Data Collection and Distribution in Sensory Networks

Thesis by

Cédric Florens

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2004
(Submitted May 3, 2004)

Acknowledgements

I would like to begin by expressing my deepest gratitude to my advisor, Prof. Robert J. McEliece for his continuous guidance and support throughout my stay at Caltech. His sharp insight and infectious enthusiasm proved to be of inestimable value in my research endeavors and in the production of this document.

I would like to thank Prof. Robert J. McEliece, Prof. Jehoshua (Shuki) Bruck, Prof. Babak Hassibi, Prof. Steven Low at Caltech, and Dr. Polly Estabrook at the Jet Propulsion Laboratory for serving on my candidacy and defense committees, and for their advice on this thesis.

I am also indebted to Dr. Loren Clare and Dr. Jay Gao at the Jet Propulsion Laboratory for their encouragement and comments on my work.

I am grateful to my officemates Amir Dana, Radhika Gowaikar, Ravi Palanki, Chaitanya Rao, Masoud Sharif, Mihailo Stojnic and Jeremy Thorpe for making the Posner Communication Lab a pleasant and productive environment.

My work was supported by a research assistantship funded by the Caltech Lee Center for Advanced Networking.

Abstract

The deployment of large-scale, low-cost, low-power, multifunctional sensory networks brings forward numerous and diverse research challenges. Critical to the design of systems that must operate under extreme resource constraints, the understanding of the fundamental performance limits of sensory networks is a research topic of particular importance. This thesis examines, in this respect, an essential function of sensory networks, viz., data collection, that is, the aggregation at the user location of information gathered by sensor nodes.

In the first part of this dissertation we study, via simple discrete mathematical models, the time performance of the data collection and data distribution tasks in sensory networks. Specifically, we derive the minimum delay in collecting sensor data for networks of various topologies such as line, multi-line, tree and give corresponding optimal scheduling strategies assuming that the amount of data observed at each node is finite and known at the beginning of the data collection phase. Furthermore, we bound the data collection time on general graph networks.

In the second part of this dissertation we take the view that the amount of data collected at a node is random and study the statistics of the data collection time. Specifically, we analyze the average minimum delay in collecting randomly located/distributed sensor data for networks of various topologies when the number of nodes becomes large. Furthermore, we analyze the impact of various parameters such as lack of synchronization, size of packet, transmission range, and channel packet erasure probability on the optimal time performance. Our analysis applies to directional antenna systems as well as omnidirectional ones. We conclude our study with a simple comparative analysis showing the respective advantages of the two systems.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

In this chapter we give a brief overview of sensory networks main characteristics, applications, and research issues. The emphasis is put on the research topics relevant to the contents of this thesis. We summarize the main contributions of our work and conclude by an outline of this dissertation.

1.1 Sensory Networks: A Brief Overview

Recent technological advances in the Very Large Scale Integration (VLSI) field have contributed much to the development of microsensor systems. These combine various sensors, signal processing capabilities, data storage capabilities, wireless (radio, infrared or optical) communication capabilities, and energy sources on a single chip [1, 2, 43]. Such 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. Fig. 1.1 illustrates the architecture of a sensor node while Fig. 1.2 illustrates a sensory network in a sensor field.

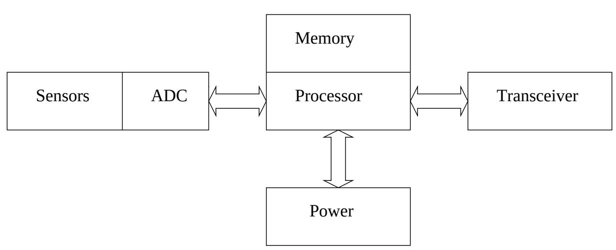


Figure 1.1: Sensor node architecture.

In the future sensor networks promise to revolutionize our lives. Pervasive wireless integrated networks will provide access to information anytime, anywhere and will be able to instantaneously respond to our actions, in a way creating smart environments

[53, 20, 21]. Potential applications for such networks are numerous and can be broadly divided into military and civilian categories. Military applications include space exploration [41], battlefield surveillance and enemy tracking [54]. Civilian applications include habitat monitoring [46, 10], environmental observation and forecast [68, 45], as well as various health applications [61].

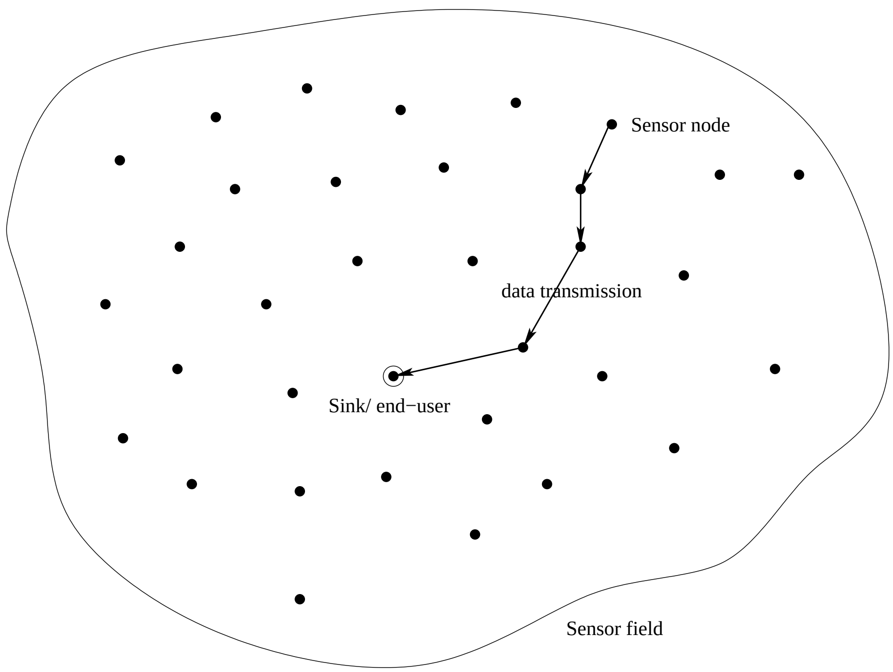


Figure 1.2: Sensor nodes scattered in field.

Sensor networks are wireless networks with unique characteristics which distinguishes them from traditional wireless networks [69]. They are designed for unattended operation, must accommodate a traffic of statistical nature, support very low data rates to the order of 1-100 kb/s, and are characterized by a predominantly unidirectional flow of data from sensor nodes to sink. Sensory networks are members of the wireless ad hoc network family, that is, they are infrastructure-less networks unlike cellular networks. But they also distinguish themselves from MANETs[1], designed to provide good throughput/delay characteristics under high mobility conditions, without much regard for energy consumption. Indeed, operation under severe constraints (lack of accessibility, limited energy resources and capabilities of nodes, absence of

[1]Mobile Ad Hoc Networks

infrastructure), not existent in more traditional networks, imposes aggressive energy management [58]. Accordingly, in many sensor network applications, energy (or equivalently lifetime) is traded against throughput/delay. Finally, 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. Accordingly the deployment of sensory networks brings forward numerous research problems [4, 70, 52].

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 [40]. The protocol stack typically used by sensor nodes is composed of a physical layer, data link layer, network layer, transport and application layers, as well as a power plane, mobility plane, and task management plane [4] as schematically illustrated in Fig. 1.3. The physical layer is responsible for frequency selection, carrier frequency generation, signal detection, modulation and data encryption [65, 11]. The data link layer is responsible for the multiplexing of data streams, data frame detection, medium access (MAC) and error control. The MAC protocol ensures the creation of the network infrastructure and efficient communication resource allocation between the sensor nodes [67, 72, 42, 7, 56, 72]. The network layer provides routing capabilities [59, 64] to the transport layer and is responsible for internetworking with external networks. Finally, the transport layer is responsible for maintaining the flow of data when and if required by the application layer. The three planes are responsible for task allocations between nodes and monitoring/managing energy consumption, mobility. An investigation of current protocol and algorithm proposals in these layers is presented in [4]. Technical issues and application requirements to be dealt with by these protocols are multiple and often specific to the class of sensory networks as mentioned earlier. Among those, efficient management of energy budget is of paramount importance to the lifetime of the networks [36, 66]. Important issues under investigation include node localization [15], clock synchronization [38, 37], fault tolerance, connectivity and coverage issues [31, 8, 48, 39], security [51], analysis of network fundamental performance limits, and hardware design [57, 11, 49, 71].

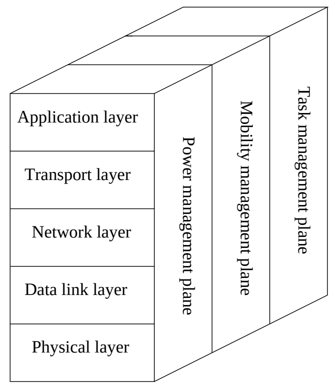


Figure 1.3: Sensory network protocol stack.

1.2 Contributions

The extreme resource constraints under which wireless sensor networks must operate, strongly motivate an understanding of the fundamental performance limits of these systems, for example in order to figure out in what areas improvements over state-of-art protocol and hardware design are possible and efforts should be directed. The main performance measure considered in the literature are capacity or throughput, power consumption and network lifetime. Results on capacity may be found in [34, 32, 5, 50, 17, 18, 47, 3]. Upper bounds on the lifetime of a sensor network are derived in [6]. Energy expenditure is considered in [13, 30, 16, 14]. Distributed compression is studied in [62, 60, 63, 55]. In most applications, sensor networks are expected to autonomously extract information about their surroundings, perform basic collective processing and transmit the collected data to the end-user for further processing and analysis. In this dissertation we study the problem in sensory networks of collecting sensor data at the network processing center. Although many protocols based on resource-efficient heuristics have been proposed for data collection in sensory networks [35], few analyses of the process have appeared. In this thesis we derive new performance (with respect to time and to a lesser extent energy) results on data-gathering sensory networks. Specifically, we derive, via simple discrete mathematical models, lower bounds on data collection time (delay) in data gathering sensory networks and exhibit algorithms that achieve those bounds [22, 23, 24, 25, 26, 27]. Most 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 [44, 28, 29, 12, 9] with respect to wireline networks and general purpose wireless networks.

1.3 Thesis Outline

This dissertation is organized as follows: This chapter reviews briefly research issues in sensory networks and summarizes our contribution to the field. In Chapter 2, we 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 which motivates the more complex model of the following chapter.

In Chapter 3, 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 on delay. We also propose a simple distributed scheduling strategy and analyze its delay performance showing that it is asymptotically optimal. Finally we extend our result to more general topologies such as multi-line networks and trees.

Chapter 4 contains our concluding remarks as well as open problems.

Chapter 2 Deterministic Sensory Networks

In this chapter we study, with respect to a simple discrete mathematical model, the data collection problem in sensory networks. 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. We refer to this network model as deterministic sensory network. More specifically, we describe optimal strategies to perform data collection and derive corresponding time performances.

This chapter is organized as follows: In section 2.1 we describe our sensor network model. We present results in deterministic sensory networks equipped with directional antenna elements in section 2.2. In section 2.3, we propose a generalization to omnidirectional systems. We present a comparison analysis of the two systems in section 2.4. Finally, we conclude in section 2.5. Miscellaneous derivations for the chapter and Pseudo-code of presented algorithms are grouped in the appendices.

2.1 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. 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 assume two distinct 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 chapter, 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' (to be defined shortly). Fig. 2.1 illustrates sensor nodes, together with gathered data, scattered in a sensor field.

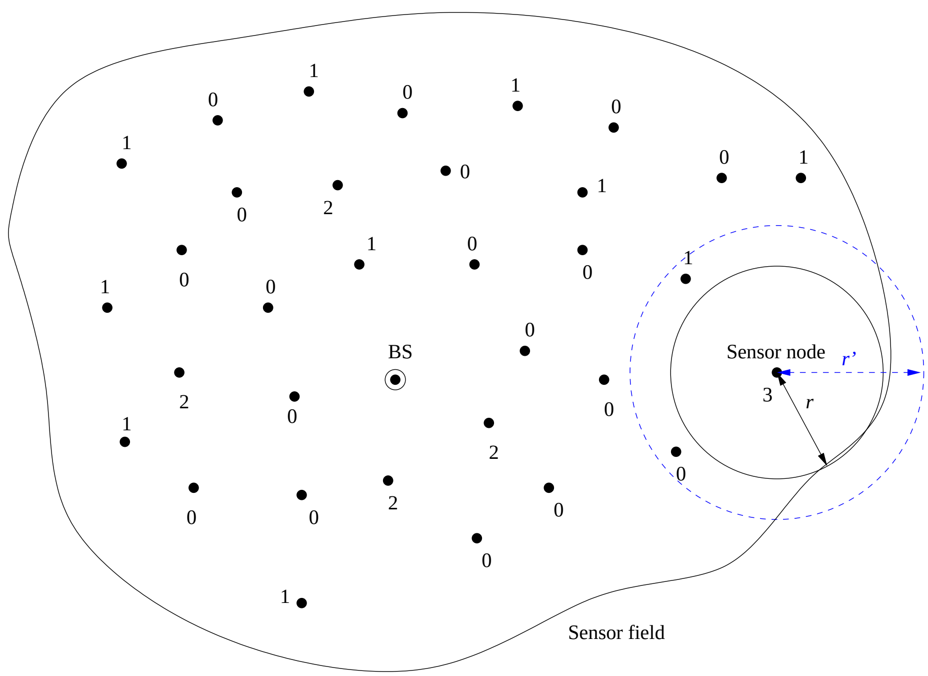


Figure 2.1: Sensor nodes, gathered data, and sensor field.

The interference model as defined in [34] 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

$$|N_i - N_j| \leq r, \; |N_k - N_j| \geq (1 + \delta)r, \; \delta > 0 \tag{2.1}$$

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. 2.2 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. 2.3 illustrates the sensory network formed by the nodes of Fig. 2.1. Nodes within transmission range are connected through a solid line while nodes within interference range are connected through a solid line or dotted line.

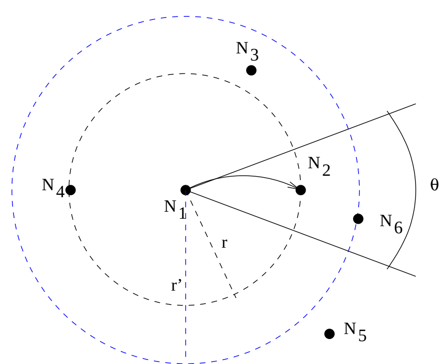


Figure 2.2: Interference model parameters.

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

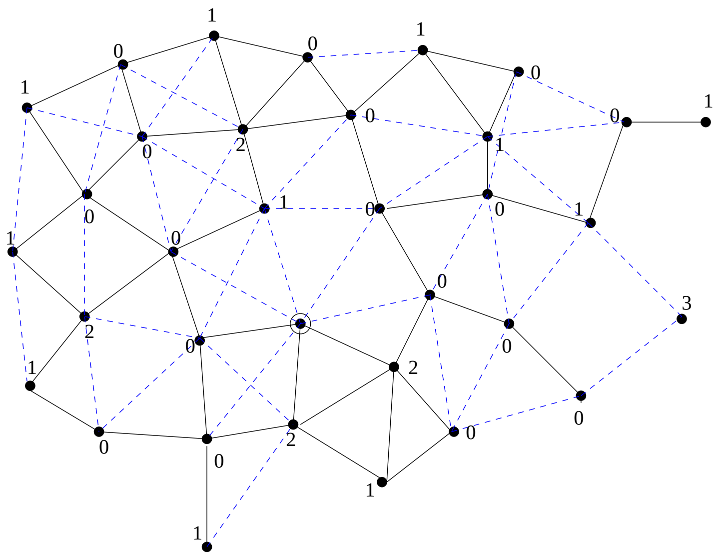


Figure 2.3: Sensory network.

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.

2.2 Directional Antenna Systems

In the following section, we focus our attention on directional antenna systems. We first study the data collection process in networks with linear topology (half-line, line, multi-line), then move to the study of networks with tree topology and conclude with a study of networks with general topology.

2.2.1 Line Networks

In this subsection, we consider a line network (an example of which is given in Fig. 2.4). A BS is placed at one end of the network. We assume sensor nodes are regularly placed along the network. We denote by d the distance between any 2 nodes. Assume each node is equipped with directional antennas allowing transmissions over a distance r where $d < r < 2d$. Further assume that δ is such that $(1 + \delta)r < 2d$. In this scenario there are two nodes (one on the left, one on the right) within transmission/interference range of any given node in the line (except for the end nodes). It is possible to extend this model to a more realistic scenario where nodes are randomly placed along a line and where different values of r, δ are considered (as long as end-to-end connectivity of the network is ensured). However, we find that simple case to be most insightful. In the following section, we consider more general scenarios. Let N_i be the node at distance i from the BS. We denote by $i \to i + 1$ a transmission from node i to node $i + 1$. Our goal is to determine the minimal duration of the collection phase and an associated optimal communication strategy.

For purpose of solving this problem we look initially at the following converse problem (which we shall refer to as the *distribution problem*); instead of nodes sending packets to the BS, assume the BS is to transmit packets to nodes. The data transfer efficiency remains our concern. This problem is of separate interest in sensor networks.

We propose the following simple algorithm for solving the distribution problem. We shall prove subsequently it is optimal. The BS is to 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). We include, in Appendix B, Algorithm 1 running at the BS.

The procedure is illustrated on an example, where $V = \{0, 1, 2, 3, 4, 5, 6, 7\}$, $E = \{(i, i + 1), 0 \leq i \leq 6\}$, $\boldsymbol{\nu} = (2, 0, 0, 0, 3, 0, 1)$, $d < r < 2d$, $(1 + \delta)r < 2d$, in Fig. 2.4.

The schedule of transmissions, as determined by Algorithm 1, is drawn below the network for the distribution and collection problems respectively. Either way it is performed in 11 TS.

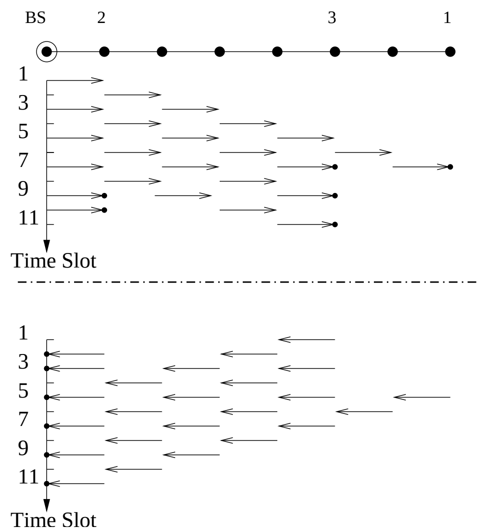


Figure 2.4: Optimal distribution and collection schedules in 8-node line network.

Next we determine the performance of our algorithm in general. Denote by T_i the last busy TS at node i in the execution of our distribution algorithm (In the previous example, we have $T_1 = 10, T_2 = 9, T_3 = 10, T_4 = 11, T_5 = 11, T_6 = 7, T_7 = 7$). Clearly then our algorithm runs in $\max_{1 \le i \le n} \{Ti\}$. T_i is a function of the distance to the BS, the number of packets destined for node i as well as the number of packets forwarded by node i. Assuming $\nu_i = 0$ for $i > n$, node i's last busy TS when running Algorithm 1 is

$$T_i = \begin{cases} i - 1 + 2\sum_{j \ge i+1} \nu_j & \text{if } \nu_i = 0 \\ \nu_1 + 2\sum_{j \ge 2} \nu_j & \text{if } i = 1 \text{ and } \nu_1 \ge 1 \\ i - 2 + 2\sum_{j \ge i} \nu_j & \text{if } i \ge 2 \text{ and } \nu_i \ge 1 \end{cases} \tag{2.2}$$

Proof. $\forall i \ge 1$, node i is idle the first $i - 1$ TS. It forwards $\sum_{j \ge i+1} \nu_j$ data packets to further nodes and receives ν_i data packets that are destined for itself. Forwarding a data packet consists in receiving that data packet and transmitting it right away and therefore a node involved in forwarding one data packet will remain busy two consecutive TS. Receiving a data packet on the other hand consumes only one TS but in our scheme forces node $i \geq 2$ to remain silent in the following TS. Therefore,

$$\nu_1 \geq 1 \Rightarrow T_1 = 2\sum_{j\geq 2}\nu_j + \nu_1$$
$$\nu_i \geq 1,\, i > 1 \Rightarrow T_i = (i-1) + 2\sum_{j\geq i+1}\nu_j + 2(\nu_i - 1) + 1$$
$$\nu_i = 0 \Rightarrow T_i = (i-1) + 2\sum_{j\geq i+1}\nu_j$$

□

We define, for a given sensor network, $T_u(\boldsymbol{\nu})$ the minimum length of a time schedule over all time schedules that perform the distribution job.

Theorem 2.2.1. *Assuming $\nu_i = 0$ for $i > n$, the minimal data collection time in the directional line network $\boldsymbol{\nu}$ of length n*[1] *is*

$$T_u(\boldsymbol{\nu}) = \max_{1\leq i\leq n-1}(i - 1 + \nu_i + 2\sum_{j\geq i+1}^{n}\nu_j) \tag{2.3}$$

Proof. Clearly the maximum of T_i is obtained over the set $\{i \geq 1 \,|\, \nu_i \neq 0\}$. Thus we have the following upper bound on $T_u(\boldsymbol{\nu})$

$$T_u(\boldsymbol{\nu}) \leq \max_{\{i\geq 1\,|\,\nu_i\neq 0\}} T_i$$

A lower bound on $T_u(\boldsymbol{\nu})$ is as follows. Assuming $\nu_i = 0$ for $i > n$, we have

$$T_u(\boldsymbol{\nu}) \geq \max_{1\leq i\leq n-1}(i - 1 + \nu_i + 2\sum_{j\geq i+1}^{n}\nu_j)$$

Indeed node i has to forward $\sum_{j\geq i+1}^{n}\nu_i$ data packets to further nodes. Forwarding one data packet consists in receiving and transmitting that data packet and therefore

[1] *Implicitly we assume that the distance to the BS of the furthest node carrying a packet is n.*

results in a two TS consumption (per forwarded packet). Besides, it is itself the destination of ν_i data packets. Each received data packet costs at least one TS. Furthermore, node i can't be active before it receives a data packet, which takes at least $i-1$ TS. Therefore, $S_i \triangleq 2\sum_{j\geq i+1}\nu_j + \nu_i + (i-1)$ is a lower bound on any time schedule for all i. Hence, $\max_{1\leq i\leq n-1} S_i$ is a lower bound on $T_u(\boldsymbol{\nu})$.

Finally, we prove that lower and upper bounds on $T_u(\boldsymbol{\nu})$ are equal and therefore the proposed schedule is optimal: Clearly $S_1 = T_1$ and $\forall i \geq 2$, $S_i = T_i$ if $\nu_i \leq 1$. On the other hand, if $\nu_i > 1$ then $T_i > S_i$ but then either $\nu_{i-1} = 0$ and then $S_{i-1} = T_i$ or $\nu_{i-1} \geq 1$ and then $S_{i-1} > T_i$. □

Corollary 2.2.2. *In the particular case where no two consecutive components of vector* $\boldsymbol{\nu}$ *equal zero, Eq. (2.3) reduces to:*

$$T_u(\boldsymbol{\nu}) = \nu_1 + 2\sum_{i\geq 2}\nu_i \tag{2.4}$$

We now return to the data collection problem. The construction of a schedule here is based on the symmetry of the operations of distribution and collection. A time schedule that is symmetric to the distribution problem's schedule with respect to a fictive horizontal axis (see example in Fig. 2.4) provides us with an optimal solution, the time to transmit data packets from nodes to the BS being the same as the time to carry out the converse operation (and being therefore minimal). In particular a transmission $i \to i+1$ occurring at TS j in the distribution problem is a transmission $i+1 \to i$ occurring at TS $T_u(\boldsymbol{\nu}) + 1 - j$ in the collection problem. Since the solution to one problem gives us the solution to the other, we only consider the distribution problem in the sequel. Note that an additional issue is raised in the data collection case; indeed the described algorithms don't require the network to be synchronous in the distribution case (so the algorithms may be run in a distributed way) whereas they do in the data collection case.

2.2.2 Toward More General Scenarios

The general line case is shown in Fig. 2.5. It consists of n randomly located sensor nodes $N_1, \ldots, N_n$ along a line and a BS N_0 at the left end of that line. It is assumed that each node's transceiver has a common transmission range r such that $r \geq \max_{0 \leq i \leq n-1} d(N_i, N_{i+1})$ where $d(N_i, N_{i+1})$ denotes the distance between nodes N_i and N_{i+1} (which ensures end to end connectivity of the network) and interference range $r' = (1+\delta)r$. Under these assumptions any given node will have in general more that one neighbor to the right (resp. left), those numbers varying from one node to the other. Particular cases of this scenario are solved in the remaining of this section. We first study the case where the transmission range is fixed and equal to one hop and the interference range is variable. We then study the case of variable transmission range.

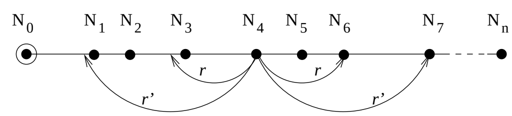


Figure 2.5: $(n+1)$-node line network where $r' = 2r$.

This analysis constitutes a generalization of the line network analysis in the previous section which allows us to study the respective impact of the transmission range and the interference range on the data collection process. We assume, for simplicity, that the number of left and right neighbors is the same (one in this case) for all nodes. Furthermore, it is convenient to imagine a line network with regularly spaced sensor nodes.

First case: variable interference range

We fix the transmission range to 1 hop and the interference range to m hops (that is $r = 1$ and $\delta = m - 1$). Note that in the previous section, m was taken to be 1. In practice m is often between 2 and 3.

The distribution strategy for the BS is to transmit ν_n data packets to node N_n first, then ν_{n-1} packets to N_{n-1}, and so on, as fast as possible while respecting the channel reuse/transceiver constraints. This strategy's time performance is $\max_i T_i$ where

$$T_1 = \begin{cases} \sum_{1\leq j}^{m-1} j\nu_j + m\sum_{j\geq m}\nu_j \text{ if } \nu_1 \geq 1 \\ 2 + m(\sum_{j\geq m}\nu_j - 1) \text{ if } \exists k \geq m \text{ such that: } \nu_1 = \ldots = \nu_{k-1} = 0, \nu_k \geq 1 \\ 2 + k(\nu_k - 1) + \sum_{k+1\leq j}^{m-1} j\nu_j + m\sum_{j\geq m}\nu_j \text{ if} \\ \qquad \exists k\, 2 \leq k < m,\ \nu_1 = \ldots = \nu_{k-1} = 0, \nu_k \geq 1 \end{cases}$$

$$T_2 = \begin{cases} \sum_{2\leq j\leq m-1} j\nu_j + m\sum_{j\geq m}\nu_j \text{ if } \nu_2 \geq 1 \\ 3 + m(\sum_{j\geq m}\nu_j - 1) \text{ if } \exists k \geq m \text{ such that: } \nu_2 = \ldots = \nu_{k-1} = 0, \nu_k \geq 1 \\ 3 + k(\nu_k - 1) + \sum_{j\geq k+1}^{m-1} j\nu_j + m\sum_{j\geq m}\nu_j \text{ if} \\ \qquad \exists k\, 2 \leq k < m,\ \nu_2 = \ldots = \nu_{k-1} = 0, \nu_k \geq 1 \end{cases}$$

$$T_i = \sum_{i\leq j\leq m-1} j\nu_j + m\sum_{j\geq m}\nu_j \text{ if } 2 < i < m$$

$$T_{m+k} = \begin{cases} k + m\sum_{j\geq m+k}\nu_j \text{ if } \nu_{m+k} \geq 1 \\ k + 1 + m\sum_{j\geq m+k}\nu_j \text{ if } \nu_{m+k} = 0 \end{cases} \qquad \text{if } k \geq 0$$

The proof follows a similar argument as the one used to prove Eq. (2.2) and is omitted. The following theorem gives a closed form expression for the minimum data collection delay. This generalizes Theorem 2.2.1.

Theorem 2.2.3. *The minimum data collection time $T_u^m(\boldsymbol{\nu})$ on directional line network $\boldsymbol{\nu}$, when the transmission range is 1 hop, and the interference range is m hops, is*

$$T_u^m(\boldsymbol{\nu}) = \begin{cases} \max_i (i - 1 + \sum_{j\geq i}^{i+m-2}(j-i+1)\nu_j + m\sum_{j\geq i+m-1}\nu_j,\ \forall m \geq 2 \\ \max_i (i - 1 + \nu_i + 2\sum_{j\geq i+1}\nu_j) \ \textit{if } m = 1 \end{cases} \tag{2.5}$$

Proof. We have $T_u^m(\boldsymbol{\nu}) \leq \max_i T_i$. A lower bound on the minimum time performance can be derived as well.

$$\forall i \geq 1,\, T_u^m(\boldsymbol{\nu}) \geq i - 1 + \sum_{j \geq i} \nu_j + \sum_{j \geq i+1} \nu_j + \ldots + \sum_{j \geq i+m-1} \nu_j \tag{2.6}$$

Indeed transmissions $i-1 \to i, i \to i+1, \ldots, i+m-2 \to i+m-1$ may not occur concurrently due to channel reuse constraints. Inequality (2.6) may be rewritten

$$\forall m \geq 2, \quad T_u^m(\boldsymbol{\nu}) \geq \max_i(S_i) \tag{2.7}$$

where

$$S_i \triangleq (i - 1 + \sum_{i \leq j \leq i+m-2} (j - i + 1)\nu_j + m \sum_{j \geq i+m-1} \nu_j) \tag{2.8}$$

The case $m = 1$ may be derived from the above formula by choosing $m = 2$. Assume there exists j_0, $1 \leq j_0 \leq n$ such that $\forall i \neq j_0$, $T_{j_0} \geq T_i \quad T_{j_0+1} < T_{j_0}$

- if $j_0 = 1$ then $\nu_1 \geq 1 \Rightarrow S_1 = T_1$
- if $j_0 = 2$ then $\nu_2 \geq 1, \nu_1 = 0 \Rightarrow T_2 - S_2 = \sum_{2 \leq j \leq m} \nu_j - 1 \geq 0$
  $\nu_1 = 0 \Rightarrow T_1 = T_2 \Rightarrow S_1 \geq T_2$. Indeed $S_1 \geq T_1$ since

$$\nu_1 \geq 1 \Rightarrow S_1 = T_1$$

$$\text{and } \nu_1 = 0 \Rightarrow S_1 - T_1 = \begin{cases} \sum_{1 \leq j \leq m-1} j\nu_j + m - 2 \geq 0 & (m \geq 2) \text{ or} \\ \sum_{1 \leq j \leq k-1} j\nu_j + k - 2 \geq 0 & (k \geq 2) \end{cases}$$

- if $2 < j_0 < m \Rightarrow \nu_i \geq 1, \nu_1 = \ldots = \nu_{i-1} = 0 \Rightarrow T_{j_0} - S_1 = -\sum_{1 \leq j \leq i-1} j\nu_j = 0$

- if $j_0 = m + k \quad k \geq 0 \Rightarrow \nu_{m+k} \geq 1 \quad \nu_k = \ldots = \nu_{k+m-1} = 0 \Rightarrow T_{m+k} - S_{k+1} = -\sum_{j \geq k+1}^{k+m-1} (j-k)\nu_j = 0$

Therefore $\max_i T_i = \max_i S_i$ and Theorem 2.2.3 follows. □

Second case: variable transmission range

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).

Theorem 2.2.4. *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 function of the transmission range h in hops is*

$$T_{min}(h, \boldsymbol{\nu_n}) = \max(S', S_1, S_2, \ldots, S_{n-h}) \tag{2.9}$$

where

$$\begin{aligned} S_i &= \sum_{j>i+h}^{n} \nu_j + \left\lfloor \frac{\sum_{j>i+h}^{n} \nu_j - 1 + (i \mod h)}{h} \right\rfloor + \left\lfloor \frac{i}{h} \right\rfloor + 1,\ 0 \le i \le n-h \\ S' &= S_0 + \max\left(\sum_{j=1}^{l} \nu_j - 1, 0 \right) + \sum_{j=l+1}^{h} \nu_j \end{aligned} \tag{2.10}$$

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. (2.9) reduces to the familiar Eq. (2.3).

Proof. The proof is similar to the proof of the case $h = 1$. Here we only outline the generalization. The proof has two parts. Firstly, we need to show that the right-hand side of (2.9) is a lower bound for the collection time. Secondly, we prove it is an upper bound as well by exhibiting 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 $i, 1 \le i \le h$ closest to the BS. They need to forward $n_h = \sum_{j>h} \nu_j$ packets. If $n_h \le h$, 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 $h + 1 \le n_h \le 2h$, 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 the algorithm's time performance and show it achieves the lower bound previously exhibited. This is similar to what was done in the case $h = 1$ and is left out here for the sake of brevity. □

In order to get a better insight into the result of Theorem 2.2.4, we give a simple illustrative example.

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 2.2.4 gives the minimum collection time as

$$T = n + \lfloor \frac{n}{h} \rfloor - 1 \tag{2.11}$$

Fig. 2.6 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.

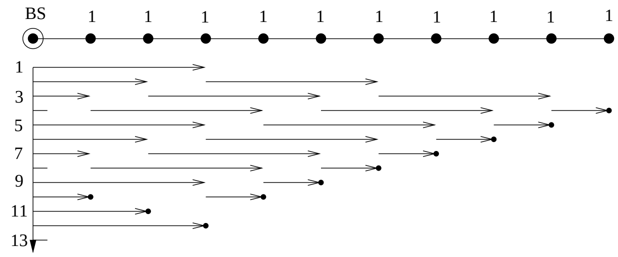


Figure 2.6: Minimum length data distribution schedule in 10-node line network with maximum transmission range of 3 hops.

2.2.3 Synchronization Problem

Up to now we have assumed that all sensor nodes and BS were synchronized. In this section, we study the impact on the data collection time of lack of synchronization in the network. In our communication model we have assumed that the transmission of a data packet consumed one TS. Let us be more precise. The transmission of a data packet is made up of a transmission phase (at the transmitting node), a propagation phase (from the transmitting node to the receiving node) and a reception phase (at the receiving node). Assuming sensor nodes are about one meter apart, that the size of a data packet is about 20 bytes and data rates are of the order of 10 kbps, we can get an idea of the duration of each phase. We find $\Delta TX = \Delta RX = 1.6 * 10^{-2}$ s which is very large compared to the propagation time $0.33 * 10^{-8}$ s. The latter may therefore be ignored for the purpose of this analysis. Then, the first half of the TS is used for transmission of the data packet at the transmitting node while the second half is used for reception of the data packet, by the receiving node. Fig. 2.7 illustrates the distribution schedule for a particular network assuming perfect synchronization (left figure, delay is 6 TS) and multiple unsynchronized cases. In the middle figure all sensor nodes are synchronized but out-of-synch with the BS and the delay is 8.5 TS. In the right figure the sensor node at distance one from the BS is out of synch with other nodes and the delay is 9 TS. We have the following theorem.

Theorem 2.2.5. *The worst-case time performance in an unsynchronized directional* *line network is*

$$T_{us}(\boldsymbol{\nu}) = \max_{1 \le i \le n-1} (2i - 3 + \nu_i + 2 \sum_{j \ge i+1}^{n} \nu_j) \tag{2.12}$$

Proof. The proof follows a similar argument as the one used to prove Theorem 2.2.1. □

The worst-case performance for the previous example is illustrated in Fig. 2.8. Delay becomes 11 TS. In conclusion the data collection time is quite sensitive to variation in clock synchronization. Our analysis shows indeed worst-case performance degradation in the order of 50 %.

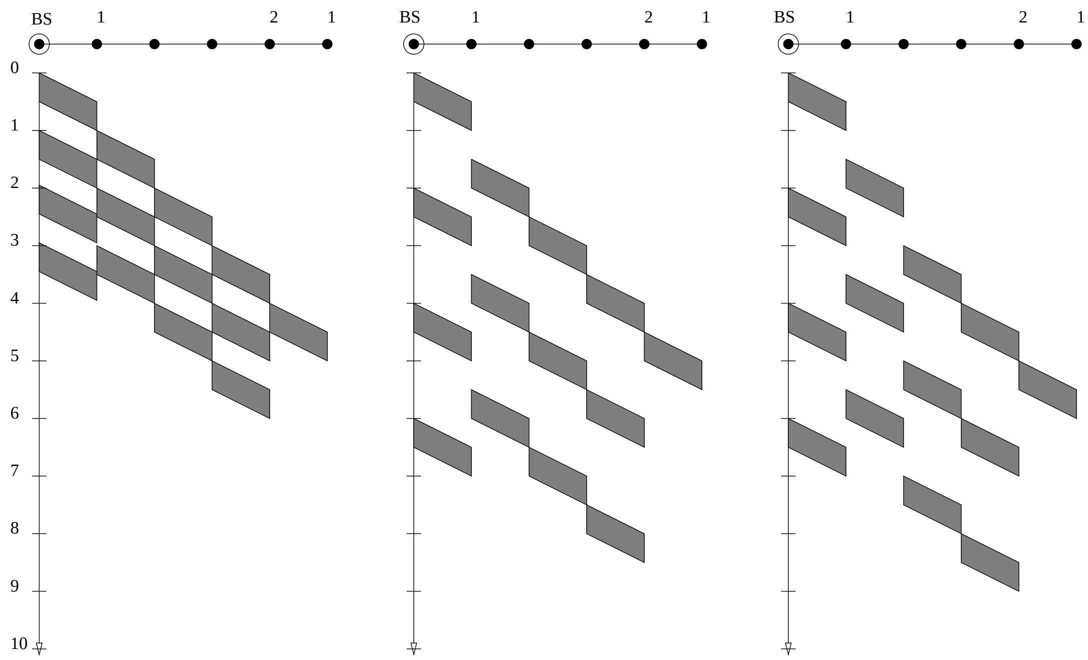


Figure 2.7: Distribution schedules in synchronized and unsynchronized linenetwork.

2.2.4 2-line Networks

Consider now a line network and place the BS anywhere on that line. This may be seen as a 2-line network $(\boldsymbol{\mu}, \boldsymbol{\nu})$. We denote by $T_u(\boldsymbol{\mu}, \boldsymbol{\nu},)$ the optimal performance achievable on a 2-line network. The scheduling procedure, a particular case of the

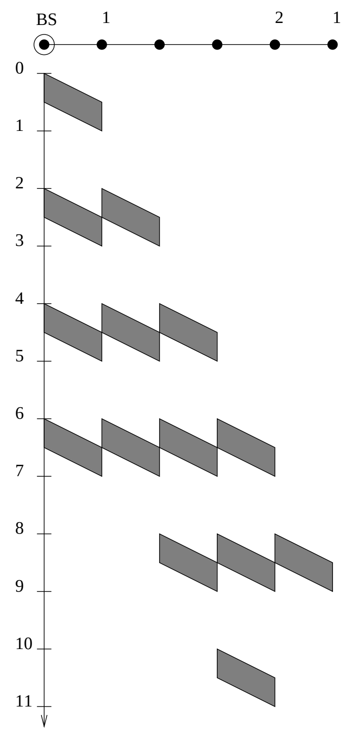


Figure 2.8: Worst-case out-of-sync distribution schedule in line network.

multi-line algorithm described in the next section, is illustrated in the example of Fig. 2.9.

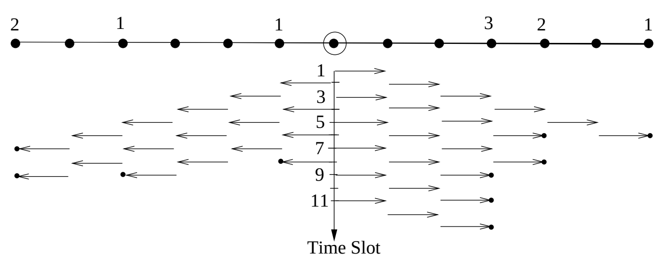


Figure 2.9: Optimal distribution schedule in 2-line sensor network.

Theorem 2.2.6. *The minimum collection time on a directional 2-line network* $(\boldsymbol{\mu}, \boldsymbol{\nu})$

is

$$T_u(\boldsymbol{\mu},\boldsymbol{\nu}) = \begin{cases} \max(T_u(\boldsymbol{\nu})+1, \sum_{i\geq 1}\mu_i+\nu_i) & \textit{if } T_u(\boldsymbol{\mu}) = T_u(\boldsymbol{\nu}) \\ \max(T_u(\boldsymbol{\nu}), \sum_{i\geq 1}\mu_i+\nu_i) & \textit{if } T_u(\boldsymbol{\nu}) > T_u(\boldsymbol{\mu}) \end{cases} \tag{2.13}$$

Proof.

$$T_u(\boldsymbol{\mu}) = T_u(\boldsymbol{\nu}) \Rightarrow T_u(\boldsymbol{\mu},\boldsymbol{\nu}) \geq \max(T_u(\boldsymbol{\mu})+1, \sum_{i\geq 1}\mu_i+\nu_i))$$

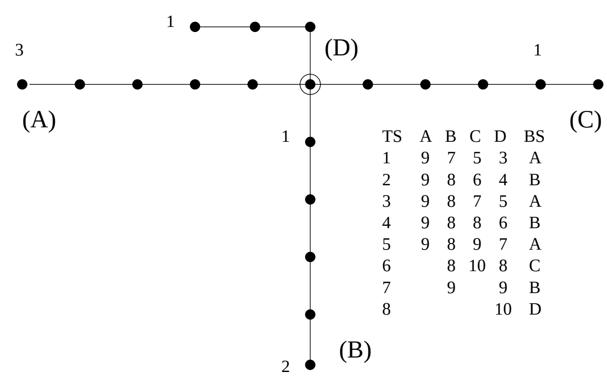
$$T_u(\boldsymbol{\mu}) > T_u(\boldsymbol{\nu}) \Rightarrow T_u(\boldsymbol{\mu},\boldsymbol{\nu}) \geq \max(T_u(\boldsymbol{\mu}), \sum_{i\geq 1}\mu_i+\nu_i))$$

It is easy to see why the above described algorithm achieves this lower bound. Consider for example the case $T_u(\boldsymbol{\mu}) = T_u(\boldsymbol{\nu})$. Either the algorithm takes $T(\boldsymbol{\mu})+1$ TS to perform the job or it takes T'_{-1} (resp. T'_1) defined as the last busy TS at distance 1 to the left (resp. to the right) from the BS. If it so T'_{-1} (resp. T'_1) equals $\sum_{i\geq 1}\mu_i+\nu_i)$. □

2.2.5 Multi-line Networks

In this section we consider multi-line networks, by which we mean multiple line of sensors meeting in one single point, the BS. Fig. 2.10 and Fig. 2.9 are examples of such networks. We describe an algorithm for distributing data in these networks. The algorithm (listed as Algorithm 2 in Appendix B), running at the BS, determines at each TS toward which line to transmit, if transmission is possible at all. The direction of transmission is greedily decided, based on estimates (one per line) of the completion time of the data transfer. Initial estimate for a given line is determined by Eq. (2.3). The legal direction associated with the biggest estimate is chosen (a legal transmission is one that respects the channel reuse constraints, so, for example, it is not legal for our algorithm to transmit in two successive TS toward a given node located at distance greater than 1 from the BS), ties being broken randomly. When no legal direction exists the BS remains idle. After a decision has been made (transmit toward a particular direction or stay idle) the estimates at each line are updated according to the following rule. If a legal direction was not chosen, its new estimate becomes its old estimate plus one. Illegal direction estimates remain unchanged. The idea is to minimize at each TS the overall estimate of the transmission time.

We illustrate the procedure on an example in Fig. 2.10. In the accompanying table, we list data transfer completion time estimates at each TS and the corresponding decision made by the BS. As previously stated the initial completion time estimates are computed using Eq. (2.3). The table reads as follows. TS 1: All 4 transmission directions are legal. The BS chooses to transmit toward line A. At TS 2, transmitting toward A is not a legal move, the legal transmission direction associated with the biggest estimate is B, etc. Along a given line, the packets destined for furthest nodes are sent first by the BS. As for the other nodes they merely forward the data packets of which they are not recipients (a packet is transmitted in the following TS that it was received). In this example the algorithm performance is 10 TS.



TS	A	B	C	D	BS
1	9	7	5	3	A
2	9	8	6	4	B
3	9	8	7	5	A
4	9	8	8	6	B
5	9	8	9	7	A
6		8	10	8	C
7		9		9	B
8				10	D

Figure 2.10: Optimal distribution schedule for BS in 4-line sensor network.

Theorem 2.2.7. *Algorithm 2 is optimal.*

Proof. We note that equivalently this algorithm picks at each TS the legal direction B_i that maximizes $T(\boldsymbol{\nu}^i)$ (that quantity being updated at each TS to take into account the packets delivered).

We first introduce a few notations and definitions: Let N denote the considered network for which one wishes to derive an optimal schedule. Let N_e denote the "equivalent" network to N (see following definition). Let P denote the considered problem

of scheduling data transfers to the nodes. Let P' denote the same problem under a relaxed set of conditions, namely that simultaneous transmission and reception (of different data packets) are allowed in a single TS at any given node. This problem is independently studied in Appendix A. Let $S(P, N)$ denote a schedule for problem P and network N. Let $S_{|BS}(P, N)$ denote the schedule of the BS derived from $S(P, N)$. Let $S^{opt}(P, N)$ denote an optimal schedule for (P, N).
In the "equivalent" network N_e of Network N the data packets along a particular line are redistributed along the corresponding line in N_e in the manner illustrated in Fig. 2.11.

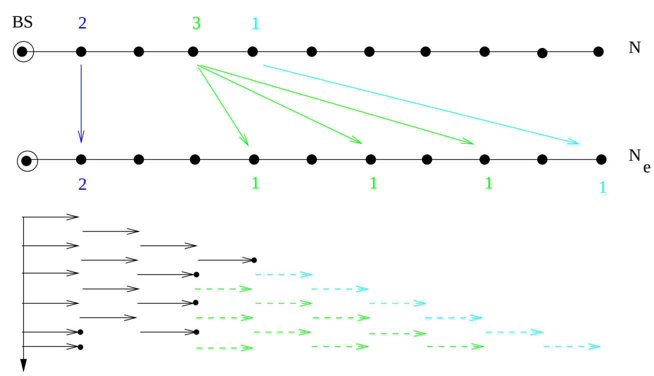


Figure 2.11: "Equivalent" network construction. Distribution schedules in network N (solid arrows) and corresponding equivalent network N_e (solid arrows + dashed arrows).

Although packets in N and N_e are distributed differently over the network, we shall see that the data collection is the same for both networks. It is in that sense that they are "equivalent." Formally the construction is as follows. To each line of N, say B_k, if ν_i^k denotes the number of data packets at distance i from the BS along B_k, $T(\boldsymbol{\nu}^k)$ denotes the length of an optimal schedule for that particular line, and T_i^k is the last busy TS at node i in the execution of Algorithm 1 for that line, associate a line in N_e, say B'_k such that, if $\nu_i'^k$ denotes the number of data packets at distance

i from the BS along B'_k, then

$$\begin{aligned} i = 1 \quad & \nu'^k_1 = \nu^k_1 \\ i \geq 2 \quad & \nu'^k_{T(\boldsymbol{\nu}^k) - T^k_i + 2j + i} = 1 \quad for \quad 0 \leq j \leq l - 1 \quad if\, \nu^k_i = l \geq 1 \\ & \nu'^k_i = 0 \quad otherwise \end{aligned} \tag{2.14}$$

By construction, N_e has the following characteristics:

- Same total number of data packets as N, same number of data packets per line, same number of lines.
- Each line carries the same workload as its corresponding line in N (i.e., $\forall k$, $T(\boldsymbol{\nu}'^k) = T(\boldsymbol{\nu}^k)$).
- Node $i > 1$ carries 0 or 1 data packet.
- Two nodes with data packets are separated by at least one node with no data packet.

Example: Consider the following 2-line network N: $B_1 : \boldsymbol{\nu^1} = (0, 4)$, $\quad B_2 : \boldsymbol{\nu^2} = (2, 0)$. Its equivalent network N_e is a 2-line network such that:
$B'_1 : \boldsymbol{\nu'^1} = (0, 1, 0, 1, 0, 1, 0, 1)$, $B'_2 : \boldsymbol{\nu'^2} = (2, 0) = \boldsymbol{\nu^2}$

Lemma 2.2.8. *There exists an optimal schedule for problem P' and network N_e.*

Proof. A construction of such a schedule is given in Appendix A. □

Lemma 2.2.9. *Let $S^{opt}(P', N_e)$ denote the optimal schedule constructed in Appendix A for problem P' and network N_e. It is possible to construct a schedule $S(P, N_e)$ for problem P and network N_e from $S^{opt}(P', N_e)$ by judiciously reordering the BS transmissions such that the two schedules have the same length.*

Proof. With the convention that furthest nodes should be served first along a given line, a schedule $S(P/P', N_e)$ is entirely defined by its restriction to the BS schedule $S_{|BS}(P/P', N_e)$. The BS schedule being a sequence of directions B_i corresponding to the lines toward which transmit at each TS as well as possible silences (corresponding to BS being idle). $S_{|BS}(P', N_e) = (B'_1, B'_1, B'_1, B'_1, B'_2, B'_2, -, -)$ is an instance of an optimal schedule for problem P' and the network described in previous example where "-" denotes a silence.

We construct $S(P, N_e)$ from $S^{opt}(P', N_e)$ by iteratively applying the following operation on S: $insert(i, j)(S)$ for $j > i \geq 1$ which returns a schedule S' where element j in schedule S was inserted between element i and $i+1$ in S. In the previous example $insert(1, 5)(S_{|BS}(P', N_e)) = (B'_1, B'_2, B'_1, B'_1, B'_1, B'_2, -, -)$.

This operation doesn't change the length of S, that is $Length(S') = Length(S)$ as long as it is not applied more than once for any i. This is a direct consequence of the fourth characteristic of an equivalent network. Next we describe the construction.

If S is a valid schedule (i.e., satisfying constraint P) we are done. Otherwise assume the first conflict occurs in position i_0 of schedule S (that is constraint P does not allow for transmission toward element i_0 followed by transmission toward element $i_0 + 1$). In the instance above, there are conflicts in $i_0 = 1, 2, 3, 5$. Further assume the first direction distinct from the one in position i_0 and that follows it is element i_1 of S. If there is no such direction then denote i_1 the position of the first silence following (it always exists by definition of N_e). Then apply $insert(i_0, i_1)(S)$. Clearly the procedure produces a new schedule S of same length. Thus the portion of the schedule S comprised between element 1 and $i_0 + 1$ satisfies P. Repeat until the schedule S satisfies constraint P. Since the number of initial conflicts is finite, this procedure ends in a finite number of steps. In the previous example these operations are in order: $insert(1, 5)$, $insert(3, 6)$ and $insert(5, 7)$. They lead to the schedule: $S_{|BS}(P, N_e)) = (B'_1, B'_2, B'_1, B'_2, B'_1, -, B'_1, -)$ □

By lemma 2.2.9, the lengths of $S(P, N_e)$ and $S^{opt}(P', N_e)$ are the same. Thus $S(P, N_e)$ is optimal. Denote it $S^{opt}(P, N_e)$. One may construct a schedule $S(P, N)$ from $S^{opt}(P, N_e)$ such that the lengths of the two schedules are the same and $S_{|BS}(P, N)$ $= S_{|BS}(P, N_e)$. If $S(P, N)$ is not optimal then there exists a schedule $S'(P, N)$ such that the length of $S'(P, N)$ is less than the length of $S(P, N)$. But from $S'(P, N)$ one may construct a schedule $S'(P, N_e)$ such that the two schedules have the same length and $S'_{|BS}(P, N_e) = S'_{|BS}(P, N)$ so the length of $S'(P, N_e)$ is less than the length of $S^{opt}(P, N_e)$, a contradiction. Thus $S(P, N)$ is optimal, which concludes the proof of Theorem 2.2.7. □

2.2.6 Tree Networks, Case Where Base Station Degree Is 1

Throughout this paragraph we assume that the degree of the root of the considered graphs is one. We define the *equivalent linear network* $(G_l, E_l, \boldsymbol{\nu_l})$ of a network $(G, E, \boldsymbol{\nu})$: If $G = \{N_0, N_1, \ldots, N_n\}$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)$ then $G_l = \{0, 1, \ldots, m\}$, $E_l = \{(i-1, i), 1 \le i \le m\}$ and $\boldsymbol{\nu_l} = (\nu_{l1}, \ldots, \nu_{lm})$ where $m = \max_i(d(N_0, N_i))$ and $\nu_{lj} = \sum_{i \,|\, d(N_0,N_i)=j} \nu_i$ We illustrate a tree network in Fig. 2.12 ($n = 14$, $m = 7$); its equivalent linear network is shown in Fig. 2.4.

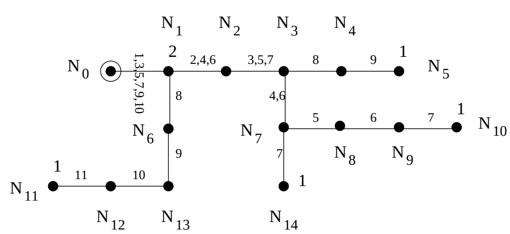


Figure 2.12: A 15-node tree network with degree of BS=1, the equivalent linear network is drawn in Fig. 2.4. Transmission time steps are written next to the edges.

The equivalent linear network's schedule may serve as a schedule for the initial tree network. Next we explain how transmission time slots for $(G_l, E_l, \boldsymbol{\nu_l})$ (determined by running algorithm 1) may be mapped onto $(G, E, \boldsymbol{\nu})$. Consider an element in E, say (N_{i_0}, N_{j_0}), such that $d(N_0, N_{i_0}) = \alpha$ (hops). Based on the number of data packets N_{j_0} has to forward, say f_{j_0}, we shall allocate transmission time slots to edge (N_{i_0}, N_{j_0}). Define $E_\alpha = \{(N_i, N_j) \in E \,|\, d(N_0, N_i) = \alpha\}$. Each packet P follows a path $path(P)$ from the BS to its destination node where $path(P)$ denotes the finite sequence of edges $(e_1, ..., e_k)$ traversed in that order by P. For convenience we shall write $path(P)$ as the sequence of vertices $(vertices(e_1), ..., vertices(e_k))$. We define

$\mathfrak{P}_\alpha = \{P \,|\, \exists e \in E_\alpha \cap path(P)\}$. We define $\mathfrak{T}_\alpha = \{\text{TS used by } (\alpha, \alpha+1) \in E_l\}$. We have $|\mathfrak{P}_\alpha| = \sum_{(N_i,N_j)\in E_\alpha} (\nu_j + f_j) = \sum_{k>\alpha} \nu_{lk} = |\mathfrak{T}_\alpha|$. Thus one may define a one-to-one correspondence g between $\mathfrak{P}_\alpha$ and $\mathfrak{T}_\alpha$ that associates the packet P with the longest path in $\mathfrak{P}_\alpha$, with the TS with the smallest index in $\mathfrak{T}_\alpha$; the packet P with second longest path, with the TS with second smallest index and so on. We finally define $\mathfrak{P}_\alpha^{(N_{i_0},N_{j_0})} = \{P \,|\, (N_{i_0}, N_{j_0}) \in path(P)\} \subseteq \mathfrak{P}_\alpha$. (N_{i_0}, N_{j_0}) is associated with time slots $g(\mathfrak{P}_\alpha^{(N_{i_0},N_{j_0})})$. In the example of Fig. 2.12, we have $\{P\} = \{P_1, P_2, \ldots, P_6\}$ where the first packet is characterized by $path(P_1) = (N_0,N_1,N_2,N_3,N_7,N_8,N_9,N_{10})$, the second one by $path(P_2) = (N_0,N_1,N_2,N_3,N_4,N_5)$, the third one by $path(P_3) = (N_0,N_1,N_6,N_{13},N_{12},N_{11})$, the fourth one by $path(P_4) = (N_0,N_1,N_2,N_3,N_7,N_{14})$, and finally the fifth and sixth ones by $path(P_5) = path(P_6) = (N_0,N_1)$. We also have $E_1 = \{(N_1,N_2),(N_1,N_6)\}$, $\mathfrak{P}_1 = \{P_1,P_2,P_3,P_4\}$, $\mathfrak{T}_1 = \{2,4,6,8\}$, and $\mathfrak{P}_1^{(N_1,N_2)} = \{P_1,P_2,P_4\}$. Thus edge (N_1,N_2) is associated with time slots $g(\mathfrak{P}_1^{(N_1,N_2)}) = \{2,4,6\}$. Thus Algorithm 1 run on the equivalent linear network provides a BS transmission schedule. Intermediate nodes simply forward data packets to further nodes as they arrive (in the TS following their arrival). This requires a routing table at junction nodes. In a centralized version of this algorithm nodes may be informed of their transmission slots. Fig. 2.12 shows such a mapping for the considered example.

Although an equivalent linear network has a reduced set of possible concurrent transmissions, this procedure produces an optimal transmission schedule. This follows from the following lemma.

Lemma 2.2.10. *Given any connected graph G such that degree of BS is one, if $t_2(G)$ denotes the time performance of a given data distribution algorithm, and ν_j denotes the number of data packets at distance j from the BS, then*

$$t_2(G) \geq \max_i (i - 1 + \nu_i + 2\sum_{j>i} \nu_j) \tag{2.15}$$

Proof. $\sum_{j\geq 1} \nu_j$ data packets must be delivered to nodes at distance greater than 1. Therefore link (0,1) is activated $\sum_{j\geq 1} \nu_j$ times and links (1,2) (all edges from a node at distance 1 from the BS to a node at distance 2) are activated $\sum_{j\geq 2} \nu_j$ times but link (0,1) can not be activated at the same time as a link (1,2), thus we have $t_2(G) \geq \sum_{j\geq 1} \nu_j + \sum_{j\geq 2} \nu_j$.

$\sum_{j\geq i} \nu_j$ data packets must be delivered to nodes at distance greater than $i > 1$. Therefore edge (0,1) is activated at least $\sum_{j\geq i} \nu_j$ times and edges (1,2) $\sum_{j>i} \nu_j$ times but link (0,1) can not be activated at the same time as a link (1,2), moreover after $\sum_{j\geq i} \nu_j + \sum_{j>i} \nu_j$ TS the last data packet sent by the BS is at distance 0 or 1 from the BS if $\nu_i > 0$ and at distance 0, 1 or 2 from the BS if $\nu_i = 0$. Indeed it takes a minimum of $2\sum_{j\geq i} \nu_j$ TS to get all the data packets out of the positions 0,1,2. Thus after $\sum_{j\geq i} \nu_j + \sum_{j>i} \nu_j$ TS whether $\nu_i > 0$ or $\nu_i = 0$ one data packet is at least $i-1$ hops away from its destination, therefore: $t_2(G) \geq \sum_{j\geq i} \nu_j + \sum_{j>i} \nu_j + i - 1$. Hence the stated result. □

2.2.7 Tree Sensor Networks, General Case

The results in the previous sections suggest the following algorithm for dealing with general tree networks.

1. Linearize the subtrees attached to the BS (with BS degree equal to 1) according to the procedure described in section 2.2.6.

2. Apply multi-line algorithm described in section 2.2.5 to the resulting multi-line system.

This procedure produces an optimal schedule. This results from Theorem 2.2.7 and Lemma 2.2.10.

Theorem 2.2.11. *If $\mathcal{T}$ is a tree network and ν_j^k denotes the number of data packets at distance j from the BS along branch k, then the minimum data collection time over $\mathcal{T}$ is*

$$T_u(\mathcal{T}) = \max_{1\leq i\leq n} (i - 1 + \sum_{j\geq i}^{n} \nu'_j) \tag{2.16}$$

where $\nu'_j = \sum_k \nu'^k_j$ and ν'^k_j is obtained from ν_j^k by equation (2.14).

Proof. This follows directly from Lemma 2.2.11. □

2.2.8 Networks with Cycles

We propose a data distribution/collection strategy on general graphs. However that strategy is not optimal in general. In this section we prove that our algorithm performs within a factor of 2 of an optimal strategy. The proposed strategy consists of two subprocedures:

1. Extract a shortest path spanning tree $\mathcal{T}_{SP}$.

2. Apply previously described distribution strategy on trees to $\mathcal{T}_{SP}$.

Note: one can show that shortest path spanning trees always exist by using Dijkstra algorithm. The following theorem provides a motivation for choosing a shortest path spanning tree. The proof follows from Theorem 2.2.11.

Theorem 2.2.12. *For any (connected) graph G, for any spanning tree $\mathcal{T}$ of G and for any shortest path spanning tree $\mathcal{T}_{SP}$ of G, the minimum data collection time over network $\mathcal{T}$, $T_u(\mathcal{T})$ satisfies*

$$T_u(\mathcal{T}_{SP}) \leq T_u(\mathcal{T}) \tag{2.17}$$

Theorem 2.2.13. *For any (connected) graph G, and any shortest path spanning tree $\mathcal{T}_{SP}$ we have*

$$\frac{T_u(\mathcal{T}_{SP})}{2} \leq T_u(G) \leq T_u(\mathcal{T}_{SP}) \tag{2.18}$$

Proof. The second inequality is clear. For a proof of the first inequality we define: $t_1(G)$ the minimum distribution time when transmission and reception are simultaneously allowed in a TS at any given node. Clearly $t_1(G) \leq T_u(G)$. By corollary A.0.4 we also have $t_1(G) = t_1(\mathcal{T}_{SP})$. Besides for any connected graph A the following inequality holds: $T_u(A) \leq 2t_1(A)$. Choose $A = \mathcal{T}_{SP}$, the inequality follows. □

These bounds are tight. The upper bound is achieved when $G = \mathcal{T}_{SP}$. As for the lower bound consider the following network G where n data packets are stored at distance k hops from the BS in node x. Further assume there are two distinct paths of length k from the BS to x.

$\mathcal{T}_{SP}$ is the line network $\boldsymbol{\nu} = (\overbrace{0, 0, ..., 0}^{k-1}, n)$. We have $T_u(G) = n + k - 1$ (for $k \geq 1$) and $T_u(\mathcal{T}_{SP}) = 2n + k - 2$ (for $k \geq 2$), thus $T_u(G)$ converges toward $T_u(\mathcal{T}_{SP})/2$ when n goes to infinity (for $k \geq 2$).

Bounds on $T_u(G)$ can also be written in the following more explicit way.

Theorem 2.2.14. *The minimum data collection time over a graph G satisfies*

$$\max_i (i - 1 + \sum_{j \geq i} \nu_j) \leq T_u(G) \leq \max_i (i - 1 + \nu_i + 2 \sum_{j \geq i+1} \nu_j) \tag{2.19}$$

Proof. We have from corollary A.0.4 $t_1(G) = \max_i (i - 1 + \sum_{j \geq i} \nu_j)$ □

Both bounds on $T_u(G)$ are achievable. The lower bound for instance is achieved in the previously considered example where $\max_i (i - 1 + \sum_{j \geq i} \nu_j) = n + k - 1$.

2.3 Omnidirectional Antenna Systems

Results on directional antenna systems may be to some extent adapted to omnidirectional antenna systems. This is the purpose of this section.

2.3.1 Line Networks

Our results readily extend to omnidirectional antenna systems. The procedure is illustrated in the example of Fig. 2.13 where $V = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $E = \{(i, i+1), 0 \leq i \leq 8\}$, $\boldsymbol{\nu} = (2, 1, 0, 0, 0, 0, 0, 1, 1)$, $d < r < 2d$, $(1 + \delta)r < 2d$.

The schedule of transmissions, as determined by Algorithm 3 in appendix B, is drawn below the network (upper schedule) for the distribution problem. It is performed in 11 TS.

Next we determine the performance of our algorithm in general. Denote T_i the last busy time slot at node i, $1 \leq i \leq n$ in the execution of our distribution algorithm (In the previous example, we have $T_1 = 10, T_2 = 8, T_3 = 7, T_4 = 8, T_5 = 9, T_6 = 10, T_7 = 11, T_8 = 11, T_9 = 9$). Clearly then our algorithm runs in $\max_{1 \leq i \leq n} \{Ti\}$. T_i is a function

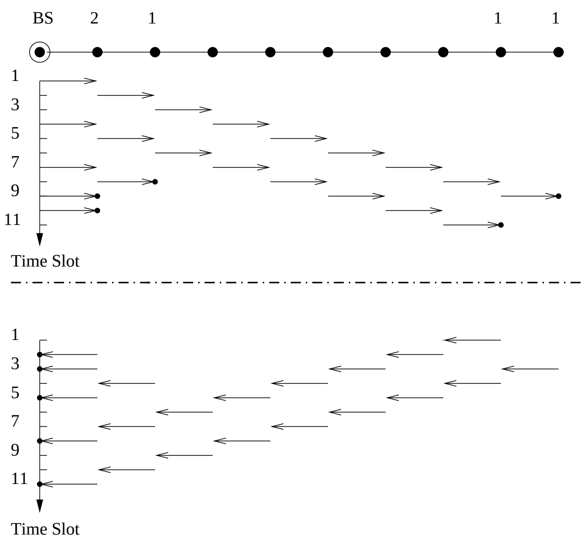


Figure 2.13: Optimal distribution and collection schedules in 10-node line network equipped with omnidirectional antennas.

of the distance to the BS, the number of data packets destined for node i (that is ν_i) and the number of data packets forwarded by node i.

Assuming $\nu_i = 0$ for $i > n$, we have

$$T_1 = \begin{cases} 3\sum_{j\geq 3}\nu_j - 1 & \text{if } \nu_1 = 0,\ \nu_2 = 0 \text{ and } \sum_{j\geq 3}\nu_j \geq 1 \\ \nu_1 + 2\nu_2 + 3\sum_{j\geq 3}\nu_j & \text{otherwise} \end{cases}$$

$$T_2 = 2\nu_2 + 3\sum_{j\geq 3}\nu_j$$

$\forall i \geq 3$

$$T_i = \begin{cases} i - 2 + 3\sum_{j>i}\nu_j & \text{if } \nu_i = 0 \text{ and } \sum_{j>i}\nu_j \geq 1 \\ i + 3\sum_{j>i}\nu_j & \text{if } \nu_i = 1 \\ i - 3 + 3\sum_{j\geq i}\nu_j & \text{if } \nu_i \geq 2 \end{cases} \tag{2.20}$$

Proof. Denote by f_i the number of data packets forwarded by node i.

If $i = 1$,

$$\nu_1 = 0, \nu_2 = 0, f_i \geq 1 \Rightarrow T_i = 3(f_i - 1) + 2 + (i - 1)$$

$$\text{otherwise}, T_i = \nu_1 + 2\nu_2 + 3(f_i - \nu_2)$$

If $i = 2$,

$$T_i = 2\nu_2 + 3(f_i - \nu_2)$$

$\forall i \geq 3$,

$$\nu_i = 0, f_i \geq 1 \Rightarrow T_i = 3(f_i - 1) + 2 + (i - 1)$$

$$\nu_1 \geq 1 \Rightarrow T_i = 3f_i + 1 + (i - 1)$$

$$\nu_i \geq 2 \Rightarrow T_i = 3f_i + 3(\nu_i - 1) + 1 + (i - 1)$$

but,

$$f_i = \sum_{j>i} \nu_j$$

hence the stated result. □

Clearly the maximum of T_i is obtained over the set $\{i \geq 1 \,|\, \nu_i \neq 0\}$. We define, for a given sensor network, $T_o(\boldsymbol{\nu})$ the minimum length of a time schedule over all time schedules that perform the distribution job. Thus we have the following result.

$$T_o(\boldsymbol{\nu}) \leq \max_{\{i \geq 1 \,|\, \nu_i \neq 0\}} T_i \tag{2.21}$$

Let's now derive a lower bound on $T_o(\boldsymbol{\nu})$. Assuming $\nu_i = 0$ for $i > n$, we have

$$T_o(\boldsymbol{\nu}) \geq \max_{1 \leq i \leq n} (i - 1 + \nu_i + 2\nu_{i+1} + 3 \sum_{j \geq i+2} \nu_j) \tag{2.22}$$

Proof. Consider node $i \geq 1$, assume there exists $k \geq i$ such that $\nu_k \geq 1$. Then

- edge $(i-1, i)$ is activated $\sum_{j \geq i} \nu_j$ TS.
- edge $(i, i+1)$-if it exists- is activated $\sum_{j \geq i+1} \nu_j$ TS.
- edge $(i+1, i+2)$-if it exists- is activated $\sum_{j \geq i+2} \nu_j$ TS.

Clearly transmissions $i-1 \rightarrow i$, $i \rightarrow i+1$, $i+1 \rightarrow i+2$, $\forall i \geq 1$ may not occur concurrently (channel reuse constraints). Besides from our initial assumptions we know that idle time of nodes $\in \{i, i+1, i+2\} \geq i-1$. Therefore,

$$T_o(\boldsymbol{\nu}) \geq \sum_{j \geq i} \nu_j + \sum_{j \geq i+1} \nu_j + \sum_{j \geq i+2} \nu_j + (i-1) \triangleq S_i$$

We have $\forall i, \quad T_o(\boldsymbol{\nu}) \geq S_i$, thus $T_o(\boldsymbol{\nu}) \geq \max_i S_i$. □

Next we prove that the lower bounds and upper bounds previously derived on $T_o(\mathbf{p})$ are in fact equal and hence that the proposed schedule is optimal.

Theorem 2.3.1. *Assuming $\nu_i = 0$ for $i > n$, we have that the minimum data collection time, in the line network $\boldsymbol{\nu}$ of length n*[2] *equipped with omnidirectional antennas, is*

$$T_o(\boldsymbol{\nu}) = \max_{1 \leq i \leq n} (i - 1 + \nu_i + 2\nu_{i+1} + 3 \sum_{j \geq i+2} \nu_j) \tag{2.23}$$

Proof. Assume there exists j such that $\forall i \neq j$, $T_j \geq T_i\,, T_{j+1} < T_j$

- if $j = 1 \Rightarrow S_1 \geq T_1 \Rightarrow T_1 = S_1$
- if $j = 2 \Rightarrow \nu_2 \geq 1$, $\nu_1 = 0 \Rightarrow T_2 - S_2 = \nu_2 + \nu_3 - 1 \geq 0 \Rightarrow T_2 \geq S_2$
  $\nu_1 = 0 \Rightarrow T_1 = T_2 \Rightarrow S_1 \geq T_1$
- if $j \geq 3 \Rightarrow \nu_{j-2} = 0$, $\nu_{j-1} = 0$, $\nu_j \geq 1$
  $S_{j-2} = j - 3 + 3\sum_{i \geq j} \nu_i$
  $\nu_j = 1 \Rightarrow T_j = S_{j-2}$
  $\nu_j \geq 2 \Rightarrow T_j = S_{j-2}$

□

Corollary 2.3.2. *In the particular case where no three consecutive components of vector $\boldsymbol{\nu}$ equal zero, Eq. (2.23) reduces to*

$$T_o(\boldsymbol{\nu}) = \nu_1 + 2\nu_2 + 3 \sum_{i \geq 3} \nu_i \tag{2.24}$$

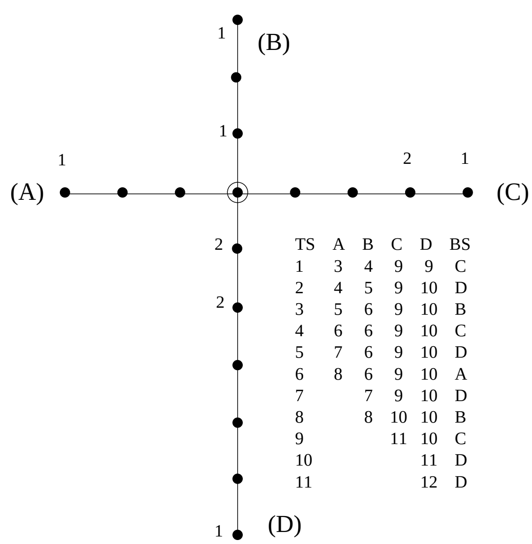
[2] *Implicitly we assume that the distance to the BS of the furthest node carrying a packet is n.*

Again the construction of a schedule for the data collection problem is based on the symmetry of the operations of distribution and collection.

Theorem 2.3.3. *The minimum data collection time over an omnidirectional line network $\boldsymbol{\nu}$, assuming the transmission range is 1 hop and the interference range is m hops, is*

$$\forall m \geq 1, \quad T_o^m(\boldsymbol{\nu}) = \max_i \, (i - 1 + \sum_{i \leq j \leq i+m} (j - i + 1)\nu_j + (m+2) \sum_{j \geq i+m+1} \nu_j) \quad (2.25)$$

Proof. The proof follows a similar argument as the one used to prove Theorem 2.2.3. ∎



TS	A	B	C	D	BS
1	3	4	9	9	C
2	4	5	9	10	D
3	5	6	9	10	B
4	6	6	9	10	C
5	7	6	9	10	D
6	8	6	9	10	A
7		7	9	10	D
8		8	10	10	B
9			11	10	C
10				11	D
11				12	D

Figure 2.14: Optimal distribution schedule for BS in 4-line sensor network.

2.3.2 Multi-line Networks

Next we illustrate the procedure to distribute data on a multi-line network on an example (Fig. 2.14). In the accompanying table, we list data transfer completion time estimates at each TS and the corresponding decision made by the BS (as to which direction to choose). As previously stated the initial completion time estimates are computed using Eq. (2.23). The table reads as follows. TS 1: All 4 transmission directions are legal. The BS chooses to transmit toward branch C (it could have chosen D as well, as ties are broken randomly). At TS 2, transmitting toward C is not a legal move, the legal transmission direction associated with the biggest estimate is D (notice that transmitting toward A or B makes the overall completion time estimate be 11 TS, whereas transmitting toward D leaves the completion time estimate unchanged (10 TS), so D is also the legal move that minimizes the estimated completion time), etc. The packets destined for furthest nodes are sent first by the BS. As for the other nodes they merely forward the data packets of which they are not recipients (a packet is transmitted in the following TS that it was received). In this example the algorithm performs in 12 TS (an obvious lower bound on the time performance is 11 TS corresponding to 11 data packets). The previously described algorithm is optimal when the number of data packets at distance 0 and 1 from the BS is zero. If it is not the case, the algorithm needs to be refined, in particular estimates ties should not be broken randomly in general. In this proof we assumed that relay sensor nodes can only perform simple receive and forward type operations in which a data packet is to be forwarded in the TS following its arrival at a relay node. Note that time performance may be further improved, if we assume that nodes have the ability to perform store and forward type operations (that is store data to be relayed). This was not the case for directional antenna systems. This is illustrated in Fig. 2.15. If the simplest relay nodes are being used the completion time is 10 TS, whereas it is as low as 9 TS when the smarter nodes are used. However, in the directional antenna case the time performance is 9 TS either way.

2.3.3 Tree Networks, Case Where Base Station Degree Is 1

Throughout this paragraph we assume that the degree of the root of the considered graphs is one. We define the *equivalent linear network* $(G_l, E_l, \boldsymbol{\nu_l})$ of a network $(G, E, \boldsymbol{\nu})$. If $G = \{N_0, N_1, \ldots, N_n\}$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)$ then $G_l = \{0, 1, \ldots, m \leq n\}$, $E_l = \{(i-1, i), 1 \leq i \leq m\}$ and $\boldsymbol{\nu_l} = (\nu_{l1}, \ldots, nu_{lm})$ where $\nu_{lj} = \sum_{i \,|\, d(N_0, N_i) = j} \nu_i$

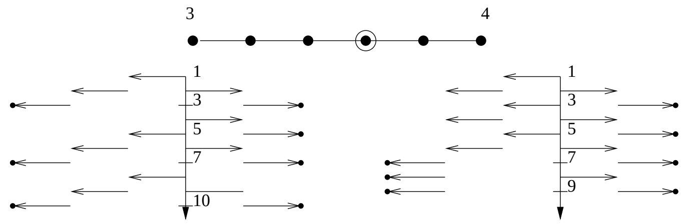


Figure 2.15: Optimal distribution schedules for BS in 2-line sensor network. Simple receive and forward sensor nodes on the right versus store and forward nodes on the left.

This definition is illustrated in Fig. 2.16 ($n = 15$, $m = 9$) and Fig. 2.13 (equivalent line network).

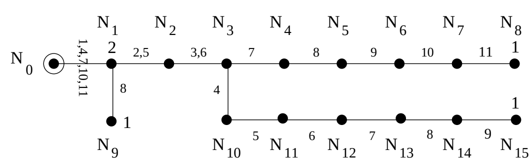


Figure 2.16: A 16-node tree network whose BS degree is 1, the equivalent linear network is drawn in Fig. 2.13. Transmission time steps are written next to the edges.

The equivalent linear network's schedule may serve as a schedule for the initial tree network. Next we explain how transmission time slots for $(G_l, E_l, \boldsymbol{\nu_l})$ (determined by running Algorithm 1) may be mapped onto $(G, E, \boldsymbol{\nu})$. Consider an element in E, say (N_{i_0}, N_{j_0}), such that $d(N_0, N_{i_0}) = \alpha$ (hops). Based on the number of data packets N_{j_0} has to forward, say f_{j_0}, we shall allocate transmission time slots to edge (N_{i_0}, N_{j_0}). Define $E_\alpha = \{(N_i, N_j) \in E \,|\, d(N_0, N_i) = \alpha\}$. Each packet P follows a path $path(P)$ from the BS to its destination node where $path(P)$ denotes the finite sequence of edges $(e_1, ..., e_k)$ traversed in that order by P. For convenience we shall write $path(P)$ as the sequence of vertices $(vertices(e_1), ..., vertices(e_k))$. We define $\mathfrak{P}_\alpha = \{P \,|\, \exists e \in E_\alpha \cap path(P)\}$. We define $\mathfrak{T}_\alpha = \{\text{TS used by } (\alpha, \alpha+1) \in E_l\}$. We have $|\mathfrak{P}_\alpha| = \sum_{(N_i,N_j)\in E_\alpha} (\nu_j + f_j) = \sum_{k>\alpha} \nu_{lk} = |\mathfrak{T}_\alpha|$. Thus one may define a one to one correspondence g between $\mathfrak{P}_\alpha$ and $\mathfrak{T}_\alpha$ that associates the packet P with the longest path in $\mathfrak{P}_\alpha$, with the TS with the smallest index in $\mathfrak{T}_\alpha$; the packet P with second longest path, with the TS with second smallest index and so on. We finally define $\mathfrak{P}_\alpha^{(N_{i_0},N_{j_0})} = \{P \mid (N_{i_0}, N_{j_0}) \in path(P)\} \subseteq \mathfrak{P}_\alpha$. (N_{i_0}, N_{j_0}) is associated with time slots $g(\mathfrak{P}_\alpha^{(N_{i_0},N_{j_0})})$. In the example of Fig. 2.16, we have $\{P\} = \{P_1,P_2,\ldots,P_5\}$ where the first packet is characterized by $path(P_1) = (N_0,N_1,N_2,N_3,N_{10},N_{11},N_{12},N_{13},N_{14},N_{15})$, the second one by $path(P_2) = (N_0,N_1,N_2,N_3,N_4,N_5,N_6,N_7,N_8)$, the third one by $path(P_3) = (N_0,N_1,N_9)$, and finally the fourth and fifth ones by $path(P_4) = path(P_5) = (N_0,N_1)$. We also have $E_1 = \{(N_1,N_2),(N_1,N_9)\}$, $\mathfrak{P}_1 = \{P_1,P_2,P_3\}$, $\mathfrak{T}_1 = \{2,5,8\}$, and $\mathfrak{P}_1^{(N_1,N_2)} = \{P_1,P_2\}$. Thus edge (N_1,N_2) is associated with time slots $g(\mathfrak{P}_1^{(N_1,N_2)}) = \{2,5\}$. Thus Algorithm 1 run on the equivalent linear network provides a BS transmission schedule. Intermediate nodes simply forward data packets to further nodes as they arrive (in the TS following their arrival). This requires a routing table at junction nodes.

Although an equivalent linear network has a reduced set of possible concurrent transmissions, this procedure produces an optimal transmission schedule. The following proof is based on the fact that transmissions that can occur in one case and not in the other are not helpful in routing data faster. This is essentially due to the fact that any route from the BS to a leaf necessarily includes link $(0, 1)$, i.e., from the BS to the unique node at distance one from the BS which constitutes a bottleneck.

Lemma 2.3.4. *Given any tree $\mathcal{T}$ such that degree of BS is one, if $t_3(\mathcal{T})$ denotes the time performance of a given data distribution algorithm, and ν_j denotes the number of data packets at distance j from the BS, then*

$$t_3(\mathcal{T}) \geq \max_i (i - 1 + \nu_i + 2\nu_{i+1} + 3 \sum_{j>i+1} \nu_j) \tag{2.26}$$

Proof. Edges at distance i from the BS are activated $\sum_{j\geq i} \nu_j$ times, edges at distance $i+1$ from the BS are activated $\sum_{j\geq i+1} \nu_j$ times and edges at distance $i+2$ from the BS are activated $\sum_{j\geq i+2} \nu_j$ times. In a given TS, the distance (to the BS) difference of any two data packets in transit is at least 3 hops. This implies in particular that no two edges whose distance difference to the BS is less than or equal to 2 hops may be activated simultaneously. □

In this proof we assumed that relay sensor nodes can only perform simple receive and forward type operations in which a data packet is to be forwarded in the TS following its arrival at a relay node. Note that time performance may be further improved, if we assume that nodes have the ability to perform store and forward type operations (that is store data to be relayed). This, again, was not the case in directional antenna systems. This is illustrated in the following example (Fig. 2.17). If the simplest relay nodes are being used, $t_3(\mathcal{T}) = 6$ TS, whereas $t_3(\mathcal{T}) = 5$ TS may be obtained with the schedule: TS 1: $N_0 \to N_1$, TS 2: $N_1 \to N_2$, TS 3: $N_0 \to N_1$, TS 4: $N_1 \to N_3$, TS 5: $N_1 \to N_2$, $N_3 \to N_4$. However, in the directional antenna case $t_2(\mathcal{T}) = 5$ TS either way.

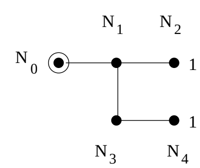


Figure 2.17: 5-node sensor network.

2.3.4 Tree Networks

The procedures described in the previous sections may be combined into a strategy for data distribution/collection on tree networks as follows.

1. Linearize the subtrees attached to the BS (with BS degree equal to 1) according to the procedure described in section 2.3.3

2. Apply multi-line distribution algorithm to the resulting multi-line system as described in section 2.3.2

One can show from previous results that this procedure is optimal on general tree networks in the same way this was proven in the directional antenna case. In the following theorem we give without proof a closed form expression for its time performance.

Time performance on tree networks:

For purpose of deriving the time performance of our strategy on tree networks, we start by defining the equivalent network N_e of a multi-line network N in the following manner: To each line B_k of N and associated data vector $\boldsymbol{\nu}^k$ corresponds a line B'_k in N_e and associated data vector $\boldsymbol{\nu}'^k$ such that

$$\begin{array}{ll} i = 1 & \nu_1'^k = \nu_1^k \\ i = 2 & \nu'^k_{T_o(\boldsymbol{\nu}^k) - T_i^k + 2j + i} = 1 \text{ for } \quad 0 \leq j \leq l-1 \quad \text{if} \quad \nu_i^k = l \geq 1 \\ i \geq 3 & \nu'^k_{T_o(\boldsymbol{\nu}^k) - T_i^k + 3j + i} = 1 \text{ for } \quad 0 \leq j \leq l-1 \quad \text{if} \quad \nu_i^k = l \geq 1 \\ & \nu_i'^k = 0 \quad \text{otherwise} \end{array} \tag{2.27}$$

Theorem 2.3.5. *If $\mathcal{T}$ is a tree and ν_j^k denotes the number of data packets at distance j from the BS along branch k, then, if $\nu_0 = \nu_1 = 0$,*

$$t_3(\mathcal{T}) = \max_i \left(i - 1 + \sum_{j \geq i} \nu'_j\right) \tag{2.28}$$

where $\nu'_j = \sum_k \nu_j'^k$ and $\nu_j'^k$ is obtained from ν_j^k by equation (2.27).

Proof. This follows from results in Appendix A and the proof of optimality of the strategy on multi-line networks, which is similar to the one in the directional antenna case. □

2.3.5 General Connected Sensor Networks

For purpose of analyzing the time performance of data distribution algorithms on general sensor networks we denote by $\mathcal{T}_{SP}(G)$ a shortest path spanning tree of the underlying network graph G. Note that one can show that shortest path spanning trees always exist by using Dijkstra algorithm. Such a tree may not be unique. The following theorem provides a motivation for choosing a shortest path spanning tree.

Theorem 2.3.6. $\forall \mathcal{T}$, *a spanning tree of* G

$$T_o(\mathcal{T}_{SP}) \leq T_o(\mathcal{T}) \tag{2.29}$$

The presence of cycles in a network G will affect the optimal time performance of distributions algorithms as compared with the optimal time performance over $\mathcal{T}_{SP}(G)$. Subsequently we attempt to quantify this phenomenon as well as giving some simple procedures to distribute data over G.
First we note that cycles may help or hurt the time performance of the optimal scheduling strategy in omnidirectional systems (in contrast with directional systems). That is $T_o(G)$ may be larger or smaller than $T_o(\mathcal{T}_{\mathcal{SP}})$ as shown in the examples of Figs. 2.18 and 2.19.

Theorem 2.3.7. *For any (connected) graph* G*, and any shortest path spanning tree* $\mathcal{T}_{SP}$

$$\frac{T_o(\mathcal{T}_{SP})}{3} \leq T_o(G) \tag{2.30}$$

Proof. Define: $t_1(G)$ the minimum distribution time when transmission and reception are simultaneously allowed in a TS at any given node. Clearly $t_1(G) \leq T_o(G)$. By corollary A.0.4 we also have: $t_1(G) = t_1(\mathcal{T}_{SP})$. Besides for any connected graph A the following inequality holds: $T_o(A) \leq 3t_1(A)$. Choose $A = \mathcal{T}_{SP}$, the inequality follows. □

Let us next give an example where the lower bound is achieved. Consider a network G where n data packets are stored at distance k hops from the BS in node x. Further assume there are three distinct paths of length k from x to BS (see Fig. 2.18 where $n = 5$, $k = 6$).

For all practical purposes, T_{SP} is the line network $\boldsymbol{\nu} = (\overbrace{0, 0, ..., 0}^{k-1}, n)$. We have $T_o(G) = n + k - 1$ (for $k \geq 1$) and $T_o(\mathcal{T}_{SP}) = 3n + k - 3$ (for $k \geq 3$), thus $T_o(G)$ converges toward $T_o(\mathcal{T}_{SP})/3$ when n goes to infinity (for $k \geq 3$).

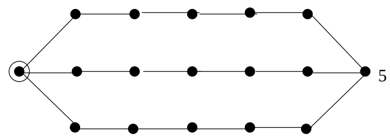


Figure 2.18: Network with cycles. $T_o(G) = 10$ TS, $T_o(\mathcal{T}_{SP}) = 18$ TS.

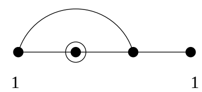


Figure 2.19: Network with cycles. $T_o(G) = 3$ TS, $T_o(\mathcal{T}_{SP}) = 2$ TS.

Strategy and Time performance:

A mere generalization of the strategy proposed for directional antenna systems, based on extracting a shortest path spanning tree of the sensor network, is not envisageable here, as such as an operation is not *physically* possible when nodes are equipped with omnidirectional antennas. We propose to transmit each data packet to its destination along any shortest path between the BS and its destination. An intermediate node will forward a data packet in the TS following its arrival along that path. Furthest nodes being served first. This is slightly different from Algorithm 1. A in the fact that the BS is not to transmit as fast as possible but according to the rule: If previous destination node is at distance greater or equal 3, stay idle 2 TS before sending next packet. If previous destination node is at distance 2 from the BS, stay idle 1 TS before sending new packet. If previous packet is at distance 1, send next packet. The time performance of that strategy is clearly $\max_i (i - 1 + \nu_i + 2\nu_{i+1} + 3\sum_{j \geq i+2} \nu_j)$. However a proof that this strategy may be implemented is required at this point.

Proof. All that is needed is a proof that given any network G equipped with omnidirectional antenna nodes, transmissions originating at any node N_1, at distance i from the BS and at any node N_2, at distance $i+3$ from the BS may occur concurrently. Note that if node N_1 can reach node N_1' and $d(N_1) = i$ then $d(N_1') \leq d(N_1) + 1$ and $d(N_1) \leq d(N_1') + 1$, therefore $d(N_1') \in \{i-1, i, i+1\}$. Assume N_1 attempts to communicate with some node N'_1 while N_2 attempts to communicate with node N'_2. One of the attempted communications fails if either there is an edge connecting N'_1 and N_2 or there is an edge connecting N_1 and N'_2. If $(N'_1, N_2) \in E_G$ then $d(N_2) = d(N'_1) + 1 \in \{i, i+1, i+2\} < i+3$ which contradicts our hypothesis. If $(N_1, N'_2) \in E_G$ then $d(N_2) = d(N'_2) + 1 \in \{i, i+1, i+2\} < i+3$ which contradicts our hypothesis. □

Corollary 2.3.8. *If ν_j denotes the total number of data packets at distance j from the BS,*

$$\max_i \,(i - 1 + \sum_{j \geq i} \nu_j) \;\leq\; T_o(G) \;\leq\; \max_i \,(i - 1 + \nu_i + 2\nu_{i+1} + 3 \sum_{j \geq i+2} \nu_j) \quad (2.31)$$

The lower bound on $T_o(G)$ is achievable. Indeed in the previously considered example $\max_i \,(i - 1 + \sum_{j \geq i} \nu_j) = n + k - 1$. The figure below shows an example where the upper bound is achieved.

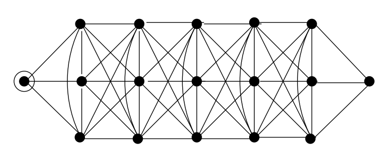


Figure 2.20: Network with cycles. $T_o(G) = T_o(\mathcal{T}_{SP}) = 18$ TS.

In general the upper bound is achieved when any node at distance i from the BS is connected to all the nodes at distance $j \in \{i-1, i, i+1\}$.

2.4 Omnidirectional/Directional Antenna Systems Comparison

The following result compares the performance of omnidirectional and directional antenna systems over a single line network.

Theorem 2.4.1. *For any line network $\boldsymbol{\nu}$ the ratio of minimum data collection times* *over a line network, assuming interference range is m times the transmission range satisfies*

$$1 \leq \frac{T_o^m(\boldsymbol{\nu})}{T_u^m(\boldsymbol{\nu})} < \begin{cases} 1.5 & m = 1 \\ 1 + \frac{2}{m} & m \geq 2 \end{cases}$$

Proof. In the case $m \geq 2$, assume there exists j_0, $1 \leq j_0 \leq n$ such that for all i, $i \neq j_0 \quad T_{j_0} \geq T_i$ and $T_{j_0+1} < T_{j_0}$. From Theorems 2.2.3 and 2.3.3,

- case: $j_0 = m + 2 + k \Rightarrow T_{j_0}^o = S_{k+1}$, $T_{j_0}^u = S_{k+3}$
  $\Rightarrow \frac{T_o^m(\boldsymbol{\nu})}{T_u^m(\boldsymbol{\nu})} = \frac{k+\sum_{j=k+1}^{k+1+m}(j-k)\nu_j+(m+2)\sum_{j\geq k+m+2}\nu_j}{k+2+\sum_{j=k+3}^{k-1+m}(j-k-2)\nu_j+m\sum_{j\geq k+m+2}\nu_j}$
  $j_0 = m{+}2{+}k \Rightarrow \nu_{k+1} = \ldots = \nu_{k+1+m} = 0 \Rightarrow \frac{T_o^m(\boldsymbol{\nu})}{T_u^m(\boldsymbol{\nu})} = \frac{k+(m+2)\sum_{j\geq k+m+2}\nu_j}{k+2+m\sum_{j\geq k+m+2}\nu_j} < \frac{m+2}{m}$

- case: $j_0 = m + 1 \Rightarrow T_{j_0}^o = S_1$, $T_{j_0}^u = S_2 \Rightarrow \frac{T_o^m(\boldsymbol{\nu})}{T_u^m(\boldsymbol{\nu})} = \frac{\sum_{j=1}^{m+1} j\nu_j+(m+2)\sum_{j\geq m+2}\nu_j}{1+\sum_{j=2}^{m}(j-1)\nu_j+m\sum_{j\geq m+1}\nu_j}$
  $j_0 = m + 1 \Rightarrow \nu_1 = \ldots = \nu_m = 0 \Rightarrow \frac{T_o^m(\boldsymbol{\nu})}{T_u^m(\boldsymbol{\nu})} = \frac{(m+1)\nu_{m+1}+(m+2)\sum_{j\geq m+2}\nu_j}{1+m\nu_{m+1}+m\sum_{j\geq m+2}\nu_j} < \frac{m+2}{m}$

- case: $1 \leq j_0 < m \Rightarrow T_{j_0}^o = S_1$, $T_{j_0}^u = S_1$
  $\Rightarrow \frac{T_o^m(\boldsymbol{\nu})}{T_u^m(\boldsymbol{\nu})} = \frac{\sum_{j=1}^{m+1} j\nu_j+(m+2)\sum_{j\geq m+2}\nu_j}{\sum_{j=1}^{m-1} j\nu_j+m\sum_{j\geq m}\nu_j} < \frac{m+2}{m}$

The case $m = 1$ follows from a similar argument. □

Note: Bounds in Theorem 2.4.1 are tight. This is clear in the case of the lower bound. As for the upper bound, consider $\boldsymbol{\nu} = \mathbf{1}_n$ (case $m = 2$), then we have $\frac{T_o^m(\boldsymbol{\nu})}{T_u^m(\boldsymbol{\nu})} = \frac{(\sum_1^{m+1} j+(m+2)\sum_{m+2}^n 1}{\sum_1^{m-1} j+m\sum_m^n 1} = \frac{(m+2)(m+1)/2+(m+2)(n-m-1)}{m(m-1)/2+m(n-m+1)} \underset{n}{\longrightarrow} \frac{m+2}{m}$

2.5 Conclusion

This work is concerned with analyzing the delay in collecting at the BS, data from sensory networks. The minimum data collection time on tree networks was derived and corresponding optimal scheduling strategies were described. We first focused our analysis on systems equipped with directional antennas and showed that more realistic hypotheses could be incorporated in our model (at the expense of the simplicity of the analysis). The study of omnidirectional antenna systems then follows under the same lines and performances of the two systems were compared on a simple line scenario.

Finally, graphs with cycles were considered and the performance of our algorithms on such graphs was compared to the optimal achievable performance. This lead to bounds on the minimum time performance of optimal data collection strategies for general graphs.

Chapter 3 Random Sensory Networks

In the previous chapter we studied the data collection problem in sensory networks, assuming the amount of data accumulated at each sensor node (characterized by a number of unit data packets) after some given observation period was finite and determined. In typical scenarios, however, the exact amount of data accumulated at each sensor node is unknown. In this chapter, we model the number of data packets as a random variable, referring to the corresponding network model as random sensory network, and analyze the delay (which is now a random variable) in collecting sensor data at the base station.

This chapter is organized as follows: We present results relative to line networks in section 3.1. In section 3.2, we present results regarding multi-line networks. In section 3.3, we compare the performance of directional and omnidirectional antenna systems. In section 3.4, we give a scaling condition on the rate at which data can be gathered by sensor nodes, for sustainable data collection. We conclude this chapter in section 3.5.

3.1 Random Line Networks

In this section, we characterize the delay in collecting random amount of data 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 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 sections 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 section, we consider the effect of error in the channel on the delay.

3.1.1 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 3.1.1. *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*

$$F_n(t) = \sum_{i=0}^{m-1} \Pr(\nu_n = i) F_{n-1}(t - 2i) \mathbf{1}_{t \geq n+2(i-1)} + \Pr(\nu_n = 0) F_{n-1}(t) \quad \textit{for } n \geq 2 \tag{3.1}$$

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

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

$$F_n(t) = \sum_{i=0}^{m-1} \Pr\{T_{min}(\boldsymbol{\nu_n}) \leq t | \nu_n = i\} \Pr(\nu_n = i) \tag{3.2}$$

To compute the conditional probability in (3.2), we use (2.3) 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

$$T_{min}(\boldsymbol{\nu_n}) \geq n - 2 + \nu_{n-1} + 2\nu_n \geq n + 2(i-1) \tag{3.3}$$

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

$$\Pr\{T_{min}(\boldsymbol{\nu_n}) \leq t | \nu_n = i\} = \Pr\{T_{min}(\boldsymbol{\nu_{n-1}}) \leq t - 2i\} \mathbf{1}_{t \geq n+2(i-1)} \tag{3.4}$$

Replacing (3.4) in (3.2), 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 (3.1). □

We can use the result of Theorem 3.1.1 to compute the CDF of $T_{min}(\boldsymbol{\nu_n})$. This is illustrated in Fig. 3.1 and Fig. 3.2. Fig. 3.1 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. 3.2 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 3.1.1 holds for any distribution of the data packets. In particular the ν_i's need not be i.i.d., however, in this chapter we deal with the case where ν_i's are independent and identically distributed.

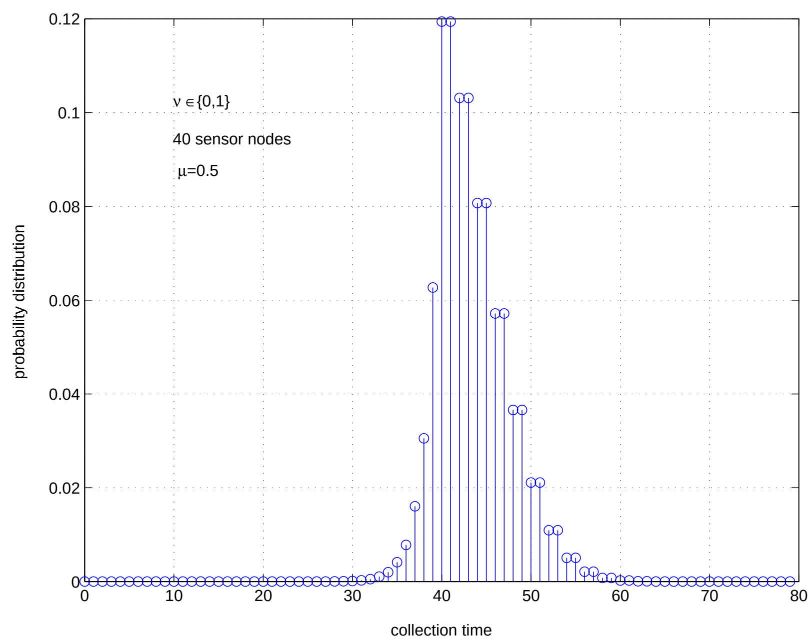


Figure 3.1: Distribution of data collection time in 40-sensor node line network. Each sensor node carries 0 or 1 data packet with probability 1/2.

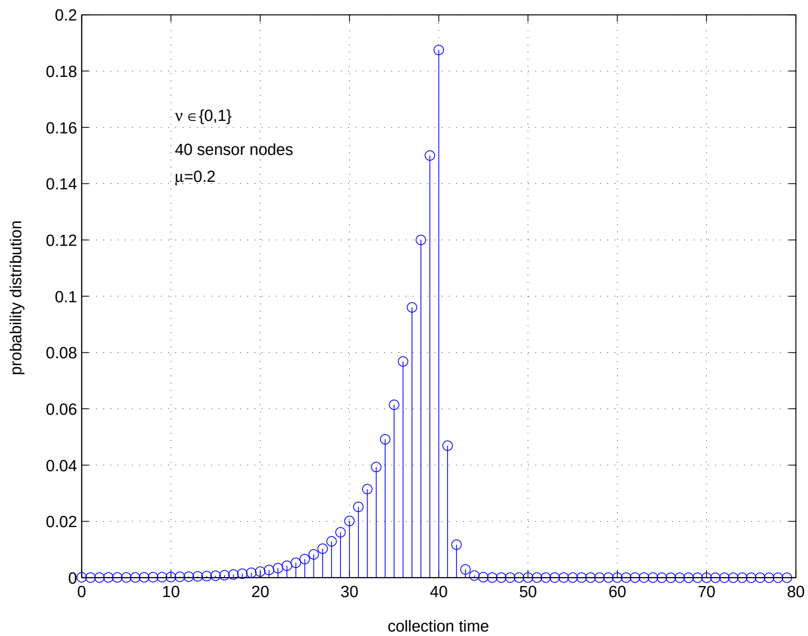


Figure 3.2: Distribution of data collection time in 40-sensor node line network. Each node in the considered network carries 0 or 1 data packet with probability 0.8 and 0.2 respectively.

Interestingly, if we plot the expected value of T_{min} as in Fig. 3.4, 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 this observation rigorously.

3.1.2 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 3.1.2. *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 are all constants independent of n. We have*

$$\lim_{n\to\infty} \frac{\mathbb{E}\{T_u\}}{n} = \begin{cases} 2\mu & \text{if } \mu \geq 1/2 \\ 1 & \text{if } \mu \leq 1/2 \end{cases} \tag{3.5}$$

Proof. We consider the case $\mu \geq 1/2$ first: Let's define $\nu_i' = \nu_i - \mu$. Using (2.3), we get

$$\begin{aligned}
\mathbb{E}\{T(\nu_{\mathbf{n}})\} &= 2\mu n + \mathbb{E}\left\{\max_{1\leq i\leq n-1}\left(i(1-2\mu) + \nu_i' + 2\sum_{i+1}^{n}\nu_j'\right)\right\} \\
&\leq 2\mu n + 2\mu - 1 + 2\mathbb{E}\left\{\max_{1\leq i\leq n}\sum_{j\geq i}^{n}\nu_j'\right\} \\
&= 2\mu n + 2\mu - 1 + 2\mathbb{E}\left\{\max_{1\leq i\leq n}\sum_{j=1}^{n+1-i}\nu_{n-j+1}'\right\} \qquad (3.6)
\end{aligned}$$

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 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 by Erdös and Kac [19] on the convergence of distribution of the maximum of partial sums.

Lemma 3.1.3. *For any* λ *and* $a > 1$,

$$\Pr\left\{\max_{1\leq i\leq n} \geq \lambda\sigma\sqrt{n}\right\} \leq \frac{a-1}{a}\Pr\left\{\sum_{j=1}^{n}\nu_j' \geq (\lambda - \sqrt{a})\sigma\sqrt{n}\right\} \qquad (3.7)$$

where $\nu_i' = \nu_i - \mu$ *and* ν_i *is as defined in Theorem 3.1.2.*

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

$$E_i = \left\{\max_{0\leq j<i} S_j \leq \lambda\sigma\sqrt{n} \leq S_i\right\} \quad i = 1, \ldots, n. \qquad (3.8)$$

which is inspired by [19]. We can then state the following inequality by the union bound,

$$\begin{aligned}
\Pr\left\{\max_{1\leq i\leq n} S_i \geq \lambda\sigma\sqrt{n}\right\} &\leq \Pr\left\{S_n > (\lambda - \sqrt{a})\sigma\sqrt{n}\right\} + \\
&\quad \sum_{i=1}^{n}\Pr\left\{E_i \cap \left(S_n \leq (\lambda - \sqrt{a})\sigma\sqrt{n}\right)\right\} \qquad (3.9)
\end{aligned}$$

To evaluate the second term in the right-hand side of (3.9), we note that $S_i \geq \lambda\sigma\sqrt{n}$

and $S_n \leq (\lambda - \sqrt{a})\sigma\sqrt{n}$ imply $S_i - S_n \geq \sqrt{a}\sigma\sqrt{n}$. Then using the fact that $S_i - S_n$ is independent of S_j for $j \leq i$, we may write

$$\begin{aligned}
\sum_{i=1}^{n} \Pr\left\{E_i \cap \left(S_n \leq (\lambda - \sqrt{a})\sigma\sqrt{n}\right)\right\} &\leq \sum_{i=1}^{n} \Pr(E_i)\Pr\left(S_i - S_n \geq \sqrt{a}\sigma\sqrt{n}\right) \\
&\leq \sum_{i=1}^{n} \Pr(E_i)\frac{E\left\{(S_i - S_n)^2\right\}}{a\sigma^2 n} \\
&= \sum_{i=1}^{n} \Pr(E_i)\frac{(n-i)\sigma^2}{a\sigma^2 n} \\
&\leq \frac{1}{a}\sum_{i=1}^{n} \Pr(E_i) \\
&\leq \frac{1}{a}\Pr\left(\max_{1\leq i\leq n} S_i \geq \lambda\sigma\sqrt{n}\right) \qquad (3.10)
\end{aligned}$$

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

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

since the events E_i are disjoint events. Therefore, Lemma 3.1.3 follows from (3.10) and (3.9). □

Now we can use Chebychev's inequality to evaluate the right-hand side of Lemma 3.1.3 as follows

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

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

$$\Pr\left(\max_{1\leq i\leq n} \sum_{j=i}^{n} \nu_j' \geq \sigma \log n\sqrt{n}\right) = O\left(\frac{1}{\log^2 n}\right) \qquad (3.11)$$

Eq. (3.11) implies that, with high probability $\max_{1\leq i\leq n} \sum_{j\geq i} \nu_j'$ is less than $\sigma \log n\sqrt{n}$.

Therefore, we may write

$$\mathbb{E}\left\{\max_{1\le i\le n-1}\sum_{j=i}^{n}\nu'_j\right\} \le \sigma\log n\sqrt{n}\,\Pr\left\{\max_{1\le i\le n-1}\sum_{j=i}^{n}\nu'_j < \sigma\log n\sqrt{n}\right\} + \\ (m-1-\mu)n\Pr\left\{\max_{1\le i\le n-1}\sum_{j\ge i}^{n}\nu'_j > \sigma\log n\sqrt{n}\right\} \\ = \sigma\log n\sqrt{n} + O\left(\frac{n}{\log^2 n}\right) \quad (3.12)$$

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. (2.3), 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

$$\mathbb{E}(T_{min}(\boldsymbol{\nu_n})) \ge 2\mu n - \mu \quad (3.13)$$

Considering (3.13) and the upper bound derived in (3.12), 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 (3.5) for $\mu \ge 1/2$.

Next, we consider the case $\mu \le 1/2$: Let's define $\nu'_i = \nu_i - 1/2$. Using (2.3), we get

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

Taking the expectation of both sides and using inequality (3.12) we get

$$\mathbb{E}(T_{min}(\boldsymbol{\nu_n})) \le n + 2\sigma\log n\sqrt{n} + O\left(\frac{n}{\log^2 n}\right) \quad (3.14)$$

On the other hand, it is clear that if there is any packet at 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

$$\mathbb{E}(T_{min}(\boldsymbol{\nu_n})) \geq (n - \log n)(1 - (\Pr(\nu_i = 0))^{\log n}) = n - O(\log n) \tag{3.15}$$

which leads to (3.5) for $\mu \leq 1/2$. □

Remark: Theorem 3.1.2 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. (3.12), the theorem goes through as long as $m = o(n)$.

Fig. 3.3 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. 3.4 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.

3.1.3 Collected Data Distribution

In this section we attempt to measure the rate at which data is being retrieved by the BS. For general distributions on the number of data packets this is a difficult problem. However, in the particular case where $\nu_i \in S_1$ we are able to do so. Let V_i denote the number of data packets collected by the BS up to time i, then we have the following theorem, if $\nu_i \in \{0, 1\}$ and $\Pr\{\nu_i = 0\} = 1/2$, $0 \leq V_i \leq \lceil i/2 \rceil$ and Algorithm 1 is used.

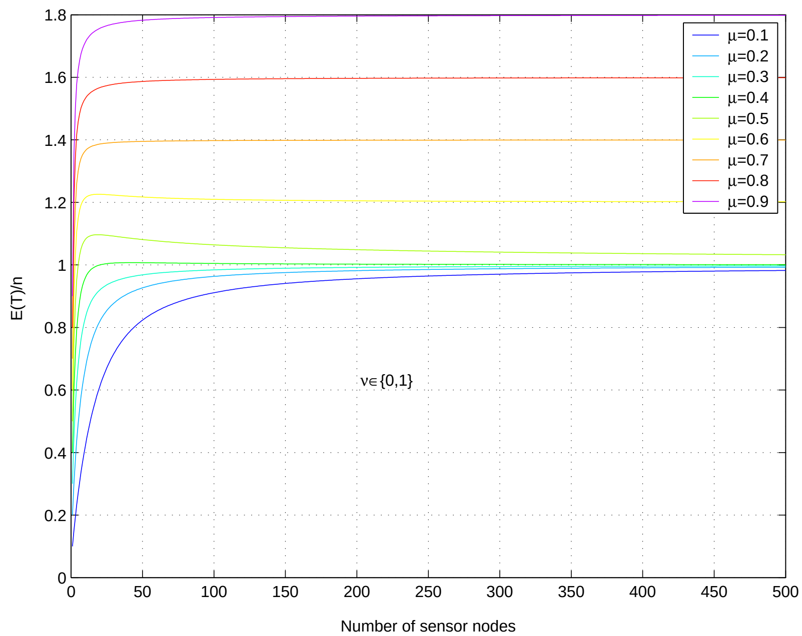


Figure 3.3: Average collection time as a function of average number of packets per node and number of nodes in line network. Nodes carry 0 or 1 data packet with probability $1-\mu$ and μ respectively.

Theorem 3.1.4. *Consider a line network consisting of N_0 sensor nodes carrying 0 or 1 data packet with probability 1/2. Let $\Pr\{V_i = j\}$ denote the probability that the BS has collected j data packets by time i, then we have $\Pr(V_0 = 0) = 1$, $\forall i \in \mathbb{N}$, $i > 2N_0$, $\forall j \in \mathbb{N}$, $\Pr(V_i = j) = \binom{N_0}{j}/2^{N_0}$ and $\forall i \in \mathbb{N}$, $1 \le i \le 2N_0$ and $\forall j \in \mathbb{N}$, $0 \le j \le \lceil i/2 \rceil$*

$$\Pr(V_i = j) = \frac{1}{2}(\Pr(V_{i-1} = j-1) + \Pr(V_{i-1} = j)) \text{ if } i \text{ is } \begin{cases} \text{even} \\ \text{odd and } j < \lceil \frac{i}{2} \rceil \end{cases} \tag{3.16}$$

$$\Pr(V_i = \lceil \frac{i}{2} \rceil) = \Pr(V_{i-1} = \lceil \frac{i-1}{2} \rceil) + 0.5 \Pr(V_{i-1} = \lceil \frac{i-2}{2} \rceil) \text{ if } i \text{ is odd and } j = \lceil \frac{i}{2} \rceil$$

Proof. By induction. □

For illustration purposes we include the example of a 7-sensor node line network

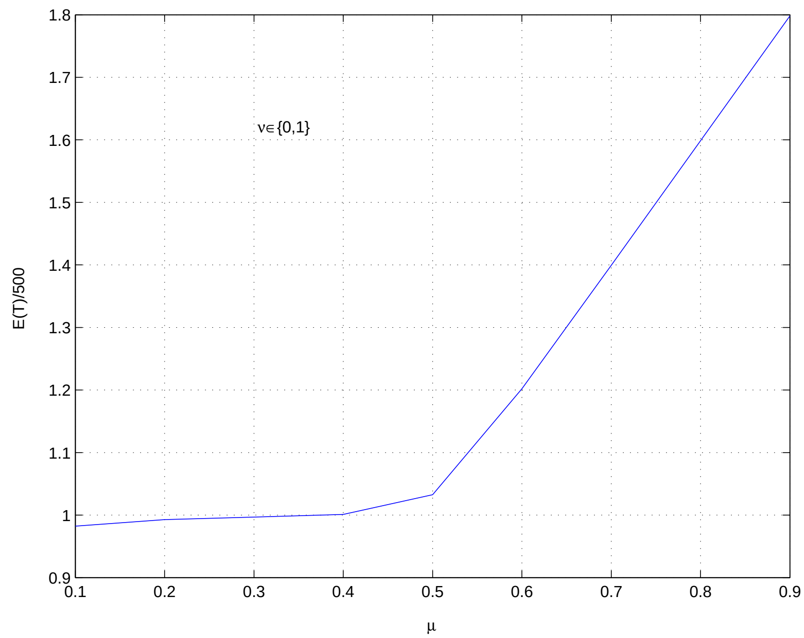


Figure 3.4: Average collection time as a function of average number of packets per node in 500-node line network. Nodes carry 0 or 1 data packet with probability $1-\mu$ and μ respectively.

in Table 3.1.

3.1.4 Multihop Case

In order to get a better insight into the result of Theorem 2.2.4, we obtain the asymptotic behavior of the expected minimum delay as n approaches infinity in the next theorem. Theorem 3.1.5, in fact, quantifies the dependency between the minimum collection time and the transmission range.

Theorem 3.1.5. *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*

$\Pr(V_i = j)$	j=0	1	2	3	4	5	6	7
i=0	1	0	0	0	0	0	0	0
1	0.5	0.5	0	0	0	0	0	0
2	0.25	0.75	0	0	0	0	0	0
3	0.125	0.5	0.375	0	0	0	0	0
4	0.0625	0.3125	0.6250	0	0	0	0	0
5	0.0156	0.1094	0.3281	0.5469	0	0	0	0
6	0.0078	0.0625	0.2188	0.4375	0.2734	0	0	0
7	0.0039	0.0352	0.1406	0.3281	0.4922	0	0	0
8	0.0020	0.0195	0.0879	0.2344	0.4102	0.2461	0	0
9	0.0010	0.0107	0.0537	0.1611	0.3223	0.4512	0	0
10	0.0005	0.0059	0.0322	0.1074	0.2417	0.3867	0.2256	0
11	0.0002	0.0032	0.0190	0.0698	0.1746	0.3142	0.4189	0
12	0.0001	0.0017	0.0111	0.0444	0.1222	0.2444	0.3666	0.2095
13	0.0001	0.0009	0.0064	0.0278	0.0833	0.1833	0.3055	0.3928

Table 3.1: Probability to have collected j packets by TS i in 7-node line network.

independent of n.

$$\lim_{n\to\infty} \frac{\mathbb{E}\{T_{min}(h, \boldsymbol{\nu_n})\}}{n} = \begin{cases} (1+\frac{1}{h})\mu & \textit{if } \mu \geq \frac{1}{h+1} \\ 1 & \textit{if } \mu \leq \frac{1}{h+1} \end{cases} \tag{3.17}$$

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 3.1.5 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 $O(\sum_i i^2\nu_i) = O(n^3)$ (worst-case) 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 $O(\sum_i i\nu_i) = O(n^2)$.

3.1.5 Packet Splitting to Improve the Average Delay

As Eq. (3.5) 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 3.1.6. *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:*

$$G_{max}(\boldsymbol{\nu_n}) = \lim_{k} G(\boldsymbol{\nu_n}, k) = \frac{\max\limits_{1\leq i\leq n-1}(i - 1 + \nu_i + 2\sum_{j\geq i+1}^{n} \nu_j)}{\nu_1 + 2\sum_{j>1}^{n} \nu_j} \tag{3.18}$$

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

$$G(\boldsymbol{\nu_n}, k) = \frac{k \max\limits_{1\leq i\leq n}(i - 1 + \nu_i + 2\sum_{j>i}^{n} \nu_j)}{\max\limits_{1\leq i\leq n}(i - 1 + k\nu_i + 2k\sum_{j\geq i+1}^{n} \nu_j)} = \frac{\max\limits_{1\leq i\leq n}(k(i-1) + k\nu_i + 2k\sum_{j>i}^{n} \nu_j)}{\max\limits_{1\leq i\leq n}(i - 1 + k\nu_i + 2k\sum_{j\geq i+1}^{n} \nu_j)} \tag{3.19}$$

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

$$\max_{1\leq i\leq n}(k_1(i-1) + k_1k_2\nu_i + 2k_1k_2\sum_{j>i}^{n} \nu_j) \geq \max_{1\leq i\leq n}(k_2(i-1) + k_1k_2\nu_i + 2k_1k_2\sum_{j>i}^{n} \nu_j)$$

which implies that $G(\boldsymbol{\nu_n}, k_1) \geq G(\boldsymbol{\nu_n}, k_2)$. The limit in (3.18) can be also easily shown using (3.19). □

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 3.1.7. *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:*

$$\lim_{n\to\infty} \frac{\mathbb{E}\{T_{min}\}}{n} = \begin{cases} 2\mu & \text{if } \mu \geq 1/2k \\ 1/k & \text{if } \mu \leq 1/2k \end{cases} \tag{3.20}$$

Proof. The proof falls along the same line as the proof of Theorem 3.1.2 substituting ν_i with $k\nu_i$, for all i, $1 \leq i \leq n$ and noting that the smaller size packets are transmitted k times faster. □

The limit in Eq. (3.20) should be compared to the data collection in the case where packets are not split as shown in Eq. (3.5). We conclude that in the asymptotic case, data splitting results in gain in the collection time for networks with low data load, i.e., $\mu \leq \frac{1}{2}$. It is also worth noting that Eqs. (3.20) and (3.5) imply that if $k \geq \frac{1}{2\mu}$ 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.

3.1.6 A Simple Distributed Suboptimal Strategy

It is important to note that the minimum collection time in (2.3) 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. 3.5.

The following theorem compares the performance of this strategy to the minimal collection time derived in (2.3).

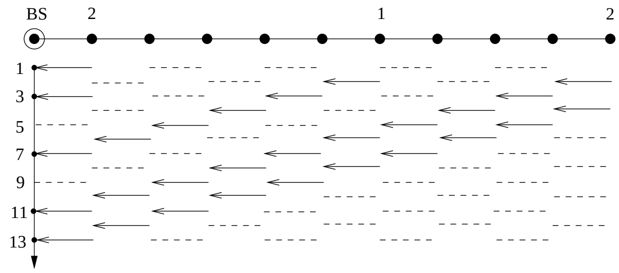


Figure 3.5: Suboptimal distributed data collection strategy.

Theorem 3.1.8. *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 this distributed scheduling strategy, denoted by $T(\boldsymbol{\nu_n})$, is:*

$$T(\boldsymbol{\nu_n}) = \max_{1\leq i\leq n}(i - 2 + 2\sum_{j\geq i-1}^{n} \nu_j) \tag{3.21}$$

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:

$$T(\boldsymbol{\nu_n}) = \max_{1\leq i\leq n}(i - 1 + 2\sum_{j\geq i}^{n} \nu_j) \tag{3.22}$$

Proof. In the rest of this chapter 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 (3.21). Eq. (3.22) 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 (2.3) and (3.22):

$$T_{min}(\boldsymbol{\nu_n}) \leq T(\boldsymbol{\nu_n}) \leq 2T_{min}(\boldsymbol{\nu_n}) - 1 \tag{3.23}$$

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 (3.23) 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 3.1.9. *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.*

$$\lim_{n\to\infty} \frac{\mathbb{E}\{T(\boldsymbol{\nu_n})\}}{n} = \begin{cases} 2\mu & \text{if } \mu \geq 1/2 \\ 1 & \text{if } \mu \leq 1/2 \end{cases} \tag{3.24}$$

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

Proof. This proof is similar to the proof of Theorem 3.1.2. □

3.1.7 Noisy 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 3.1.6. Fig. 3.6 illustrates the process. This is the same network as shown in Fig. 3.5 but it is now affected by three erasures (each shown by a crossed arrow). The new transmission time is 15 TS, an increase of 2 TS.

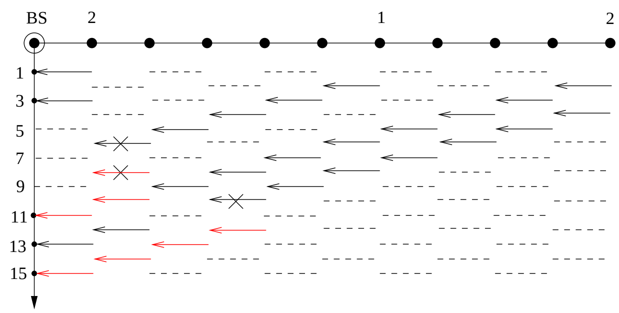


Figure 3.6: Data collection in line network under the assumption of an erasure channel. An erased packet is marked with a cross.

Theorem 3.1.10. *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*

$$T(p, \boldsymbol{\nu_n}) = (1-p)^{\sum_{i=1}^{n} i\nu_i} \sum_{k \geq 0} p^k \sum_{\sum_i e_i \chi(\nu_i > 0) = k} \prod_{i \geq 1}^{n} \binom{i\nu_i + e_i - 1}{i\nu_i - 1} T(\boldsymbol{\nu_n} + \boldsymbol{e_i}) \quad (3.25)$$

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. 3.7. 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 [57]. Theorem 3.1.10 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 3.1.11. *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:*

$$1 \leq \frac{\mathbb{E}(T(p, \boldsymbol{\nu_n}))}{\mathbb{E}(T(0, \boldsymbol{\nu_n}))} \leq 1 + O(np) \tag{3.26}$$

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 by α_i for $i = 1, \ldots, \sum i\nu_i$ the number of extra time slots needed to transmit the packet at the i'th transmission, we may write

$$T(p, \boldsymbol{\nu_n}) \leq \sum_{j=1}^{\sum_{i=1}^{n} ip_i} \alpha_j + T(0, \boldsymbol{\nu_n}) \tag{3.27}$$

where α_i has geometric distribution, i.e.,

$$\Pr(\alpha_i) = p^{i-1}(1-p) \Rightarrow \mathbb{E}(\alpha_i) = \frac{p}{1-p} \tag{3.28}$$

Taking expectation of both sides of (3.27), we obtain,

$$\frac{\mathbb{E}(T(p))}{\mathbb{E}(T(0))} \leq \frac{p \sum_{i=1}^{n} ip_i}{(1-p)\mathbb{E}(T(0))} + 1 \tag{3.29}$$

which completes the proof of our theorem. □

In particular Theorem 3.1.11 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.

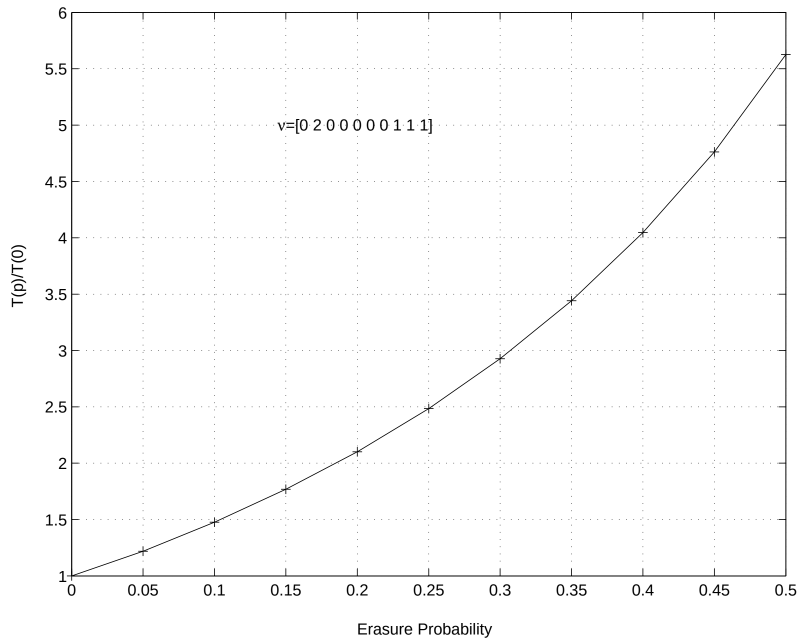


Figure 3.7: Ratio $\frac{T(p)}{T(0)}$ as a function of p in line network.

3.2 Random Multi-line 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. 3.8. 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.

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 multi-line network with $L \geq 2$ lines of length n_0, we have

$$\mathbb{E}(T_{min}^{L,n_0}) \geq n_0 L \mathbb{E}(\nu_i) \tag{3.30}$$

which follows by taking the expectation of both sides of the inequality $T_{min}^{L,n_0} \geq$ (num-

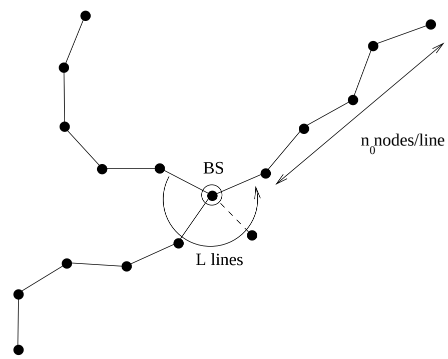


Figure 3.8: Multi-line Network.

ber of packets in network). In what follows, we shall prove that as L increases, the expected collection time converges toward this lower bound.

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, collect all the packets to the nodes within distance one of the

BS which at most takes $(m-1)(2n_0-3)$ TS, and second, send the packets of the nodes at distance one from the BS to BS.

Theorem 3.2.1. *Consider a multi-line 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*

$$n_0 L\mu \leq \mathbb{E}(T_{min}^{L,n_0}) \leq n_0 L\mu + O\left(\frac{1}{L}\right) + (m-1)(2n_0-3)(1-p_{m-1})^{L/2}. \qquad (3.31)$$

In particular,

$$i) \quad \text{if } L > (2+O(1)) \log_\alpha n_o, \ \lim_{n_o \to \infty} \mathbb{E}(T_{min}^{L,n_0}) - n_o L\mu = 0 \qquad (3.32)$$

$$ii) \quad \text{if } 2\log_\alpha n_o > L \text{ and } \lim_{n_o \to \infty} L = +\infty, \ \lim_{n_o \to \infty} \frac{\mathbb{E}(T_{min}^{L,n_0})}{n_o L\mu} = 1 \qquad (3.33)$$

$$iii) \quad \text{if } L = cte, n_o L\mu \leq \mathbb{E}(T_{min}^{L,n_0}) \leq n_o L\mu(1+\epsilon) \qquad (3.34)$$

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. (3.30) 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 the following upper bound for the delay.

$$T_{min}^{L,n_0} \leq \max\{\eta, \ (m-1)(2n_0-3)\} + \sum_{i=1}^{(m-1)(2n_0-3)} e_i \qquad (3.35)$$

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

$$\begin{aligned}\Pr(e_{2k} = 0) &\geq \Pr(\text{having at least m} - 1 \text{ packets at distance } k) \\ &\geq \Pr(\text{at least one node at distance k has m} - 1 \text{ packets}) \\ &= 1 - (1 - p_{m-1})^{L/2} \end{aligned} \tag{3.36}$$

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

$$\Pr\left((m-1)(2n_0-3) \leq \eta\right) \geq 1 - O\left(\frac{1}{n_0 L}\right) \tag{3.37}$$

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 (3.35) to get

$$\begin{aligned}\mathbb{E}(T_{min}^{L,n_0}) &\leq \mathbb{E}(\eta) + (m-1)(2n_0-3)\Pr(\eta \leq (m-1)(2n_0-3)) + \sum_{i=1}^{(m-1)(2n_0-3)} \Pr(e_i = 1) \\ &\leq n_0 L\mu + O\left(\frac{1}{L}\right) + (m-1)(2n_0-3)(1-p_{m-1})^{L/2} \end{aligned} \tag{3.38}$$

that completes the first part proof. □

Theorem 3.2.1 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 $c(n) \to \infty$ [33]. 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_0^l for $l = 1, \ldots, L$ (instead of n_0 for all l's) the lower bounds on the expected delay becomes $\mathbb{E}(T_{min}^{L,n_0}) \geq \mu \sum_{l=1}^{L} n_0^l$. We can further find an upper bound by replacing n_0 by max n_0^l in (3.35) and noting that $\mathbb{E}(\eta)$ is equal to the lower bound. The result follows in a similar fashion. Therefore as long as $(\max n_0^l)m = o\left(\frac{1}{1-p_{m-1}}\right)^L$ and L grows to infinity, the expected delay converges to $E\{\eta\}$. In Fig. 3.9 the difference between average collection time and average packet number in the network for multi-line 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 carries either 0 or 1 packet with probability $1 - \mu$ and μ respectively. The exact collection time for a particular instance is known and given by Eq. (3.11) and this is averaged over multiple instances (20000) to yield Fig. 3.9.

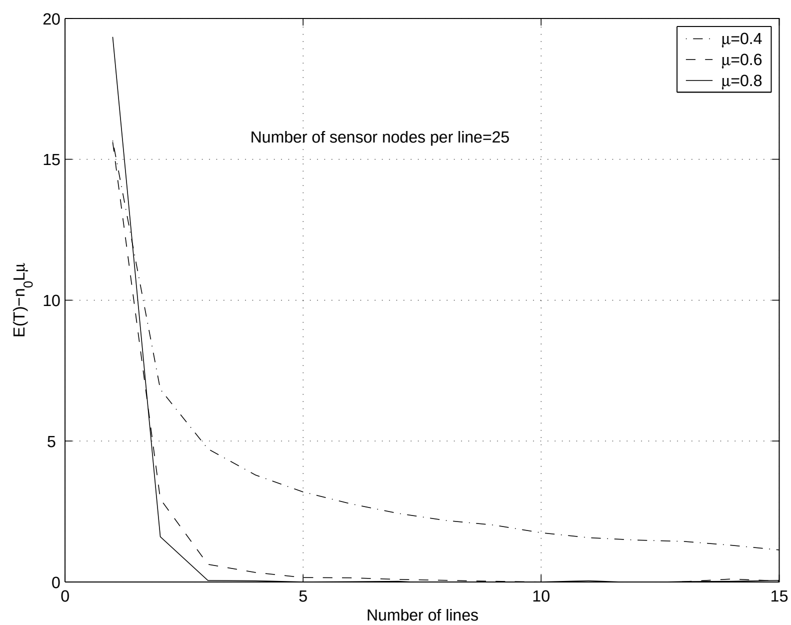


Figure 3.9: Difference between expected delay and average number of packets in network as a function of average number of packets per node and number of lines in multi-line network (25 nodes per line). Nodes carry 0 or 1 data packet with probability $1 - \mu$ and μ respectively.

3.2.1 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 results on multi-line networks. In this section we discuss the implications of previous results for networks of more general topologies.

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 3.1.2.

If the degree of the BS is 1, It is shown in the previous chapter 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. Furthermore, 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 multi-line networks to tree topologies (indeed given what what said before, a tree may be thought of as a multi-line network).

Finally the previous results give some intuition about the asymptotic 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 [33] 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 before this network behaves like a multi-line 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, $\pi r^2(n)n = O(\log(n))$ and either condition i) or condition ii) of Theorem 3.2.1 applies.

3.3 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 neighbors. This in turns increases the length of the optimum data collection schedule (when compared to directional systems). In fact we know from Theorem 2.3.1 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:

$$T_o(\boldsymbol{\nu_n}) = \max_{1 \leq i \leq n-2} (i - 1 + \nu_i + 2\nu_{i+1} + 3 \sum_{j \geq i+2}^{n} \nu_j) \tag{3.39}$$

where $\boldsymbol{\nu_n} = (\nu_1, \ldots, \nu_n)$. We know from Theorem 2.4.1 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. 2.4 the minimum data collection time becomes 14 TS, a 40 % increase.

In the following sections, we present results for the delay analysis in networks equipped with omnidirectional antennas. Results are analogous to the results stated in section 3.2 and we omit proofs for the sake of brevity.

3.3.1 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 3.3.1. *Let $F_n(t)$ be the CDF of the minimum delay $T_o(\boldsymbol{\nu_n})$, i.e. $F_n(t) = \Pr\{T_o(\boldsymbol{\nu_n}) \leq t\}$. Then $F_n(t)$ satisfies the following recursion*

$$F_n(t) = \sum_{i=1}^{m-1} \Pr(\nu_n = i) F_{n-1}(t-3i)\mathbf{1}_{t \geq n+3(i-1)} + \Pr(\nu_n = 0) F_{n-1}(t) \quad \forall n \geq 3 \quad (3.40)$$

where

$$\mathbf{1}_{t \geq t_0} = \begin{cases} 1 & \textit{if} \quad t \geq t_0 \\ 0 & \textit{otherwise} \end{cases}$$

and,

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

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

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

$$F_n(t) = \sum_{i=0}^{m-1} \Pr\{T(\boldsymbol{\nu_n}) \leq t | \nu_n = i\} \Pr(\nu_n = i) \quad (3.41)$$

To compute the conditional probability in (3.41), we use (3.39) and the fact that for all $k = 1, \ldots, n-1$, $T(\boldsymbol{\nu_n}) \geq k - 1 + \nu_k + 2\nu_{k+1} + 3\sum_{j=k+2}^{n} \nu_j$. Therefore replacing $k = n-2$ and assuming $\nu_n = i$, we get

$$T(\nu_{\mathbf{n}}) \geq n - 3 + \nu_{n-2} + 2\nu_{n-1} + 3\nu_n \geq n + 3(i-1) \quad (3.42)$$

Thus if $t < n + 3(i-1)$, then $\Pr\{T(\boldsymbol{\nu_n}) \leq t | \nu_n = i\} = 0$. Using the definition of the function $\mathbf{1}_{t \geq t_0}$, for any i we may then write the conditional probability as

$$\Pr\{T(\boldsymbol{\nu_n}) \leq t | \nu_n = i\} = \Pr\{T(\boldsymbol{\nu_{n-1}}) \leq t - 3i\}\mathbf{1}_{t \geq n+3(i-1)} \tag{3.43}$$

Replacing (3.43) in (3.41), we get

$$\begin{aligned} F_n(t) = {} & F_{n-1}(t) \Pr(\nu_n = 0) + \\ & \sum_{i \geq 1}^{m-1} \Pr\{T(\boldsymbol{\nu_{n-1}}) \leq t - 3i)\}\mathbf{1}_{t \geq n+3(i-1)} \Pr(\nu_n = i) \end{aligned}$$

which leads to (3.40). □

We can use the result of Theorem 3.3.1 to compute the CDF of $T_o(\boldsymbol{\nu_n})$. This is illustrated in Fig. 3.10 which shows the distribution of the delay $T_o(\boldsymbol{\nu_n})$ in 40-sensor node line networks in which each node carries either 0 or 1 packet with probability 0.7 and 0.3 respectively. It is also worth noting that the result of Theorem 3.3.1 holds for any distribution of the data packets. In particular the ν_i's need not be i.i.d., however, in this chapter we deal with the case that ν_i's are independent and identically distributed. Interestingly, if we plot the expected value of T_o as in Fig. 3.12, 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.

3.3.2 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 3.3.2. *Let ν_i's be i.i.d. random variables $\nu_i \in S_m$ with mean μ, variance*

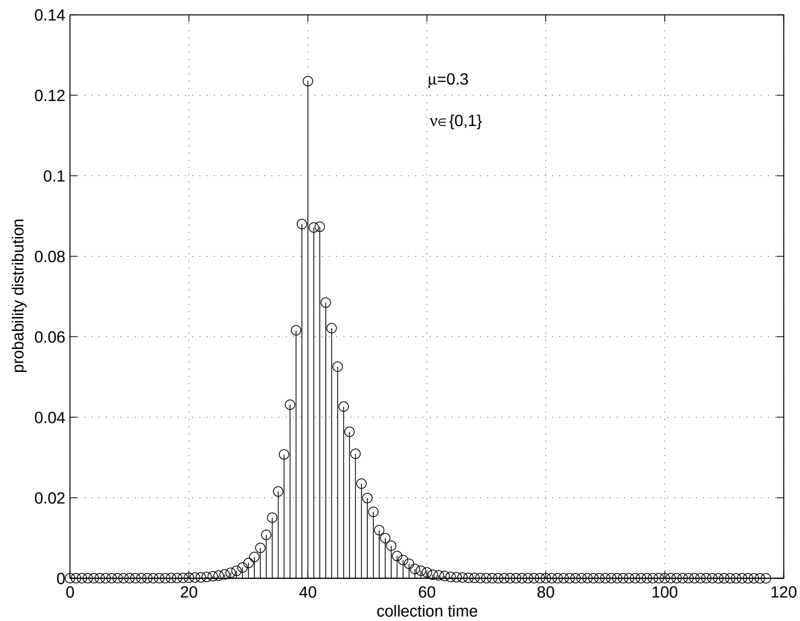


Figure 3.10: Distribution of data collection time in 40-sensor node line network. Nodes carry 0 or 1 data packet with probability 0.7 and 0.3 respectively.

σ^2 where μ, σ^2, m are all constants independent of n. We have

$$\lim_{n\to\infty} \frac{\mathbb{E}\{T_o\}}{n} = \begin{cases} 3\mu & \textit{if } \mu \geq 1/3 \\ 1 & \textit{if } \mu \leq 1/3 \end{cases} \tag{3.44}$$

Proof. We consider the case $\mu \geq 1/3$ first: Let's define $\nu_i' = \nu_i - \mu$. Using (3.39), we get

$$\begin{aligned} \mathbb{E}\{T(\nu_{\mathbf{n}})\} &= 3\mu n + \\ & \mathbb{E}\left\{\max_{1\leq i\leq n-2}\left(i(1-3\mu) + \nu_i' + 2\nu_{i+1}' + 3\sum_{i+2}^{n} \nu_j'\right)\right\} \\ & \leq 3\mu n + 3\mu - 1 + 3\mathbb{E}\left\{\max_{1\leq i\leq n} \sum_{j\geq i}^{n} \nu_j'\right\} \\ & = 3\mu n + 3\mu - 1 + 3\mathbb{E}\left\{\max_{1\leq i\leq n} \sum_{j=1}^{n+1-i} \nu_{n-j+1}'\right\} \end{aligned} \tag{3.45}$$

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 find a bound for $\mathbb{E}(\max_{1\leq i\leq n} \sum_{j\geq i}^{n} \nu_j')$, we first state the following lemma which is proved based on Erdös and Kac [19] where a convergence theorem for the distribution of the maximum of partial sums was proven. It is worth noting convergence in distribution does not imply convergence in the mean and so we cannot directly use the result of Erdös and Kac[1]. To simplify the notation, let's first define $S_i = \sum_{j=1}^{n-i+1} x_j = \sum_{j\geq i}^{n} \nu_j'$ where $x_i = \nu_{n-j+1}^{'}$.

Lemma 3.3.3. *For any λ,*

$$\Pr\left\{\max_{1\leq i\leq n} S_i \geq \lambda\sigma\sqrt{n}\right\} \leq 2\Pr\left\{S_n \geq (\lambda - \sqrt{2})\sigma\sqrt{n}\right\}. \tag{3.46}$$

Proof. We first define the events E_i as

$$E_i = \left\{\max_{0\leq j<i} S_j \leq \lambda\sigma\sqrt{n} \leq S_i\right\} \quad i = 1, \ldots, n. \tag{3.47}$$

which is inspired by [19]. We can then state the following inequality by the union bound.

$$\Pr\left\{\max_{1\leq i\leq n} S_i \geq \lambda\sigma\sqrt{n}\right\} \leq \Pr\left\{S_n > (\lambda - \sqrt{2})\sigma\sqrt{n}\right\} + \sum_{i=1}^{n} \Pr\left\{E_i \cap \left(S_n \leq (\lambda - \sqrt{2})\sigma\sqrt{n}\right)\right\} \tag{3.48}$$

To evaluate the second term in the right hand side of (3.48), we note that $S_i \geq \lambda\sigma\sqrt{n}$ and $S_n \leq (\lambda - \sqrt{2})\sigma\sqrt{n}$ imply $S_i - S_n \geq \sqrt{2}\sigma\sqrt{n}$. Then using the fact that $S_i - S_n$

[1] It is quite easy to come up with an example that the distribution of a random variable converges to $f(x)$ but its mean does not converge to $\int xf(x)dx$

is independent of S_j for $j \le i$, we may write

$$\begin{aligned}
&\sum_{i=1}^{n} \Pr\left\{E_i \cap \left(S_n \le (\lambda - \sqrt{2})\sigma\sqrt{n}\right)\right\} \\
&\qquad \le \sum_{i=1}^{n} \Pr(E_i)\Pr\left(S_i - S_n \ge \sqrt{2}\sigma\sqrt{n}\right) \\
&\qquad \le \sum_{i=1}^{n} \Pr(E_i)\frac{E\left\{(S_i - S_n)^2\right\}}{2\sigma^2 n} \\
&\qquad = \sum_{i=1}^{n} \Pr(E_i)\frac{(n-i)\sigma^2}{2\sigma^2 n} \\
&\qquad \le \frac{1}{2}\sum_{i=1}^{n} \Pr(E_i) \\
&\qquad \le \frac{1}{2}\Pr\left(\max_{1\le i\le n} S_i \ge \lambda\sigma\sqrt{n}\right) \qquad (3.49)
\end{aligned}$$

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

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

Therefore Lemma 3.3.3 follows from (3.49) and (3.48). □

Now we can use Chebychev's inequality to evaluate the right-hand side of Lemma 3.3.3 as follows.

$$\begin{aligned}
\Pr\left\{S_n = \sum_{i=1}^{n} \nu_i^{'} \ge (\lambda - \sqrt{2})\sigma\sqrt{n}\right\} &\le \frac{n\sigma^2}{(\lambda-\sqrt{2})^2\sigma^2 n} \\
&\le \frac{1}{(\lambda-\sqrt{2})^2} \qquad (3.50)
\end{aligned}$$

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

$$\Pr\left(\max_{1\le i\le n} \sum_{j=i}^{n} \nu_i^{'} \ge \sigma \log n\sqrt{n}\right) = O\left(\frac{1}{\log^2 n}\right) \qquad (3.51)$$

Eq. (3.51) implies that, with high probability, $\max_{1\leq i\leq n} \sum_{j\geq i} \nu'_j$ is less than $\sigma \log n\sqrt{n}$. Therefore, we may write

$$\begin{aligned}
&\mathbb{E}\left\{\max_{1\leq i\leq n-1}\sum_{j=i}^{n}\nu'_j\right\} \\
&\leq \sigma\log n\sqrt{n}\,\Pr\left\{\max_{1\leq i\leq n-1}\sum_{j=i}^{n}\nu'_j < \sigma\log n\sqrt{n}\right\} + \\
&(m-1-\mu)n\,\Pr\left\{\max_{1\leq i\leq n-1}\sum_{j\geq i}^{n}\nu'_j > \sigma\log n\sqrt{n}\right\} \\
&= \sigma\log n\sqrt{n} + O\left(\frac{n}{\log^2 n}\right)
\end{aligned} \tag{3.52}$$

which follows from the fact that $\nu'_i \leq m-1-\mu$. We now derive a lower bound on $E(T_o(\boldsymbol{\nu_n}))$. Using Eq. (3.39), we get $T_o(\boldsymbol{\nu_n}) \geq \nu_1 + 2\nu_2 + 3\sum_{j\geq 3}^{n}\nu_j$. Taking the expectation of both sides, we get

$$\mathbb{E}(T(\boldsymbol{\nu_n})) \geq 3\mu n - \mu \tag{3.53}$$

Considering (3.53) and the upper bound derived in (3.52), we deduce that

$$3\mu n - \mu \leq \mathbb{E}(T(\nu_{\mathbf{n}})) \leq 3\mu n + 3\mu - 1 + 3\sigma\log n\sqrt{n} + O\left(\frac{n}{\log^2 n}\right)$$

which leads to (3.44) for $\mu \geq 1/3$.

The case $\mu \leq 1/3$ follows along the same line and it can be shown that

$$n - 1 - \mu \leq \mathbb{E}(T(\boldsymbol{\nu_n})) \leq n + 3\sigma\log n\sqrt{n} + O\left(\frac{n}{\log^2 n}\right) \tag{3.54}$$

which leads to (3.44) for $\mu \leq 1/3$.

□

Remark: Theorem 3.3.2 can be easily generalized to the case that ν_i's are independent and have mean $\mu_i \geq \frac{1}{3}$ 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. The theorem goes through as long as $m = o(n)$.

Figs. 3.11 and 3.12 illustrate the behavior of the (minimum) average collection time on a line network equipped with omnidirectional antennas. It is assumed that a given sensor node has collected 0 or 1 data packet with probability μ and $1-\mu$ respectively (equivalently that the average number of packets per node is μ). Figs. 3.11 and 3.12 were obtained through the application of Theorem 3.3.1, which means that they are an exact computation of the CDF. They both confirm the asymptotic result proven in Theorem 3.3.2.

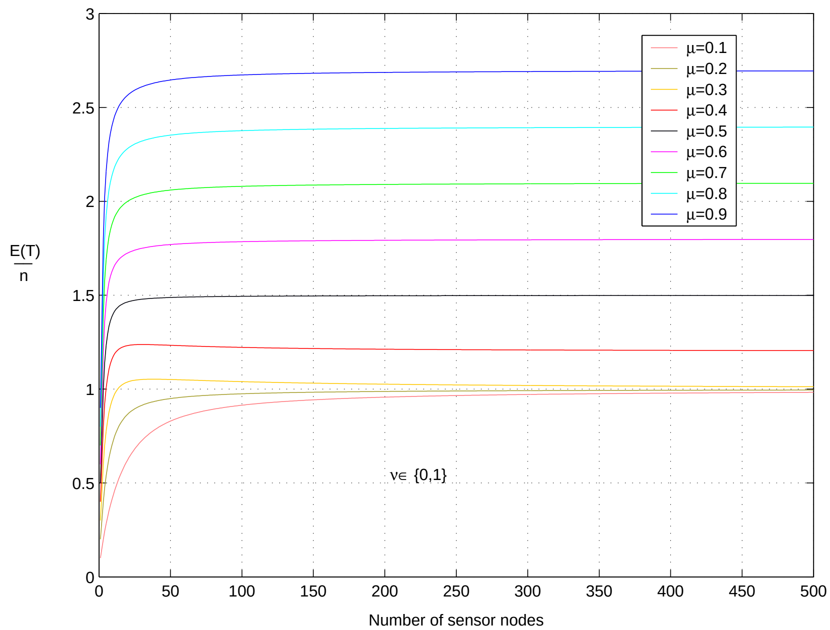


Figure 3.11: Average collection time as a function of average number of packets per node and number of nodes in line network equipped with omnidirectional antennas. Nodes carry 0 or 1 packet with probability $1 - \mu$ and μ respectively.

In omnidirectional antenna systems, data transmissions generate interference at all surrounding nodes. In a line network, in particular, this implies that a packet transmission to the left (or right) neighbor creates interference at both the left and right neighbors. This in turns increases the length of the optimum data collection schedule (when compared to directional systems). So time efficiency may be improved by using directional antenna systems. In order to get a better intuition on how the

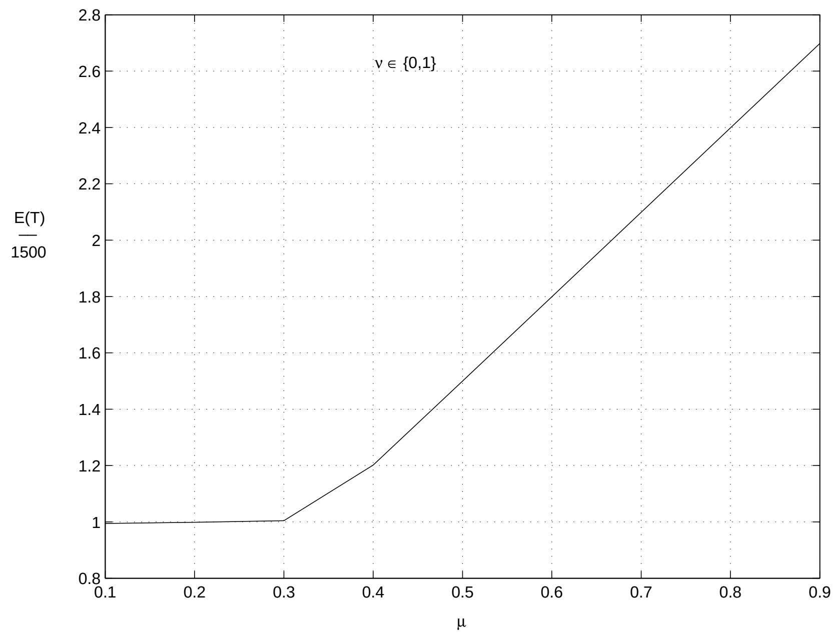


Figure 3.12: Average collection time as a function of average number of packets per node in 1500-node line network equipped with omnidirectional antennas. Nodes carry 0 or 1 packet with probability $1-\mu$ and μ respectively.

two systems perform relative to each other, we give the following comparative result for a line network.

Theorem 3.3.4. *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, if T_o (resp. T_u) denotes the minimum collection time on a line network equipped with omnidirectional (resp. directional) antennas*

$$\lim_{n\to\infty} \frac{\mathbb{E}\{T_o\}}{\mathbb{E}\{T_u\}} = \begin{cases} 1 & \text{if } \mu \geq 1/3 \\ 3\mu & \text{if } 1/3 \leq \mu \leq 0.5 \\ 3/2 & \text{if } \mu \geq 0.5 \end{cases} \tag{3.55}$$

3.3.3 Multi-line/Omnidirectional Case

Theorem 3.3.5. *Consider a multi-line network with L lines of length n_0, and ν_i's are i.i.d. chosen from $\mathcal{S}_m$ such that $\forall k$, $0 \le k \le m-1$, $\Pr(\nu_i = k) = p_k$ where $p_{m-1} \neq 0$. Further assume that $\mathbb{E}(\nu_i) = \mu$ Then*

$$n_0 L\mu \le \mathbb{E}(T_o) \le n_0 L\mu + O\left(\frac{1}{L}\right) + (3n_0(m-1)-2)\,(1-p_{m-1})^{L/3} \tag{3.56}$$

In particular,

$$i) \quad \text{if } L > (3+O(1))\log_\alpha n_o, \ \lim_{n_o\to\infty} \mathbb{E}(T_o) - n_o L\mu = 0 \tag{3.57}$$

$$ii) \quad \text{if } 3\log_\alpha n_o > L \text{ and } \lim_{n_o\to\infty} L = +\infty, \ \lim_{n_o\to\infty} \frac{\mathbb{E}(T_o)}{n_o L\mu} = 1 \tag{3.58}$$

$$iii) \quad \text{if } L = cte, n_o L\mu \le \mathbb{E}(T_o) \le n_o L\mu(1+\epsilon) \tag{3.59}$$

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

Proof. This follows by taking the expectation from both sides of the inequality $T \ge$ (number of packets in network). ☐

In what follows, we prove that as L increases, the expected collection time converges toward this lower bound.

To prove our asymptotic result, we use a suboptimal procedure to collect the data at the BS: we divide the network into three subnetworks $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$. Line $l \in \mathcal{S}_1$ if $l \equiv 0 \pmod 3$, $l \in \mathcal{S}_2$ if $l \equiv 1 \pmod 3$, $l \in \mathcal{S}_3$ if $l \equiv 2 \pmod 3$. For $l \in \mathcal{S}_1$, nodes at distance $d \equiv 1 \pmod 3$ from the BS transmit toward the BS at time slots $t \equiv 1 \pmod 3$, nodes at distance $d \equiv 2 \pmod 3$ transmit at times $t \equiv 0 \pmod 3$, nodes at distance $d \equiv 2 \pmod 3$ transmit at times $t \equiv 2 \pmod 3$. For $l \in \mathcal{S}_2$, nodes at distance $d \equiv 1 \pmod 3$ from the BS transmit toward the BS at time slots $t \equiv 2 \pmod 3$, nodes at distance $d \equiv 2 \pmod 3$ transmit at times $t \equiv 1 \pmod 3$, nodes at distance $d \equiv 0 \pmod 3$ transmit at times $t \equiv 0 \pmod 3$. And so on. In a given subnetwork multiple nodes at distance 1 from the BS may carry packets. Since the BS can only receive one packet at a time, we assume the presence of some mechanism that ensures that only one packet is transmitted to the BS and other available packets are stored for later transmission. This strategy is followed until all packets have reached a node at distance 1 from the BS. At this point packets are simply transmitted to the BS in turn so that the BS doesn't become idle until all packets have been collected. Furthermore we require that a given packet leaves its source node at the same TS it would have if all nodes were carrying $m-1$ packets (worst case scenario) and not sooner. With this scheduling and assuming each node has at most $m-1$ packets, it is clear that after $3n_0(m-1)-2$ TS, all the packets are within distance one from the BS.

Proof. The lower bound follows from Theorem 3.30. For the purpose of deriving an upper bound, we define the random variable $e_i \in \{0,1\}$, for $i = 1,\ldots,3n_0(m-1)-2$, 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 the previously described scheduling strategy, if the total number of packets is greater than $3n_0(m-1)-2$, 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 $1 \leq t \leq 3n_0(m-1)-2$ which is equal to $\sum_{i=1}^{3n_0(m-1)-2} e_i$. Therefore, we have the following upper bound for the delay

$$T \leq \max\{\eta,\ 3n_0(m-1)-2\} + \sum_{i=1}^{3n_0(m-1)-2} e_i \tag{3.60}$$

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

$$\begin{aligned}
\Pr(e_{3k} = 0) &\geq \Pr(\text{having at least m} - 1 \text{ packets at distance } k) \\
&\geq \Pr(\text{at least 1 node at distance k has m} - 1 \text{ pckts}) \\
&= 1 - (1 - p_{m-1})^{L/3}
\end{aligned} \tag{3.61}$$

A similar expression can be written for $\Pr(e_{3k+1/2} = 0)$. 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

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

which implies that $\Pr(\eta \leq 3n_0(m-1) - 2) \leq O\left(\frac{1}{n_0 L}\right)$. Now we can take the expectation from both sides of (3.60) to get

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

that completes the proof for (3.56). □

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 $\mathbb{E}(T) \geq \mu \sum_{l=1}^{L} n_0^l$. We can further find an upper bound by replacing n_0 by max ${n^l}_0$ in (3.60) 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 $E\{\eta\}$. Fig. 3.13 shows the average (minimum) collection time in a multi-line network equipped with omnidirectional antennas. In our scenario each line has at most 25 sensor nodes. The average number of packets per node is μ as well with a maximum of 1 packet per node. Those results were obtained through Monte Carlo simulations with 40000 iterations and confirm the convergence shown in Theorem 3.3.5.

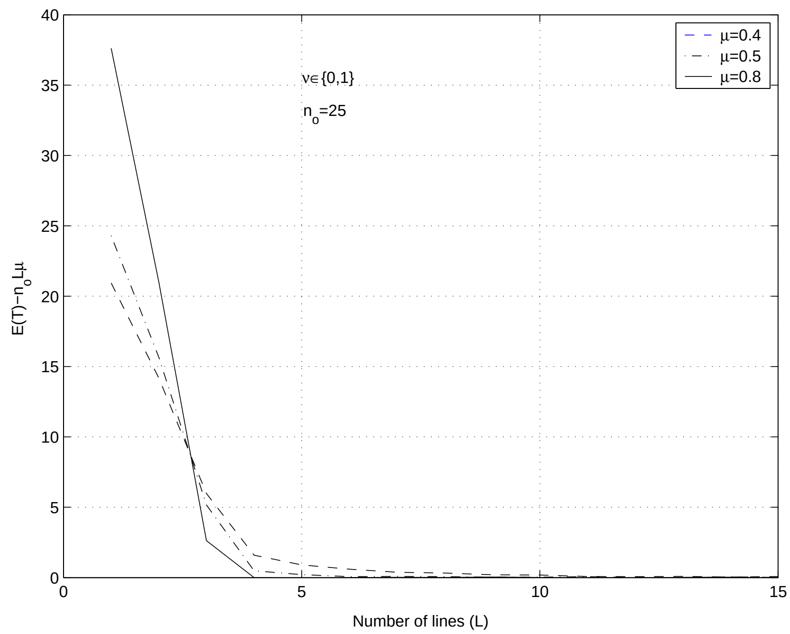


Figure 3.13: Omnidirectional Line Network, $\nu \in \{0,1\}$, $\mathbb{E}(T) - n_0 L\mu$ as a function of L and μ.

3.4 Dynamic Data Collection in Linear Sensor Networks

In stationary state, after nodes have organized themselves into a network, the operation of a sensor network can be broken down into two main phases. 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 (BS) located within the sensor network.

we assume that while data is being collected by the BS from sensor nodes, those sensor nodes keep gathering new data to be transmitted at a later time. We specifically assume that each sensor node collects data according to a Poisson distribution with mean λ.

Let $\boldsymbol{\nu}^0$ denote the initial data vector. Then $T(\boldsymbol{\nu}^0)$ denotes the corresponding data collection time and can be calculated according to Eq. (2.3). During that period new data is collected at sensor nodes and by the end of that period we have a new data vector $\boldsymbol{\nu}^1$. The corresponding data collection time is $T(\boldsymbol{\nu}^1)$. And so on. The data collected during a given transfer phase was collected by the sensor nodes during the previous transfer phase. We would like to find the maximum rate at which data may be gathered such that the system is stable. In a stable system the collection time remains bounded or equivalently the sensor node buffer size is bounded.

Theorem 3.4.1. *Data collection on a linear network consisting of n sensor nodes gathering data packets according to a Poisson distribution with rate λ is sustainable with high probability over the long term for large n iff $\lambda = o(1/n)$. In particular we have*

$$\lambda = o(1/n) \Rightarrow \exists K > 0\, K < \infty \text{ such that } \lim_n \Pr\{\nu_i^k \leq K\} = 1\, \forall i\, 1 \leq i \leq n \tag{3.62}$$

Proof. The distribution of the packets at the end of observation phase $k-1$ is

$$\Pr\{\nu_i^k = j\} = \exp(-\lambda T(\boldsymbol{\nu}^{k-1}))\frac{(\lambda T(\boldsymbol{\nu}^{k-1}))^j}{j!} \tag{3.63}$$

Therefore,

$$\Pr\{\nu_i^k \leq K\} = \sum_{j=0}^{K} \exp(-\mu)\frac{\mu^j}{j!} \tag{3.64}$$

where $\mu = \lambda T(\boldsymbol{\nu}^{k-1})$.

$$\Rightarrow \Pr\{\max_i \nu_i^k \leq K\} = \exp(-\mu n)\left(\sum_{j=0}^{K}\frac{\mu^j}{j!}\right)^n \tag{3.65}$$

$$= \exp(-\mu n)\left(\exp(\mu) - \frac{\mu^{k+1}}{(k+1)!}\exp(\theta\mu)\right)^n,\ 0 \leq \theta \leq 1 \tag{3.66}$$

$$= \left(1 - \frac{\mu^{k+1}}{(k+1)!}\exp((\theta-1)\mu)\right)^n \tag{3.67}$$

$$\geq \left(1 - \frac{\mu^{k+1}}{(k+1)!}\right)^n \tag{3.68}$$

Choose

$$\frac{\mu^{k+1}}{(k+1)!} = \frac{1}{n \log n} \tag{3.69}$$

$$\Rightarrow K = -\frac{\log(n \log n)}{\log \mu} \tag{3.70}$$

But we know from Eq. (2.3) that

$$T(\boldsymbol{\nu}) = O(n) \tag{3.71}$$

therefore choose

$$\lambda = \frac{1}{n^{1+\epsilon}} \tag{3.72}$$

$$\Rightarrow K = 1/\epsilon \tag{3.73}$$

Therefore if $\lambda < 1/n$, $\Pr\{\max_i \nu_i^k \leq K\}$ converges to 1 when the number of nodes becomes large for some finite K. On the other it is easy to see that if $\lambda = 1/n$, $\Pr\{\max_i \nu_i^k \leq K\}$ converges to 0 for any positive, finite K from Eq. (3.67). The theorem follows. □

3.5 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 as demonstrated in the previous chapter. 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, we derive 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.

Chapter 4 Conclusions and Future Directions

Data collection in an important communication primitive in sensory networks. In this dissertation we studied the data collection process and its fundamental performance limits. Specifically,

- In the deterministic case, we exhibited optimal scheduling strategies to collect data on trees and derived corresponding minimal data collection times. Furthermore we bounded time collection on networks with cycles.
- In the random case, we derived the expected value of the minimum collection time in trees.
- We studied the impact of hop length, packet splitting, packet erasure, and lack of synchronization in line networks. Furthermore we found scaling conditions on the rate at which data may be gathered by sensor nodes for sustainable data collection over time.

Future directions for our research include:

- Consider the presence of multiple base stations in the network, and quantify the corresponding gains in data collection time.
- We studied the impact of noise in the channel on our data collection strategies. Those strategies are optimal in the absence of noise. Are those strategies still optimal in the presence of noise? Can we come up with strategies that are less sensitive to noise?
- Similarly we studied the impact of partial lack of synchronization among the sensor nodes. Are the strategies presented in this thesis still optimal in the absence of clock synchronization?

- Define energy expended as $energy(h) = \sum_{N_i} l(N_i)^\alpha$, where α is a positive constant greater than 2, h is the maximal node transmission range, and $l(N_i)^\alpha$ denotes the energy expended by node N_i ($l(N_i)$ is taken to be the length in hops of a transmission initiated by node N_i). In this thesis, we derived the minimum collection time at the point of minimum energy (one hop transmissions). We further studied minimum collection time as a function of transmission range. An interesting follow-up would be to derive optimal schedules with respect to time *and* energy for a given transmission range (larger than 1) and subsequently the trade-off between minimum delay and energy expended in data collection. This is illustrated in the next two figures where $h = 3$ and a line network is considered. Fig. 4.1 shows an optimal data collection schedule with respect to time, while Fig. 4.2 shows a schedule optimal with respect to time and energy. If$\alpha = 2$, the energy expended in the first schedule is 5*9+2*4+2=55 energy units while it is only 2*9+2*4+11=37 in the second schedule.

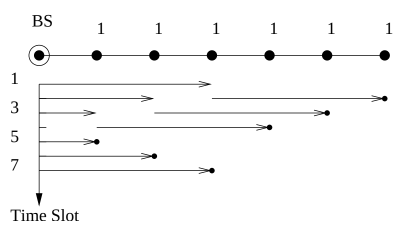


Figure 4.1: Optimal distribution schedule with respect to time that is suboptimal with respect to energy.

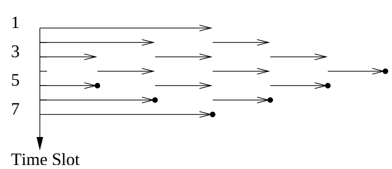


Figure 4.2: Optimal schedule with respect to time and energy.

Appendix A A Preliminary Result

In the following section we assume that a network equipped with directional nodes may receive *and* transmit a data packet during any given TS (whereas so far we had assumed that it was only possible to receive *or* transmit a data packet in a given TS). Although such networks may seem artificial and not practical for the time being, the results that follow allow us to gain some insight into more complex systems.

The purpose of this section is the construction of an optimal strategy for collecting data as well as deriving a closed form expression for time performance. We obtain both for *any* general connected graphs. To that end, we first go through a series of successive building steps.

Lower Bound on the Time Performance of Data Distribution Algorithms

Lemma A.0.1. *Given any connected graph G, if $t_1(G)$ denotes the time performance of a given data distribution algorithm, and ν_j denotes the number of data packets at distance j from the BS, then*

$$t_1(G) \geq \max_i \, (i - 1 + \sum_{j \geq i} \nu_j) \tag{A.1}$$

Proof. $\sum_{j\geq 1} \nu_j$ data packets must be delivered to nodes at distance greater than 1. Since the BS can only transmit one data packet at a time, we have: $t_1(G) \geq \sum_{j\geq 1} \nu_j$.

$\sum_{j\geq i} \nu_j$ data packets must be delivered to nodes at distance greater than $i > 1$. After $\sum_{j\geq i} \nu_j$ TS the last data packet sent by the BS is at distance one from the BS and therefore at least $i - 1$ extra TS are required for it to reach its destination, thus: $t_1(G) \geq \sum_{j\geq i} \nu_j + i - 1$. Hence the stated result. □

Achievability of Lower Bound

1) Line Network:

The purpose of this section is to prove that the lower bound derived in the previous section is achievable on a line network. We shall show in the next section achievability on general connected graphs based on this result.

The algorithm:

The BS is to 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). This algorithm is illustrated by an example in Fig. A.1.

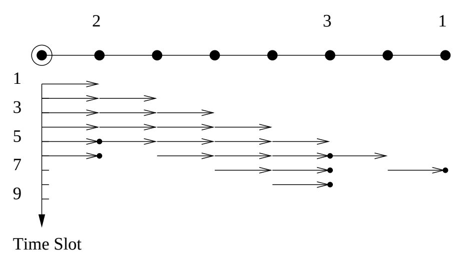


Figure A.1: Optimal distribution schedule for BS in line network equipped with directional antennas and ability to receive and transmit in the same TS. The completion time is 8 TS.

Proof of optimality and time performance:

Denote T_i the last busy time slot at node i in the execution of our algorithm. Clearly then our algorithm runs in $\max_{1 \le i \le n} \{Ti\}$. T_i is a function of the distance to the BS, the number of data packets destined for node i and the number of data packets forwarded by node i.

Lemma A.0.2.

$$T_i = \begin{cases} i + \sum_{j>i} \nu_j & \mathit{if}\ \nu_i \le 1 \\ i - 1 + \sum_{j \ge i} \nu_j & \mathit{if}\ \nu_i \ge 1 \end{cases} \tag{A.2}$$

Proof.

$$\nu_i \leq 1 \Rightarrow T_i = (f_i + 1) + (i - 1)$$
$$\nu_i > 1 \Rightarrow T_i = (f_i + 1) + \nu_i - 1) + (i - 1)$$
$$f_i = \text{number of packets forwarded by } i = \sum_{j>i} \nu_j$$

□

Lemma A.0.3. *Define:* $S_i = \sum_{j \geq i} \nu_j + i - 1$, *then* $\max_i S_i = \max_i T_i$

Proof. Indeed S_i is a lower bound for all i. So $\max_i S_i \leq \max_i T_i$, but $S_i = T_i$ if $\nu_i \geq 1$. Since clearly $\max_i T_i$ occurs in i such that $\nu_i \geq 1$, we have $\max_i S_i = \max_i T_i$, i.e., the algorithm is optimal. □

2) General Connected Graphs:

By using the shortest routes (from the BS) to the sensor nodes, the algorithm previously described on line networks may be used on general (connected) graphs. The performance time of that algorithm is then $\max_i T_i$ where T_i is defined in Lemma A.0.2 and ν_j is the number of data packets at distance j from the BS. The next corollary follows from Lemma A.0.3.

Corollary A.0.4. *The minimum data collection time* $t_1(G)$ *on any connected graph* G *is*

$$t_1(G) = \max_{1 \leq i \leq n} (i - 1 + \sum_{j \geq i} \nu_j) \tag{A.3}$$

The following corollary follows from Corollary A.0.4.

Corollary A.0.5.

$$\forall \mathcal{T} \text{ a spanning tree of } G, \quad t_1(\mathcal{T}_{SP}) \leq t_1(\mathcal{T}) \tag{A.4}$$

Appendix B Algorithms

Algorithm 1 (for directional antenna systems) and Algorithm 3 (for omnidirectional ones), running at the BS, optimally distribute data in a line network. Given a line network $Network = \boldsymbol{\nu}$, they dictate the BS actions at each TS: Remain idle ($action = 0$) or transmit ($action = 1$). The result is stored in the vector *action*. When an action is chosen the right packet is to be handed over to the BS for transmission. One might assume that there is a stack of data packets correctly ordered with respect to the distance to the BS and that that stack is being updated after each BS action so that a packet is popped off the stack as it is transmitted. Algorithm 2 (for directional antenna systems) and Algorithm 4 (for omnidirectional ones), running at the BS, optimally distribute data in a multiline network. The input to Algorithms 2 and 3 is a n by m matrix $Network$ where n is the number of lines and m is the maximum number of nodes per line. It is further assumed that the vector Est_trans_time of size n is initialized with the respective $T(\boldsymbol{\nu})$ of each line.

Algorithm 1 Determines BS actions in line networks

input: Network
output: action

step $\leftarrow$ 1, legal$\leftarrow$ 1, packts_left $\leftarrow \sum_i$Network(i)
while packts_left $\neq 0$ **do**
  if legal **then**
    action(step) $\leftarrow$ 1
    packts_left $\leftarrow$ packts_left-1
    legal $\leftarrow$ 0
  else
    action(step) $\leftarrow$ 0
    legal $\leftarrow$ 1
  end if
  if packts_left $<$ Network(1) **then**
    legal $\leftarrow$ 1
  end if
  step $\leftarrow$ step+1
end while

Algorithm 2 Determines BS actions in multi-line networks

input: Network
output: action

```
step ← 1, prev_legal ← ones(1,n), legal ← ones(1,n), packts_left ← ∑_{i,j} Network(i,j)
∀i packts_left_for_branch(i)← ∑_j Network(i,j)
while packts_left ≠ 0 do
  (y,ind)=max(Est_trans_time.*legal)
  if y=0 then
    for i=1 to nb_of_branches do
      if packts_left_for_branch(i) ≠ 0 then
        ind=i
      end if
    end for
    action(step) ← 0
  else
    action(step) ← ind
    packts_left ← packts_left-1
    packts_left_for_branch(ind) ← packts_left_for_branch(ind)-1
  end if
  legal← ones(1,nb_of_branches)
  for i=1 to nb_of_branches do
    if packts_left_for_branch(i)=0 then
      legal(i) ← 0
    end if
  end for
  tabtest ← sum(Network(ind,1:nb_of_nodes))-Network(ind,1)
  if (tabtest > 0 & action(step) ≠ 0) then
    if packts_left_for_branch(ind) ≥ Network(ind,1) then
      legal(ind)← 0
    end if
  end if
  for i=1 to nb_of_branches do
    if (prev_legal(i)=1 & i ≠ ind) then
      Est_trans_time(i) ← Est_trans_time(i)+1
    end if
  end for
  prev_legal ← legal
  step ← step+1
end while
```

Algorithm 3 Determines BS actions

input: Network
output: action

```
step ← 1, packts_left1 ← Network(1), packts_left2 ← Network(2), packts_left3 ← ∑_{i≥3} Network(i), packts_left ← ∑_i Network(i)
while packts_left ≠ 0 do
    while packts_left3 ≠ 0 do
        action(step) ← 1
        action(step+1) ← 0
        action(step+2) ← 0
        step=step+3
        packts_left3=packts_left3-1
    end while
    while packts_left2 ≠ 0 do
        action(step) ← 1
        action(step+1) ← 0
        step=step+2
        packts_left2=packts_left2-1
    end while
    while packts_left1 ≠ 0 do
        action(step) ← 1
        step=step+1
        packts_left1=packts_left1-1
    end while
    packts_left ← packts_left-1
end while
```

Algorithm 4 determines BS actions in multi-line network

input: Network
output: action

```
step ← 1, prev_legal ← ones(1,n), legal ← ones(1,n)
packts_left ← ∑_{i,j} Network(i,j)
∀i packts_left_for_branch(i)← ∑_j Network(i,j)
while packts_left ≠ 0 do
   (y,ind)=max(Est_trans_time.*legal)
   if y=0 then

      for i=1 to nb_of_branches do

         if packts_left_for_branch(i) ≠ 0 then
            ind=i
         end if
      end for
      action(step) ← 0
   else
      action(step) ← ind
      packts_left ← packts_left-1
      packts_left_for_branch(ind) ← packts_left_for_branch(ind)-1
   end if
   legal← ones(1,nb_of_branches)
   for i=1 to nb_of_branches do

      if packts_left_for_branch(i)=0 then
         legal(i) ← 0
      end if
   end for
   tabtest ← sum(Network(ind,1:nb_of_nodes))-Network(ind,1)
   if (tabtest > 0 & action(step) ≠ 0) then

      if packts_left_for_branch(ind) ≥ Network(ind,1) then
         legal(ind)← 0
      end if
   end if
   for i=1 to nb_of_branches do
      if (prev_legal(i)=1 & i ≠ ind) then
         Est_trans_time(i) ← Est_trans_time(i)+1
      end if
   end for
   prev_legal ← legal
   step ← step+1
end while
```

Bibliography

[1] J. R. Agre, *et al.,* "Development Platform for Self-Organizing Wireless Sensor Networks," *Proceedings of SPIE, Unattended Ground Sensor Technologies and Applications,* Vol. 3713, 1999.

[2] J. Agre, and L. Clare, "An Integrated Architecture for Cooperative Sensing Networks," *Computer,* Vol. 33, 2000.

[3] M. Ahmed, and G. J. Pottie, "Information Theory of Wireless Sensor Networks: the n-helper Gaussian Case," *Proceedings of IEEE International Symposium on Information Theory (ISIT'00),* 2000.

[4] I. F. Akyildiz, *et al.,* "A survey on Sensor Networks," *IEEE Communications Magazine,* Aug. 2002.

[5] J. Barros, and S. D. Servetto, "Coding Theorems for the Sensor Reachback Problem with Partially Cooperating Nodes," *American Mathematical Society (AMS), Discrete Mathematics and Theoretical Computer Science (DIMACS) series on Network Information Theory.*

[6] M. Bhardwaj, A. Chandrakasan, and T. Garnett, "Upper Bounds on the Lifetime of Sensor Networks." *Proceedings of IEEE International Conference on Communications,* 2001.

[7] V. Bharghavan *et al.,* "MACAW: A Media Access Protocol for Wireless LAN," *Proceedings of the SIGCOMM 94 Conference of Communications architectures, protocols and applications*, Aug. 1994.

[8] E. S. Biagioni, and G. Sasaki, "Wireless Sensor Placement for Reliable and Efficient Data Collection," *Proceedings of Hawaii International Conference on System Science (HICSS'03),* Hawaii, 2003.

[9] J. Bruck,*et al.* , "Efficient Algorithms for All-to-All Communications in Multi-Port Message-Passing Systems," *IEEE Transactions on Parallel and Distributed Systems,* Vol. 8, No. 11, Nov. 1997.

[10] A. Cerpa, *et al.,* "Habitat Monitoring: Application Driver for Wireless Communications Technology," *Proceedings of the 2001 ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean,* Apr. 2001.

[11] C. Chien, I. Elgorriaga, and C. McConaghy, "Low-Power Direct-Sequence Spread-Spectrum Modem Architecture for Distributed Wireless Sensor Networks," *Proceedings of International Symposium on Low Power Electronics and Design (ISLPED'01),* Huntington Beach,CA, Aug. 2001.

[12] B. S. Chlebus *et al.,* "Deterministic Broadcasting in Ad-hoc Radio Networks,"*Distributed computing,* 2002.

[13] S. Cui, A. J. Goldsmith and A. Bahai, "Energy-Efficiency of MIMO and Cooperative MIMO Techniques in Sensor Networks," *To appear in IEEE Journal on Selected Areas in Communications,* 2004.

[14] A. F. Dana, and B. Hassibi, "On the Power Efficiency of Sensory and Ad-hoc Wireless Networks," *Submitted to IEEE Transactions on Information Theory,* 2003.

[15] L. Doherty, K. S. J. Pister, and L. E. Ghaoui, "Convex Position Estimation in Wireless Sensor Networks," *Proceedings of IEEE Symposium on Information and Communication (INFOCOM'01),* Anchorage, AK, Apr. 2001.

[16] E. Duarte-Melo, and M. Liu, "Analysis of Energy Consumption and Lifetime of Heterogeneous Wireless Sensor Networks", *Proceedings of the IEEE Global Telecommunications Conference (Globecom'02),* Nov. 2002.

[17] E. J. Duarte-Melo and M. Liu, "Data-Gathering Wireless Sensor Networks: Organization and Capacity," *Computer Networks (COMNET) Special Issue on Wireless Sensor Networks,* Vol. 43, Issue 4, Nov. 2003.

[18] E. J. Duarte-Melo, M. Liu and A. Misra, "A Modeling Framework for Computing Lifetime and Information Capacity in Wireless Sensor Networks," *2nd WiOpt: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks,* Mar. 2004.

[19] P. Erdös and M. Kac, "On Certain Limit Theorems of the Theory of Probability," *Bulletin of the American Mathematical Society,* 1946.

[20] D. Estrin, *et al.,* "Embedding the Internet," *Communications of the ACM,* Vol. 43, May 2000.

[21] D. Estrin, *et al.,* "Connecting the Physical World with Pervasive Computing," *IEEE Pervasive Computing,* Vol. 1, No. 1, Jan.-Mar. 2002.

[22] C. Florens and R. J. McEliece, "Scheduling Algorithms for Wireless Ad-hoc Sensor Networks," *Proceedings of the IEEE Global Telecommunications Conference (Gobecom'02)*, Nov. 2002.

[23] C. Florens and R. J. McEliece, "Packets Distribution Algorithms for Sensor Networks," *Proceedings of IEEE Symposium on Information and Communication (INFOCOM'03),* 2003.

[24] C. Florens, M. Franceschetti, and R. J. McEliece, "Lower Bounds on Data Collection Time in Sensory Networks," *Journal of Selected Areas in Communications,* 2004.

[25] C. Florens, M. Sharif, and R. J. McEliece, "Random Sensory Networks: A Delay Analysis,"*Proceedings of the 40th Annual Allerton Conference on Communication, Control and Computing* 2004.

[26] C. Florens, M. Sharif, and R. J. McEliece, "Random Sensory Networks: A Delay Analysis,"*Submitted to IEEE Transactions on Information Theory,* 2004.

[27] C. Florens, M. Sharif, and R. J. McEliece, "Delay Issues in Linear Sensory Networks,"*Proceedings of International Symposium on Information Theory (ISIT'04),* 2004.

[28] P. Fraigniaud, "Minimum-Time Broadcast under Edge-Disjoint Paths Modes," *Proceedings of International Conference on Fun with Algorithms,* 2001.

[29] P. Fraigniaud, "Approximation algorithms for Minimum-Time Broadcast under the vertex-disjoint Paths Modes," *Proceedings of European Symposium on Algorithms,* 2001.

[30] J. Gao, "Analysis of Energy Consumption for Ad-Hoc Wireless Sensor Networks Using a Bit-Meter-per-Joule Metric," *IPN Progress Report 42-150,* Jet Propulsion Laboratory, Aug. 2002.

[31] Y. Gao, K. Wu, and F. Li, "Analysis on the Redundancy of Wireless Sensor Networks," *Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications,* San Diego, CA, USA, 2003.

[32] M. Gastpar, and M. Vetterli, "On the Capacity of Large Gaussian Relay Networks," *Submitted to IEEE Transactions on Information Theory,* 2003.

[33] P. Gupta, and P. R. Kumar, "Critical Power for Asymptotic Connectivity," *Proceedings of the IEEE 37th Conference on Decision and Control,* 1998.

[34] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks,"*IEEE Transactions on Information Theory,* Vol. 46, Mar. 2000.

[35] W. R. Heinzelman, J. Kulic, and H. Balakrishnan, "Adaptive Protocols for Information Dissemination in Wireless Sensor Networks," *Proceedings of the ACM MobiCom'99,* Seattle, WA, 1999.

[36] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," *IEEE Proceedings of Hawaii international Conference on Systems Science,* Jan. 2000.

[37] A. Hu, and S. D. Servetto, "Optimal Detection for a Distributed Transmission Array," *Proceedings of IEEE International Symposium on Information Theory (ISIT'03),* Yokohama, Japan, Jun.-Jul. 2003.

[38] A. Hu, and S. D. Servetto, "Asymptotically Optimal Time Synchronization in Dense Sensor Networks," *Proceedings of ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'03),* San Diego, California, USA. Sept. 2003.

[39] C. Huang, and Y. Tseng, "The Coverage Problem in a Wireless Sensor Network," *Proceedings of ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'03),* San Diego, CA, Sept. 2003.

[40] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking* Aug. 2000.

[41] E. Jennings *et al.,* "Space based multihop networking," *Submitted to Special Issue of Computer Networks (Elsevier),* 2004.

[42] J. H. Ju and V. O. K. Li, "TDMA Scheduling Design of Multihop Packet Radio Networks Based on Latin Squares," *IEEE Journal on Selected Areas in Communications,* Vol. 17, No. 8, Aug. 1999.

[43] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Next Century Challenges: Mobile Networking for Smart Dust," *Proceedings of the fifth annual ACM/IEEE International Conference on Mobile Computing and Networking,* Aug. 1999.

[44] F. T. Leighton, B. M. Maggs, and S. B. Rao, "Packet Routing and Job-Shop Scheduling in 0(Congestion+Dilation) Steps," *Combinatorica*, Vol. 14, 1994.

[45] J. Lundquist, D. Cayan, and M. Dettinger, "Meteorology and Hydrology in Yosemite National Park: A Sensor Network Application," *Proceedings of Information Processing in Sensor Networks (IPSN'03)*, Apr. 2003.

[46] A. Mainwaring, *et al.,* "Wireless Sensor Networks for Habitat Monitoring," *Proceedings of ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'02),* Atlanta, GA, Sept. 2002.

[47] D. Marco, *et al.,* "On the Many-to-One Transport Capacity of a Dense Wireless Sensor Network and the Compressibility of its Data," *Proceedings of International Workshop on Information Processing in Sensor Networks (IPSN),* Apr. 2003.

[48] S. Meguerdichian, *et al.,* "Coverage Problems in Wireless Ad-Hoc Sensor Networks," *Proceedings of the Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'01),* Vol. 3, Apr. 2001.

[49] R. Min, *et al.,* "Low-Power Wireless Sensor Networks," *Fourteenth International Conference on VLSI Design,* 2001.

[50] C. Peraki, and S. D. Servetto, "On the Maximum Stable Throughput Problem in Random Networks with Directional Antennas," *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc),* Annapolis, MD, Jun. 2003.

[51] A. Perrig, *et al.,* "SPINS: Security Protocols for Sensor Networks," *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom'01),* 2001.

[52] G. J. Pottie, "Multi-node Processing Problems in Distributed Sensor Networks," *IEEE International Symposium on Information Theory (ISIT'98),* Cambridge, MA, Aug. 1998.

[53] G. Pottie, "Wireless Sensor Networks," *Proceedings of Information Theory Workshop,* Jun. 1998.

[54] G. J. Pottie, and W. J. Kaiser, "Wireless Integrated Network Sensors," *Commununications of the ACM,* Vol. 43, May 2000.

[55] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed Compression in a Dense Microsensor Network," *IEEE Signal Processing Magazine,* Mar. 2002.

[56] M. Pursley, "The Role of Spread Spectrum in Packet Radio Networks," *Proceedings of the IEEE*, Jan. 1987.

[57] J. M. Rabaey, *et al.,* "PicoRadio Supports Ad Hoc Ultra Low Power Wireless Networking," *IEEE Computer,* Vol. 33, July 2000.

[58] V. Raghunathan, *et al.* , "Energy-Aware Wireless Microsensor Networks,"IEEE Signal Processing Magazine, Mar. 2002.

[59] E. M. Royer, and C. -K. Toh, "A Review of Current Routing Protocols for Ad-Hoc Wireless Networks," *IEEE Personal Communications Magazine,* Apr. 1999.

[60] A. Scaglione, and S. D. Servetto, "On the Interdependence of Routing and Data Compression in Multi-Hop Sensor Networks," *ACM/Kluwer Journal on Mobile Networks and Applications (MONET)—Selected (and revised) papers from ACM MobiCom 2002,* To appear.

[61] L. Schwiebert, S. K. S. Gupta, and J. Weinmann, "Research Challenges in Wireless Networks of Biomedical Sensors," *Mobile Computing and Networking*, 2001.

[62] S. D. Servetto, "Quantization with Side Information: Lattice Codes, Asymptotics, and Applications in Sensor Networks," *Submitted to IEEE Transactions on Information Theory,* Mar. 2002.

[63] S. D. Servetto, "Distributed Signal Processing Algorithms for the Sensor Broadcast Problem," *Proceedings of the 37th Annual Conference on Information Sciences and Systems (CISS),* Baltimore, MD, Mar. 2003.

[64] S. D. Servetto, and G. Barrenechea, "Constrained Random Walks on Random Graphs: Routing Algorithms for Large Scale Wireless Sensor Networks," *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'02),* Atlanta, GA, Sept. 2002.

[65] E. Shih *et al.* , "Physical Layer Driven Protocol and Algorithm Design for Energy-Efficient Wireless Sensor Networks," *Proceedings of the 7th ACM In-*

ternational Conference on Mobile Computing and Networking (MobiCom'01), Rome, Italy, Jul. 2001.

[66] A. Sinha, and A. Chandrakasan, "Dynamic Power Management in Wirelss Sensor Networks," *IEEE Design & Test,* Mar.-Apr. 2001.

[67] K. Sohrabi,*et al.,* "Protocols for Self-Organization of a Wireless Sensor Network," *IEEE Personal Communications Magazine*, Vol. 7, No. 5, Oct. 2000.

[68] D. Steere, *et al.,* "Research Challenges in Environmental Observation and Forecasting Systems," *Proceedings of the 6th annual international Conference on Mobile Computing and Netwworking (MobiCom'00),* 2000.

[69] S. Tilak, N. B. Abu-ghazaleh, and W. Heinzelman, "A Taxonomy of Wireless Micro-Sensor Network Models," *ACM Mobile Computing and Communications Review (MC2R),* 2002.

[70] M. Tubaishat, S. Madria, "Sensor Networks: An Overview," *IEEE Potentials,* Apr.-May 2003.

[71] M. A. M. Vieira, *et al.,* "Survey on Wireless Sensor Network Devices," *Proceedings of IEEE International Conference on Emerging Technologies and Factory Automation (ETFA'03),* Vol. 1, Sept. 2003.

[72] A. Woo, and D. Culler, "A Transmission Control Scheme for Media Access in Sensor Networks," *Proceedings of the 7th annual international Conference on Mobile Computing and Netwworking (MobiCom'01,* Rome, Italy, Jul. 2001.