\boldsymbol{\nu_n}))$: From Eq. (2), we get $T_{min}(\boldsymbol{\nu_n}) \ge \nu_1 + 2\sum_{j\ge 2}^{n}\nu_j$.

Taking the expectation of both sides, we get:

$$\mathbb{E}(T_{min}(\boldsymbol{\nu_n})) \ge 2\mu n - \mu \tag{15}$$

Considering (15) and the upper bound derived in (14), we deduce that

$$2\mu n - \mu \le \mathbb{E}(T(\nu_{\mathbf{n}})) \le 2\mu n + 2\mu - 1 + 2\sigma\log n\sqrt{n} + O\left(\frac{n}{\log^2 n}\right)$$

which leads to (7) for $\mu \ge 1/2$.

Next we consider the case $\mu \leq 1/2$: Let's define $\nu'_i = \nu_i - 1/2$. Using (2), we get,

$$\begin{aligned} T_{min}(\nu_{\mathbf{n}}) &= \max_{1 \leq i \leq n-1} \left(n - \frac{1}{2} + \nu'_i + 2 \sum_{i+1}^{n} \nu'_j \right) \\ &\leq n - \frac{1}{2} + 2 \max_{1 \leq i \leq n-1} \sum_{i}^{n} \nu'_j \end{aligned}$$

Taking the expectation of both sides and using inequality (14) we get:

$$\mathbb{E}(T_{min}(\boldsymbol{\nu_n})) \leq n + 2\sigma \log n \sqrt{n} + O\left(\frac{n}{\log^2 n} \right) \tag{16}$$

On the other hand, it is clear that if there is any packet in distance r, it takes at least r TS to be collected. Furthermore the probability that there are no packets in the last $\log n$ nodes of the line network is $1 - (\Pr(\nu_i = 0))^{\log n}$. Therefore, noting that $\Pr(\nu_i = 0)$ is a fixed number, we may write,

$$\mathbb{E}(T_{min}(\boldsymbol{\nu_n})) \geq (n - \log n)(1 - (\Pr(\nu_i = 0))^{\log n}) = n - O(\log n), \tag{17}$$

which leads to (7) for $\mu \leq 1/2$.

□

Remark: Theorem 2 can be easily generalized to the case that ν_i's are independent and have mean $\mu_i \geq \frac{1}{2}$ and variance σ_i^2 and $\nu_i \leq m - 1$ where m is a constant. In fact we can assume m is also going to infinity as well. Considering Eq. (14), the theorem goes through as long as $m = o(n)$.

Fig. 5 shows the ratio of the average delay to the number of sensor nodes, i.e. $\mathbb{E}(T_{min}(\boldsymbol{\nu_n}))/n$, for a line network where each sensor node carries 0 or 1 data packet with probabilities $1 - \mu$ and μ respectively as a function of the number of sensor nodes n in the network and the average number of packets per node μ. Fig. 6 shows the ratio of the average delay to the number of sensor nodes in a line network (where again each node carries either 0 or 1 packet with probabilities $1-\mu$ and μ respectively) for a fixed number of sensor nodes (500) as a function of the average number of packets per node μ.

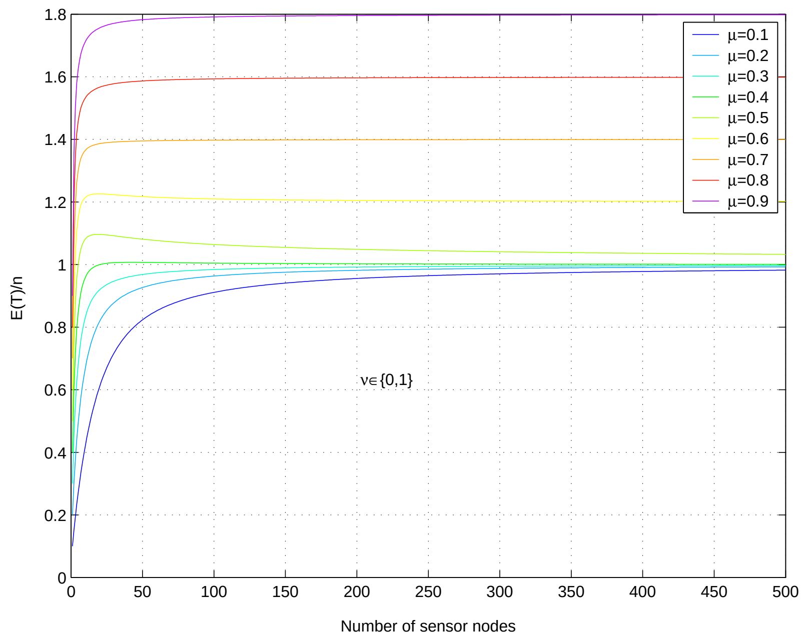

Fig. 5: Average collection time as a function of average number of packets per node and number of nodes in line network. Nodes carry 0 or 1 data packet with probability $1-\mu$ and μ respectively.

3.3 Multihop Case

In this section, we consider the problem of scheduling when each node is allowed to use up to h hops. Of course, a longer transmission range leads to faster data collection. This is quantified in the following theorem where the minimum data collection time $T_{min}(h, \boldsymbol{\nu_n})$ is expressed as a function of the transmission range h (hops). This Theorem is basically a generalization of the result of Eq. (2) where $h = 1$.

Theorem 4. *For a one-sided line network of length n in which the* ith *node has ν_i packets and is equipped with directional antennas, the minimum collection time of the packets at the BS as a*

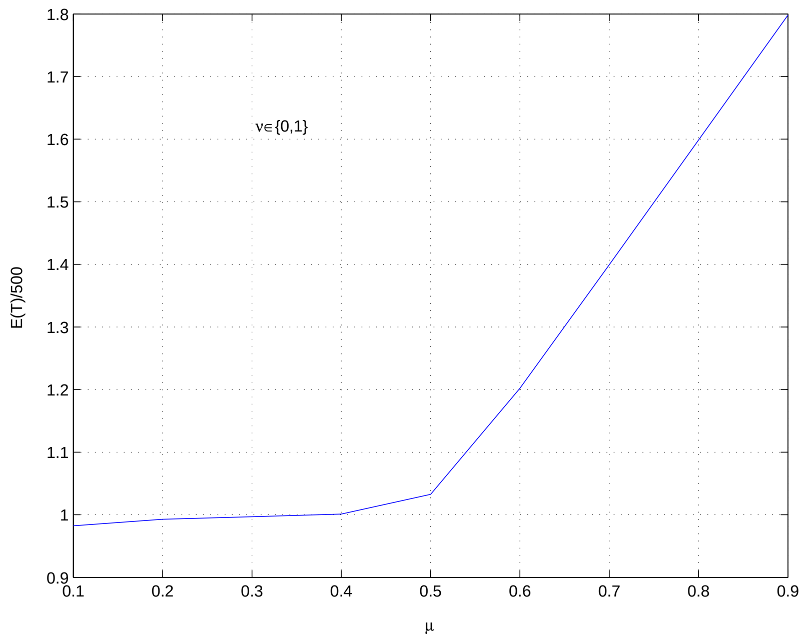

Fig. 6: Average collection time as a function of average number of packets per node in 500-node line network. Nodes carry 0 or 1 data packet with probability $1-\mu$ and μ respectively.

function of the transmission range h in hops is:

$$T_{min}(h, \boldsymbol{\nu_n}) = \max(S', S_{h+1}, S_{h+2}, \ldots, S_{n-h}) \tag{18}$$

where

$$\begin{aligned} S_i &= \sum_{j>i+h}^{n} \nu_j + \left\lfloor \frac{\sum_{j>i+h}^{n} \nu_j - 1 + (i \bmod h)}{h} \right\rfloor + \left\lfloor \frac{i}{h} \right\rfloor + 1,\ 0 \le i \le n-h \\ S' &= S_0 + \max\left(\sum_{j=1}^{l} \nu_j - 1, 0 \right) + \sum_{j=l+1}^{h} \nu_j \end{aligned} \tag{19}$$

where l is the unique solution to $l + n_0 = 0 \mod h$ such that $0 \le l \le h-1$.

Remark: Note that when $h = 1$, Eq. (18) reduces to the familiar Eq. (2) proved in [10].

Proof. This theorem was proven in [10] when $h = 1$. Here we only outline the generalization.

The proof has two parts. First we need to show that the RHS of (18) is a lower bound for the collection time. Second we prove it is an upper bound as well by exibiting a schedule with this time performance.

In order to show that the right hand side is a lower bound, we first consider the h nodes $i, 1 \leq i \leq h$ closest to the BS. They need to forward $n_h = \sum_{j>h} \nu_j$ packets. If $n_h \leq h$, this can be done in $n_h + 1$ TS or more. This takes exactly $n_h + 1$ TS if all packets to be distributed are located at node $h + 1$ and more otherwise. If $h + 1 \leq n_h \leq 2h$, this can be done in $n_h + 2$ TS or more. So in general it takes at least $n_h + \lfloor \frac{n_h - 1}{h} \rfloor + 1$ TS. More generally if $n_{i,h}$ denotes the number of packets to be forwarded by the h nodes $j, i + 1 \leq j \leq i + h$, it can be shown that it takes at least $n_{i,h} + \lfloor \frac{n_{i,h} + (n_{i,h} \mod h) - 1}{h} \rfloor + \lfloor \frac{i}{h} \rfloor + 1$ TS to do so. Therefore the maximum of the previous expression over i gives a lower bound for the data collection time performance. We are not done though. Indeed this lower bound is not achievable when there are packets to be distributed at distance i where $i, 1 \leq i \leq h$. An additional lower bound may be derived to handle this case by reconsidering the first h nodes. They must not only forward $\sum_{j>h} \nu_j$ packets, but also receive $\sum_{j \leq h} \nu_j$ packets. The lower bound S_0 may be adjusted (to S') to take this fact into account.

A possible (optimal) schedule for the distribution problem is as follows: It consists of transmitting data packets first to the furthest node, then to the second furthest node and so on as fast as possible until all packets at distance greater than h have been served. Packets at distance $i, 1 \leq i \leq h$ are served in the reversed order, i.e, from closest to the BS to furthest. To prove this is indeed optimal we compute its time performance and shows it achieves the lower bound previously exhibited. This is similar to what was done in [10] and is left out here for the sake of brevity. □

In order to get a better insight into the result of Theorem 4, we give a simple illustrative example before obtaining the asymptotic behavior of the expected minimum delay as n approaches infinity in the next theorem. Theorem 5, in fact, quantifies the dependency between the minimum collection time and the transmission range.

Example: We consider a line network of length n, where each node carries exactly one data packet and has a transmission range of $h \leq n$ hops. Direct application of Theorem 4 gives the minimum collection time as:

$$T = n + \lfloor \frac{n}{h} \rfloor - 1 \tag{20}$$

Fig. 7 shows an instance of this network: $n = 10$ and $h = 3$. Hence the data collection time is 12TS. The associated distribution schedule accompanies the figure.

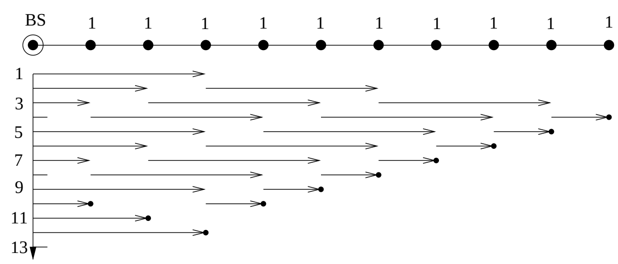

Fig. 7: Minimum length data distribution schedule of a 10-node line network when maximum transmission range is 3 hops.

Theorem 5. *Let h be the transmission range, let ν_i's be i.i.d. random variables $\nu_i \in \{0, 1, ..., m-1\}$ with mean μ and variance σ^2 where h, m, μ, σ^2 are constants independent of n.*

$$\lim_{n\to\infty} \frac{\mathbb{E}\{T_{min}(h, \boldsymbol{\nu_n})\}}{n} = \begin{cases} (1 + \frac{1}{h})\mu & \text{if } \mu \geq \frac{1}{h+1} \\ 1 + \frac{1}{h} & \text{if } \mu \leq \frac{1}{h+1} \end{cases} \tag{21}$$

Proof. The Theorem follows by using the same machinery as in the proof of Theorem 2 and we omit the proof for the sake of brevity. □

We can now evaluate the gain in increasing the transmission range of a sensor node. Theorem

5 shows that a maximum gain of 2 on the collection time may be obtained by increasing the transmission range (in the limit when h approaches infinity) from $h = 1$. One should note however that this gain necessitates a significant amount of energy, in fact in the order of $O(\sum_i i^2 \nu_i) = O(n^3)$ if the energy expanded is taken to be proportional to the square of the distance traveled by a packet, whereas the minimum energy expanded (case $h = 1$) is of the order $O(\sum_i i\nu_i) = O(n^2)$.

3.4 Packet Splitting to Improve the Average Delay

As Eq. (7) implies, if the network is under-loaded (i.e., $\mu \leq \frac{1}{2}$), the ratio of the expected collection time to the expected number of packets in the network is $\frac{1}{\mu}$ and is rather high. One approach to decrease this ratio for small μ is to artificially increase the expected number of packets at each node by splitting each packet into k packets with length $\frac{1}{k}$ times of the original one. Clearly, this increases μ by a factor of k, and therefore, can potentially decrease the delay. It is also worth noting that the time needed to send the smaller size packets is $\frac{1}{k}$ of the time to send the original packets.

In this section we examine the potential gain obtained by splitting data packets into sub-packets. As a first step, we prove that the delay is a decreasing function of k in the next theorem.

Theorem 6. *Given a line network $\boldsymbol{\nu_n}$ there is a gain $k \geq G(\boldsymbol{\nu_n}, k) \geq 1$ in splitting the data packets into k sub-packets. Furthermore $G(\boldsymbol{\nu_n}, k)$ is a non-decreasing function of k and the maximum achievable gain is:*

$$G_{max}(\boldsymbol{\nu_n}) = \lim_k G(\boldsymbol{\nu_n}, k) = \frac{\max\limits_{1 \leq i \leq n-1} (i - 1 + \nu_i + 2\sum_{j \geq i+1}^{n} \nu_j)}{\nu_1 + 2\sum_{j>1}^{n} \nu_j} \tag{22}$$

Proof. In general if each item is split into k sub-items, the gain $G(\boldsymbol{\nu_n}, k)$ satisfies:

$$G(\boldsymbol{\nu_n}, k) = \frac{k \max\limits_{1 \le i \le n} (i - 1 + \nu_i + 2\sum_{j>i}^{n} \nu_j)}{\max\limits_{1 \le i \le n} (i - 1 + k\nu_i + 2k\sum_{j \ge i+1}^{n} \nu_j)} = \frac{\max\limits_{1 \le i \le n} (k(i-1) + k\nu_i + 2k\sum_{j>i}^{n} \nu_j)}{\max\limits_{1 \le i \le n} (i - 1 + k\nu_i + 2k\sum_{j \ge i+1}^{n} \nu_j)} \tag{23}$$

It is easy to check that $1 \le G(\boldsymbol{\nu_n}, k) \le k$. Furthermore $G(\boldsymbol{\nu_n}, k)$ is a non-decreasing function of k. Indeed, if $k_1 \ge k_2$, we can write,

$$\max_{1 \le i \le n} (k_1(i-1) + k_1 k_2 \nu_i + 2k_1 k_2 \sum_{j>i}^{n} \nu_j) \ge \max_{1 \le i \le n} (k_2(i-1) + k_1 k_2 \nu_i + 2k_1 k_2 \sum_{j>i}^{n} \nu_j)$$

which implies that $G(\boldsymbol{\nu_n}, k_1) \ge G(\boldsymbol{\nu_n}, k_2)$. The limit in (22) can be also easily shown using (23). □

Next we derive the average collection time in random sensor network in the limit when n goes to infinity and when packets have been split into k sub-packets.

Theorem 7. *Let ν_i's be i.i.d. random variables $\nu_i \in S_m$ with mean μ, variance σ^2 where μ, σ^2, m are all constants independent of n. If each packet is split into k sub-packets we have:*

$$\lim_{n \to \infty} \frac{\mathbb{E}\{T_{min}\}}{n} = \begin{cases} 2\mu & \text{if } \mu \ge 1/2k \\ 1/k & \text{if } \mu \le 1/2k \end{cases} \tag{24}$$

Proof. The proof falls along the same line as the proof of Theorem 2 substituting ν_i with $k\nu_i$, for all i, $1 \le i \le n$ and noting that the smaller size packets are transmitted k times faster. □

The limit in Eq. (24) should be compared to the data collection in the case where packets are not split as shown in Eq. (7). We conclude that in the asymptotic case, data splitting results in gain in the collection time for networks with low data load, i.e. $\mu \le \frac{1}{2}$. It is also worth noting that Eqs. (24) and (7) imply that if $k \ge \frac{1}{2\mu}$ there is no gain in further increasing k; the expected delay remains the same as k further increases. For example, if $\mu = \frac{1}{5}$, the expected delay behaves like n,

$\frac{1}{2}n$, and $\frac{2}{5}n$ for $k = 1$, $k = 2$, and $k \geq 3$, respectively. In other words, increasing k beyond $\frac{1}{2\mu}$ does not lead to any improvement on the scaling law of the average delay.

3.5 A Simple Suboptimal Strategy

It is important to note that the minimum collection time in (2) is achieved under the assumption that each sensor node has a perfect knowledge of the network topology and data packets locations. A more practical strategy that does not require knowledge of the packets locations and therefore can be run in a distributed fashion is as follows: nodes at odd (resp. even) distance from the BS transmit to their closest neighbors toward the BS at odd (resp. even) TS. It is illustrated in Fig. 8.

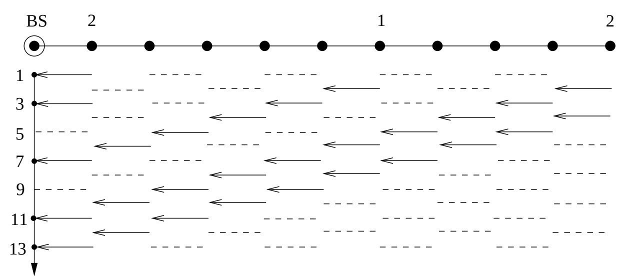

Fig. 8: Suboptimal data collection strategy described in section 3.5.

The following theorem compares the performance of this strategy to the minimal collection time derived in (2).

Theorem 8. *For a one-sided line network of length n in which the i'th node has ν_i packets and is equipped with directional antennas, the collection time of the packets at the BS under simple scheduling strategy, denoted by $T(\boldsymbol{\nu_n})$, is:*

$$T(\boldsymbol{\nu_n}) = \max_{1 \leq i \leq n} (i - 2 + 2 \sum_{j \geq i-1}^{n} \nu_j) \tag{25}$$

This further assumes that the closest, third closest, etc... edges to the BS are activated at TS 1, *3,... whereas the second closest, fourth closest,... edges are activated at TS 2, 4,... . In the opposite case the data collection time is:*

$$T(\boldsymbol{\nu_n}) = \max_{1\leq i\leq n}(i-1+2\sum_{j\geq i}^{n}\nu_j) \tag{26}$$

Proof. In the rest of this paper we refer to the closest edge to the BS as edge 1, second closest as edge 2 and so on. Assume TS 1, 3, 5,... are respectively allotted to edges 1,2,3,.... That is nodes 1, 3, 5... can only transmit at TS 1, 3, 5,... and receive at TS 2, 4, 6.... The BS may receive at most 1 packet/TS at TS 1, 3, 5,.... Either it is busy at all TS$\geq$ 1, or it is busy at all those TS$\geq$ 3, or at all TS$\geq$ 5, etc. In general if the BS is busy at all TS $\geq i$ and the packet received at TS i comes from node i or $i-1$ the data collection time is $i-2+2\sum_{j\geq i-1}^{n}\nu_j$ TS. This completes the proof for (25). Eq. (26) follows similarly. □

The aforementioned absence of knowledge (packets location) translates into a delay cost $T(\boldsymbol{\nu_n})-T_{min}(\boldsymbol{\nu_n})\geq 0$. More generally we have the following relationship between $T(\boldsymbol{\nu_n})$ and $T_{min}(\boldsymbol{\nu_n})$, which follows from (2) and (26):

$$T_{min}(\boldsymbol{\nu_n}) \leq T(\boldsymbol{\nu_n}) \leq 2T_{min}(\boldsymbol{\nu_n}) - 1 \tag{27}$$

The worst performance of this simple strategy relative to the optimal strategy occurs when n packets are located at distance 1 from the BS (Indeed $T_{min}=n$ and $T=2n-1$ then). However, on average, achieving the upper bound in (27) is unlikely and we have the following asymptotic comparative result, according to which the simple scheduling strategy is asymptotically optimal with respect to time:

Theorem 9. *Let ν_i's be i.i.d. random variables $\nu_i\in\{0,1,...,m-1\}$ with mean μ and variance σ^2*

where μ, σ^2, m *are constants independent of* n.

$$\lim_{n\to\infty} \frac{\mathbb{E}\{T(\boldsymbol{\nu_n})\}}{n} = \begin{cases} 2\mu & \textit{if } \mu \geq 1/2 \\ 1 & \textit{if } \mu \leq 1/2 \end{cases} \tag{28}$$

That is $\lim_{n\to\infty} \frac{\mathbb{E}\{T(\boldsymbol{\nu_n})-T_{min}(\boldsymbol{\nu_n})\}}{n} = 0$.

Proof. This proof is similar to the proof of Theorem 2. □

3.6 Imperfect Channel

In this final section we introduce noise in the channel. Specifically we model the channel as an erasure channel with erasure probability p and measure the time performance degradation as a function of p. We assume that a node is instantaneously informed that a packet has not reached its (intermediate) destination and immediately retransmits the erased packet at the next available TS (that is 2 TS later). For reasons discussed in section 3 we focus on the simple scheduling strategy introduced in subsection 3. 5. Fig. 9 illustrates the process. This is the same network as shown in Fig. 8 but it is now affected by three erasures (each shown by a crossed arrow). The new tranmission time is 15 TS, an increase of 2 TS.

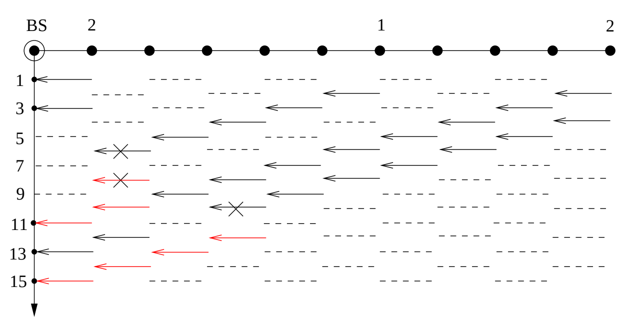

Fig. 9: Suboptimal data collection strategy described in section III.5. Erasure channel. An erased packet is marked with a cross.

Theorem 10. *Given a probability p of packet erasure, the data collection time $T(p, \boldsymbol{\nu_n})$ on a line network $\boldsymbol{\nu_n}$ when the simple scheduling strategy is used is:*

$$T(p, \boldsymbol{\nu_n}) = (1-p)^{\sum_{i=1}^{n} i\nu_i} \sum_{k \geq 0} p^k \sum_{\sum_i e_i \chi(\nu_i > 0) = k} \prod_{i \geq 1}^{n} \binom{i\nu_i + e_i - 1}{i\nu_i - 1} T(\boldsymbol{\nu_n} + \boldsymbol{e_i}) \tag{29}$$

Proof. The collection time may be expressed as an average of collection times. The probability that the entire collection process is not affected by any error is $(1-p)^{\sum_{i=1}^{n} i\nu_i}$. In that case the collection time is $T(\boldsymbol{\nu_n})$. The probability that the collection process is affected by exactly k errors is $(1-p)^{\sum_{i=1}^{n} i\nu_i} p^k$. Notice that a packet erasure along a specific edge increases the collection time from $T(\boldsymbol{\nu_n})$ to $T(\boldsymbol{\nu_n} + \boldsymbol{e_i})$ where $\boldsymbol{e_i}$ is the vector of length n whose ith component is 1 and other components are 0 and where i is the source node for the packet. For a given source node there are $\binom{i\nu_i + e_i - 1}{i\nu_i - 1}$ choices of e_i erasures. One needs to consider all the possible schedules with exactly k erasures. This can be done by solving the equation $\sum_i e_i \chi(\nu_i > 0) = k$. □

In order to see the impact of the erasure probability on the data collection time the ratio $T(p)/T(0)$ is plotted for increasing values of p for a specific line network $\boldsymbol{\nu} = (0, 2, 0, 0, 0, 0, 0, 1, 1, 1)$ in Fig. 10. It shows a degradation of 50% for an erasure probability $p = 0.1$. Our model shows that multihopping can have disastrous effects on the collection time in presence of noise. Note however that in networks with more general topology this needs not be, since in that case a node may choose to forward data to the neighbors with the best channels [13]. Theorem 10 allows for an exact computation of the delay incurred by a specific network, given a packet erasure probability, however the overall insight provided by it is limited. In the following Theorem, instead of considering the expected delay for a specific network, we consider a random line network and obtain an upper bound for the expected delay as a function of the packet erasure probability:

Theorem 11. *Let ν_i's be i.i.d. random variables $\nu_i \in \{0, 1, ..., m-1\}$ with mean μ and variance*

σ^2 *where* μ, σ^2, m *are constants independent of* n *then:*

$$1 \leq \frac{\mathbb{E}(T(p, \boldsymbol{\nu_n}))}{\mathbb{E}(T(0, \boldsymbol{\nu_n}))} \leq 1 + O(np) \tag{30}$$

Proof. In order to find an upper bound for the expected delay, we may use any strategy in scheduling. Here, we assume that whenever an erasure occurs, the transmitting node retransmits the packet until it gets through and all the other nodes remain silent at that period. Denoting α_i for $i = 1, \ldots, \sum i\nu_i$ as the number of extra time slots needed to transmit the packet at the i'th transmission, we may write,

$$T(p, \boldsymbol{\nu_n}) \leq \sum_{j=1}^{\sum_{i=1}^{n} ip_i} \alpha_j + T(0, \boldsymbol{\nu_n}) \tag{31}$$

where α_i has geometric distribution, i.e.

$$\Pr(\alpha_i) = p^{i-1}(1-p) \Rightarrow \mathbb{E}(\alpha_i) = \frac{p}{1-p} \tag{32}$$

Taking expectation of both sides of (31), we obtain,

$$\frac{\mathbb{E}(T(p))}{\mathbb{E}(T(0))} \leq \frac{p\sum_{i=1}^{n} ip_i}{(1-p)\mathbb{E}(T(0))} + 1 \tag{33}$$

which completes the proof of our theorem. □

In particular Theorem 11 implies that for networks of large size, a probability of erasure p of order $o(\frac{1}{n})$ does not significantly affect the time performance of the data collection process.

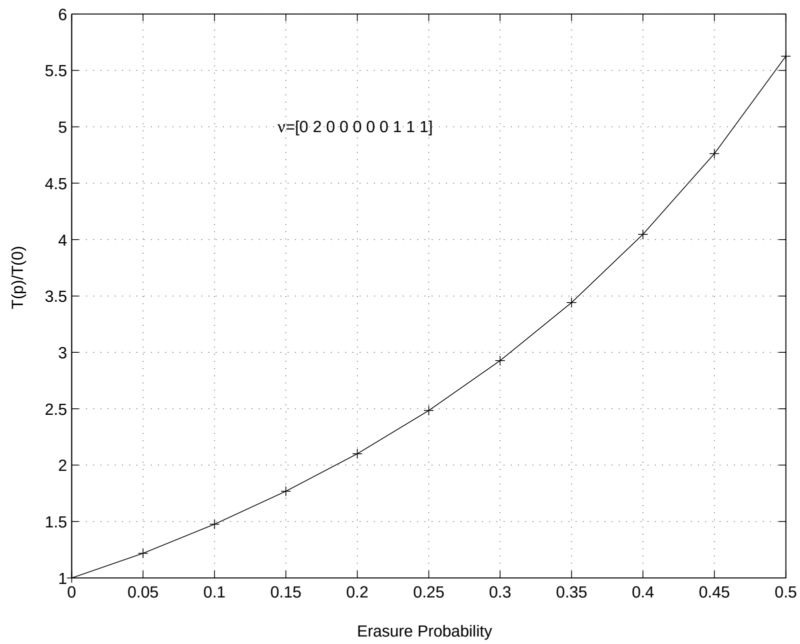

Fig. 10: Ratio $\frac{T(p)}{T(0)}$ as a function of p for line network $\boldsymbol{\nu} = (0, 2, 0, 0, 0, 0, 0, 1, 1, 1)$.

4 Random Multiline Networks

In this section, we consider a more general network, i.e. a network consisting of $L \geq 2$ lines. For simplicity we assume each line has n_0 nodes. This is illustrated in Fig. 11. Furthermore each node carries $\nu \in S_m$ packets with probability distribution $(p_0, p_1, \ldots, p_{m-1})$. We will later argue that the results for the more general case follows along the same line of this simple case.

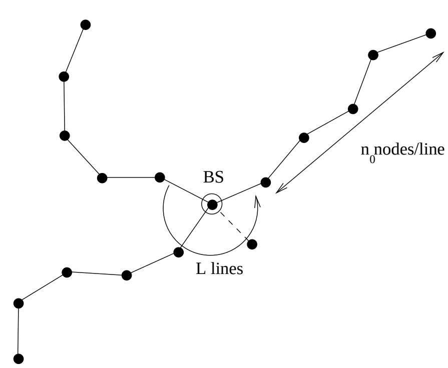

Fig. 11: Multiline Network.

It is quite easy to state a lower bound for the average delay. Assuming ν_i's are i.i.d., and denoting T_{min}^{L,n_0} as the minimum data collection time for a multiline network with $L \geq 2$ lines of length n_0, we have,

$$\mathbb{E}(T_{min}^{L,n_0}) \geq n_0 L \mathbb{E}(\nu_i) \tag{34}$$

which follows by taking the expectation of both sides of the inequality $T_{min}^{L,n_0} \geq$ (number of packets in network). In what follows, we shall prove that as L increases, the expected collection time converges toward this lowerbound.

To prove our asymptotic result, we describe a suboptimal procedure to collect the data at the BS: we may divide the network into two subnetworks $\mathcal{S}_1$ consisting of odd lines and $\mathcal{S}_2$ consisting of even lines. For $l \in \mathcal{S}_2$, nodes at even distance from the BS transmit toward the BS at even time slots and nodes at odd distance from the BS transmit toward the BS at odd time slots. If $l \in \mathcal{S}_1$ the opposite happens, i.e. nodes at even distance transmit toward the BS at odd time slots and vice versa. However if at a given TS multiple nodes at distance 1 from the BS carry data packets, only one packet (randomly chosen from all available packets) gets transmitted to the BS (since this BS can only receive one packet at a time). Remaining packets are stored for later transmission. This strategy is followed until all packets in the network have reached the BS or a node at distance one from the BS. At this point packets at distance one from the BS are simply transmitted to the BS in turn so that the BS does not become idle until all packets have been collected.

With this scheduling and assuming each node carries at most $m-1$ data packets it is clear that after $(m-1)(2n_0-3)$ TS (assuming that $\nu_i \in S_2$), all the packets are within distance one of the BS (since it is true in the worst case where each node carries exactly $m-1$ packets). Therefore, we may think of data collection as two separate phases: first collecting all the packets to the nodes with distance one of the BS which at most takes $(m-1)(2n_0-3)$ TS, and in the second phase, the nodes within distance one of BS are the only nodes with packets and they try to send their packets to BS.

Theorem 12. *Consider a multiline network with $L \geq 2$ lines of length n_0, and ν_i's are i.i.d. chosen from $\{0, 1, ..., m-1\}$ with an arbitrary distribution. Let $\forall k$, $0 \leq k \leq m-1$, $\Pr(\nu_i = k) = p_k$ where $p_{m-1} \neq 0$. Further assume that $\mathbb{E}(\nu_i) = \mu$ Then*

$$n_0 L\mu \leq \mathbb{E}(T_{min}^{L,n_0}) \leq n_0 L\mu + O\left(\frac{1}{L}\right) + (m-1)(2n_0-3)(1-p_{m-1})^{L/2}. \tag{35}$$

In particular,

$$i) \quad \text{if } L > (2+O(1))\log_\alpha n_o, \lim_{n_o\to\infty} \mathbb{E}(T_{min}^{L,n_0}) - n_o L\mu = 0 \tag{36}$$

$$ii) \quad \text{if } 2\log_\alpha n_o > L \text{ and } \lim_{n_o\to\infty} L = +\infty, \lim_{n_o\to\infty} \frac{\mathbb{E}(T_{min}^{L,n_0})}{n_o L\mu} = 1 \tag{37}$$

$$iii) \quad \text{if } L = cte, n_o L\mu \leq \mathbb{E}(T_{min}^{L,n_0}) \leq n_o L\mu(1+\epsilon) \tag{38}$$

where $\alpha = \frac{1}{1-p_{m-1}}$ and ϵ is a constant independent of n_o when L is fixed.

Proof. The lower bound follows from Eq. (34) and noting that $\mathbb{E}(\nu_i) = \mu$. To prove the upper bound, we use the suboptimal scheduling described before to collect the data packets. We also define the random variable $e_i \in \{0, 1\}$, for $i = 1, \ldots, (m-1)(2n_0-3)$, such that $e_i = 0$ if the BS is busy at TS i, and $e_i = 1$ if it is not.

Considering the steps in collecting packets in the network with our scheduling, if the total number of packets is greater than $(m-1)(2n_0-3)$, then the time needed to collect the data packets is equal to the total number of packets in the network (denoted by η) plus the number of times that the BS was not busy during $1 \leq t \leq (m-1)(2n_0-3)$ which is equal to $\sum_{i=1}^{(m-1)(2n_0-3)} e_i$. Therefore, we can write an upper bound for the delay as,

$$T_{min}^{L,n_0} \leq \max\{\eta,\ (m-1)(2n_0-3)\} + \sum_{i=1}^{(m-1)(2n_0-3)} e_i. \tag{39}$$

To find an upper bound for the expected delay, we have to find $\Pr(e_i = 1)$ and $\Pr(\eta \leq (m-1)(2n_0-3))$. To find an upper bound for the expected delay, we find $\Pr(e_i = 1)$ and $\Pr(\eta \leq (m-1)(2n_0-3)$. It is clear that

$$\begin{aligned}
\Pr(e_{2k} = 0) &\geq \Pr(\text{having at least m} - 1 \text{ packets at dist. } k) \\
&\geq \Pr(\text{at least one node at dist. k has m} - 1 \text{ packets}) \\
&= 1 - (1 - p_{m-1})^{L/2}
\end{aligned} \tag{40}$$

A similar expression can be written for $\Pr(e_{2k+1} = 0)$. Furthermore, using Chebychev's inequality and noting that η is the total number of packets in the network, i.e. $\eta = \sum_{i=1}^{n_0 L} \nu_i$, we may write,

$$\Pr\left((m-1)(2n_0-3) \leq \eta\right) \geq 1 - O\left(\frac{1}{n_0 L}\right) \tag{41}$$

which implies that $\Pr(\eta \leq (m-1)(2n_0-3)) \leq O\left(\frac{1}{n_0 L}\right)$. Now we can take the expectation from both sides of (39) to get,

$$\begin{aligned}
\mathbb{E}(T_{min}^{L,n_0}) &\leq \mathbb{E}(\eta) + (m-1)(2n_0-3)\Pr(\eta \leq (m-1)(2n_0-3)) + \sum_{i=1}^{(m-1)(2n_0-3)} \Pr(e_i = 1) \\
&\leq n_0 L \mu + O\left(\frac{1}{L}\right) + (m-1)(2n_0-3)(1-p_{m-1})^{L/2}
\end{aligned} \tag{42}$$

that completes the proof for the first part. □

Theorem 12 shows that either $i)$ the difference of the expected delay and the average number of packets is converging to zero as $L \to \infty$ and n_0 grows slower than 2^{2L} (that is equivalently, L grows faster than $O(\log n)$) or at least that ii) the ratio of the expected delay to the average number of packets converges toward 1 as long as L goes to infinity. It is a reasonable hypothesis in general. Indeed as the number of sensor nodes per unit of observation area increases, noting that L is the number of sensors within reach of the BS, it can be shown that L scales like $\log n + c(n)$ where $c(n) \to \infty$ [14]. Therefore, fixing the area of the network, having n goes to infinity, and noting that $n_0 = n/L$, the aforementioned condition is satisfied. We will come back to that in the next section which deals with more general topologies.

In the more general case where the number of sensors per line is ${n^l}_0$ for $l = 1, \ldots, L$ (instead of n_0 for all l's) the lower bounds on the expected delay becomes $\mathbb{E}(T^{L,n_0}_{min}) \geq \mu \sum_{l=1}^{L} n^l_0$. We can further find an upper bound by replacing n_0 by max ${n^l}_0$ in (39) and noting that $\mathbb{E}(\eta)$ is equal to the lower bound. The result follows in a similar fashion. Therefore as long as $(\max {n^l}_0)m = o\left(\frac{1}{1-p_{m-1}}\right)^L$ and L grows to infinity, the expected delay converges to $E\{\eta\}$. In Fig. 12 the difference between average collection time and average packet number in the network for multiline networks is plotted as the function of the number of lines for various average number of packets per node (and a fixed number of nodes per line, $n_0 = 25$) using Monte Carlo simulation. Each instance of a random network has L lines of n_0 nodes. Each node caries either 0 or 1 packet with probability $1 - \mu$ and μ respectively. The exact collection time for a particular instance is known and given in [10] and this is averaged over multiple instances (20000) to yield Fig. 12.

4.1 Delay analysis for More General Topologies

Insightful results about the delay in collecting data from sensory networks forming more general topologies may be inferred from sections 3 and 4. In this section we discuss the implications of previous results for networks of more general topologies.

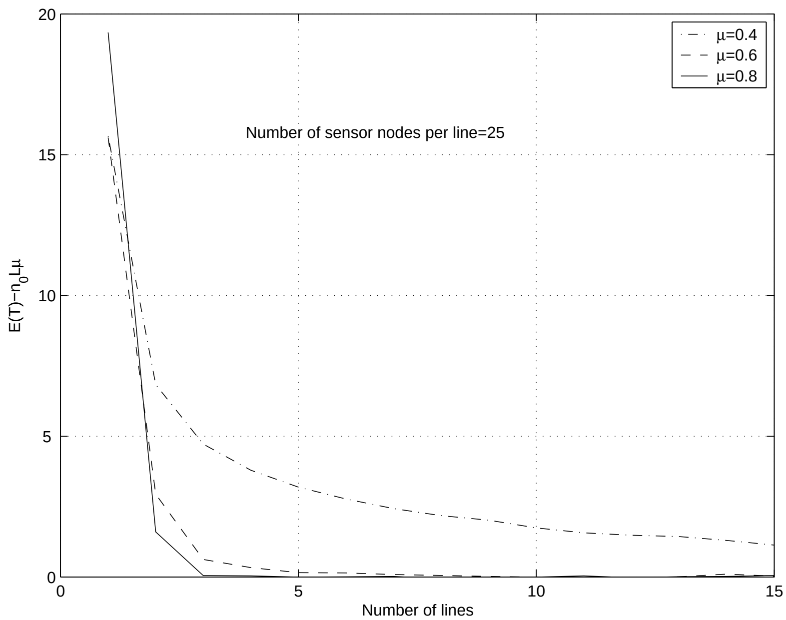

Fig. 12: Difference between expected delay and average number of packets in network ($\mathbb{E}(T_{min}^{L,n_0}) - n_0 L\mu$) as a function of average number of packets per node and number of lines in multiline network (25 nodes per line). Nodes carry 0 or 1 data packet with probability $1 - \mu$ and μ respectively.

Clearly for a sensor network of any topology, the expected minimum collection delay satisfies: $\mathbb{E}(T) \geq n\mathbb{E}(\nu_i)$ where n is the number of sensor nodes in the network. However in the particular case where only a single path exists from the sensors to the BS (i.e the degree of the BS is one) this lower bound is not tight and may be improved to: $2n\mathbb{E}(\nu_i)$ using Theorem 1.

If the degree of the BS is 1, It is shown in [11] that the network may be thought of as a line network -for analysis purposes- by combining nodes at the same distance from the BS without impeding the time performance of optimal data collection strategy. In the resulting "linearized" network the number of data packets at a given distance from the BS is the sum of the packets at that distance in the original network. Consequently results in section 3.1 may be applied to this type of networks to derive the exact delay distribution. Results in section 3.2 hold as well. That is the delay is $2n\mathbb{E}(\nu_i)$ asymptotically in the first order.

If the degree of the BS is greater than 1, it is straightforward to extend the previous results on multiline networks to tree topologies (indeed given what what said before a tree may be thought of as a multiline network).

Finally the previous results give some intuition about the asymtotic average minimum collection time in a random sensory network. Consider a disk of radius 1 and a network of n sensors randomly located on that disk. Assume the BS is placed at the center of that disk. We know from [14] that the minimum transmission range $r(n)$ must satisfy $\pi r^2(n) = \frac{\log(n)+c(n)}{n}$ where $c(n) \to \infty$ to insure network connectivity as n goes to infinity. We can then argue that the average collection delay converges toward the average number of packets in the network when the number of sensors is large. Indeed, a shortest path spanning tree of the considered network rooted at the BS may be extracted. From what was said begore this network behaves like a multiline network as far as delay is concerned and noticing that the maximum distance of a sensor to the BS (the distance being the length in number of hops of a shortest path to the BS) grows like $\frac{1}{r(n)} = O(\sqrt{\frac{n}{\log(n)}})$ and L is the number of packets within reach of the BS, that is $\pi r^2(n)n = O(\log(n)) =$ either condition $i)$ or condition $ii)$ of Theorem 12 applies.

5 Comparison of Omnidirectional and Directional Systems

The previous analysis of directional antenna systems may be extended to omnidirectional systems. In these systems, nodes are equipped with omnidirectional antennas generating interference for all surrounding nodes. In particular in a line network this implies that a packet transmission to the left (or right) neighbor creates interference at both the left and right neigbors. This in turns increases the length of the optimum data collection schedule (when compared to directional systems). In fact it was shown in [11] that the minimum data collection time $T_o(\boldsymbol{\nu_n})$ over a line network of length n

equipped with omnidirectional antennas in which the ith node has ν_i packets becomes:

$$T_{o,min}(\boldsymbol{\nu_n}) = \max_{1 \le i \le n-2} (i - 1 + \nu_i + 2\nu_{i+1} + 3 \sum_{j \ge i+2}^{n} \nu_j) \tag{43}$$

where $\boldsymbol{\nu_n} = (\nu_1, \ldots, \nu_n)$. It was shown in [10] that this represents a maximum increases of 50% over the data collection time achieved by a directional antenna system for the same considered line network. In the example of Fig. 1 the minimum data collection time becomes 14 TS, a 40% increase.

In the following sections, we presents results for the delay analysis for networks equipped with ominidirectional antennas. Results are analogous to the results stated in section 3 and we omit the proofs for the sake of brevity.

5.1 Delay Distribution

In this section we derive, by means of a recursion, the cumulative distribution function (CDF) of $T_o(\boldsymbol{\nu_n})$ for a line network. Let's assume that ν_i's are i.i.d. random variables chosen from the set $S_m = \{0, 1, \ldots, m-1\}$.

Theorem 13. *Let $F_n(t)$ be the CDF of the minimum delay $T_o(\boldsymbol{\nu_n})$, i.e. $F_n(t) = \Pr\{T_o(\boldsymbol{\nu_n}) \le t\}$. Then $F_n(t)$ satisfies the following recursion*

$$F_n(t) = \sum_{i=1}^{m-1} \Pr(\nu_n = i) F_{n-1}(t - 3i) \mathbf{1}_{t \ge n+3(i-1)} + \Pr(\nu_n = 0) F_{n-1}(t) \quad \forall n \ge 3 \tag{44}$$

where

$$\mathbf{1}_{t \ge t_0} = \begin{cases} 1 & \text{if} \quad t \ge t_0 \\ 0 & \text{otherwise,} \end{cases}$$

and,

$$F_1(t) = \begin{cases} \sum_{i=0}^{t} \Pr(\nu_1 = i) & \text{if } t < m-1 \\ 1 & \text{otherwise,} \end{cases}$$

$$F_2(t) = \sum_{i=1}^{m-1} \Pr(\nu_2 = i) F_1(t-2i) \mathbf{1}_{t \geq 2i} + \Pr(\nu_2 = 0) F_1(t).$$

5.2 Asymptotic Analysis of the Average Delay

In this subsection, we study the asymptotic behavior of the minimum average delay in collecting data from a line network as the number of nodes becomes large.

Theorem 14. *Let ν_i's be i.i.d. random variables $\nu_i \in S_m$ with mean μ, variance σ^2 where μ, σ^2, m are all constants independent of n. We have*

$$\lim_{n \to \infty} \frac{\mathbb{E}\{T_o\}}{n} = \begin{cases} 3\mu & \text{if } \mu \geq 1/3 \\ 1 & \text{if } \mu \leq 1/3 \end{cases} \tag{45}$$

5.3 Multiline/omnidirectional Case

Theorem 15. *Consider a multiline network with L lines of length n_0, and ν_i's are i.i.d. chosen from $\mathcal{S}_m$ such that $\forall k$, $0 \leq k \leq m-1$, $\Pr(\nu_i = k) = p_k$ where $p_{m-1} \neq 0$. Further assume that $\mathbb{E}(\nu_i) = \mu$ Then*

$$n_0 L \mu \leq \mathbb{E}(T_o) \leq n_0 L \mu + O\left(\frac{1}{L}\right) + (3n_0(m-1) - 2)\left(1 - p_{m-1}\right)^{L/3}. \tag{46}$$

In particular,

$$i) \quad \text{if } L > (3 + O(1)) \log_\alpha n_o, \ \lim_{n_o \to \infty} \mathbb{E}(T_o) - n_o L \mu = 0 \tag{47}$$

$$ii) \quad \text{if } 3 \log_\alpha n_o > L \text{ and } \lim_{n_o \to \infty} L = +\infty, \ \lim_{n_o \to \infty} \frac{\mathbb{E}(T_o)}{n_o L \mu} = 1 \tag{48}$$

$$iii) \quad \text{if } L = cte, n_o L \mu \leq \mathbb{E}(T_o) \leq n_o L \mu (1 + \epsilon) \tag{49}$$

where, $\alpha = \frac{1}{1 - p_{m-1}}$ *and* ϵ *is a constant independent of* n_o *when* L *is fixed.*

6 Conclusion

This work is concerned with characterizing the delay in collecting data from sensory networks at the BS. Under the assumption that the number of data packets accumulated by a sensor node is a random variable, we give lower and upper bounds for the average delay and derive the asymptotic behavior of this quantity as the number of nodes becomes large. Note that if the number of packets at each node is deterministic, the exact delay can be derived for tree topologies [11]. However, using probabilistic approach, we showed that asymptotically the average delay converges to the expected number of packets in the network for a tree with multiple connections to the BS. We further argued that this holds for sensory networks of randomly located nodes in a disk as well. Furthermore this paper derives exact relationships between data collection time and transmission range, data packet size and channel noise in the simple line scenario. To develop intuition these relationships are studied in the asymptotic case where the number of sensor nodes becomes large. Remarkably we show that multihopping does not lead to significant deterioration of the time efficiency of the data collection process. Indeed the latter deteriorates by a maximum factor of 2 when compared to direct transmission. This seems like a relatively low cost to pay in comparison to the energy saving realized by multihopping, which is of the order of the number of sensor nodes in the network. On the other hand our model shows that multihopping can have disastrous effects on the collection time in presence of noise. Note however that in networks with more general topology this needs not be, since in that case a node may choose to forward data to the neighbors with the best channels.

References

[1] J. R. Agre, L. P. Clare, G. J. Potty, and N. P. Romanov, "Development platform for self-organizing wireless sensor networks," *Proceedings of SPIE, unattended ground sensor technologies and applications*, vol. 3713, pp. 257–268, 1999.

[2] Y. Sankarasubramaniam I. F. Akyildiz, W. Su and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, Aug. 2002.

[3] P. Gupta and P.R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, Mar. 2000.

[4] M. Gastpar and M. Vetterli, "On the capacity of large gaussian relay networks," *IEEE Transactions on Information Theory*, 2003.

[5] A. F. Dana and B. Hassibi, "On the power efficiency of sensory and ad-hoc wireless networks," *submitted to IEEE Transactions on Information Theory.*

[6] F. T. Leighton, B. M. Maggs, and S. B. Rao, "Packet routing and job-shop scheduling in 0(congestion+dilation) steps," *Combinatorica*, vol. 14, pp. 167–186, 1994.

[7] P. Fraigniaud, "Minimum-time broadcast under edge-disjoint paths modes," *International conference on fun with algorithm*, 2001.

[8] P. Fraigniaud, "Approximation algorithms for minimum-time broadcast under the vertex-disjoint paths modes," *European symposium on algorithms*, 2001.

[9] B. S. Chlebus, L. Gasieniec, A. Gibbons, A. Pelc, and W. Rytter, "Deterministic broadcasting in ad hoc radio networks," *Distributed computing*, 2002.

[10] R. J. McEliece C. Florens, M. Franceschetti, "Lower bounds on data collection time in sensory networks," *Journal of Selected Areas in Communications*, 2004.

[11] C. Florens and R. McEliece, "Scheduling algorithms for wireless ad-hoc sensor networks.," *Proceedings of the IEEE Global Telecommunications Conference 2002*, Nov. 2002.

[12] P. Erdos and M Kac, "On certain limit theorems of the theory of probability," *Bull. Amer. Math. Soc.*, pp. 292–301, 1946.

[13] Jan Rabaey, "Reliable operation despite ultra low power - the ultimate quest in wireless sensor networks," *Caltech RF and Microwave Seminar*, Mar. 2004.

[14] P. Gupta and P.R. Kumar, "Critical power for asymptotic connectivity," *Proc. IEEE 37th Conf. on Decision and Control*, pp. 1106–1110, 1998.