The Dissertation Committee for Seung Jun Baek
certifies that this is the approved version of the following dissertation:

Spatial Modelling and Analysis of Wireless Ad-hoc and Sensor Networks: An Energy Perspective

Committee:

Gustavo de Veciana, Supervisor

Aristotle Arapostathis

Ross Baldick

Lili Qiu

Sanjay Shakkottai

Spatial Modelling and Analysis of Wireless Ad-hoc and Sensor

Networks: An Energy Perspective

by

Seung Jun Baek; B.S., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2006

Acknowledgments

Spatial Modelling and Analysis of Wireless Ad-hoc and Sensor Networks: An Energy Perspective

Publication No. _____

Seung Jun Baek, Ph.D.
The University of Texas at Austin, 2006

Supervisor: Gustavo de Veciana

This dissertation focuses on modelling and analyzing the spatial characteristics of traffic in these networks so as to extend network lifetime for various application/traffic scenarios.

In the first part of the dissertation we consider large-scale sensor networks that systematically sample a spatio-temporal field. Firstly we formulate a distributed compression problem subject to aggregation costs to a single sink. We show that the optimal solution is based on ordering sensors according to aggregation costs. Next we consider a hierarchical model for a sensor network including sinks, compressors and sensors. We show that the optimal organization is associated with the Johnson-Mehl tessellation induced by nodes' locations. Our analysis and simulations show the proposed scheme can yield 8-28% energy savings depending on the compression ratio.

In the second part of the dissertation we investigate the use of proactive multipath routing in ad hoc wireless networks. The focus is on optimizing tradeoffs between the increased energy cost associated with spreading traffic and the improved spatial balance of energy burdens. We show how its optimization depends on the relative values of the energy reserves/storage, replenishing rates, and network load characteristics. In particular, we show that the degree of spreading should roughly scale as the square root of the bits-meters load offered by a session. Simulation

results confirm that proactive multipath routing decreases the probability of energy depletion by orders of magnitude versus that of a shortest path routing scheme when the initial energy reserve is high.

In the third part of the dissertation we consider a large sensor network with mobile sinks. The network makes use of aggregation nodes (AGNs), for compression and/or data fusion of locally sensed data. Since the aggregated data may cause a concentration of energy burdens when routed to sinks, we use proactive multipath routing between AGNs to mobile sinks. We show that the scale of aggregation and degree of spreading can be optimized. Particularly if the sensed data is bursty in space and time, then one can reap substantial benefits from aggregation and balancing.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

An emerging vision for the future is that of a physical (or virtual) world embedded with networked sensors and actuators. The interweaving of local sensing/actuation, communication, and computation with decision-making and control has broad applicability, including: transportation, environmental monitoring/oversight, structural monitoring, health care, and national security. Progress is being made toward increasing the diversity of available sensors, while furthering the technologies permitting the deployment of small, inexpensive devices operating on limited battery power and interconnected via wireless links [1–4]. A key challenge for such applications lies in devising system architectures to realize distributed sensing tasks subject to hard system constraints, e.g., energy.

Among the numerous elements that affect the energy consumption characteristics of large-scale ad-hoc/sensor networks, the *spatial loads* play a critical role. Due to radio power limitations, nodes must rely on hop-by-hop dissemination of information to extend the physical coverage area of such networks. The relaying of information between sources and sinks imposes processing and transmission overheads on intermediate nodes, leaving *spatial traces* of energy burdens.

To study the relationship between spatial traffic patterns, energy consumption and network architectures, we list the decisive factors that affect such traffic patterns:

- **Engineering of traffic:** one can devise network architectures that engineer the traffic in order to achieve energy efficiency and/or meet the needs of the application. Such schemes may induce *structural* flow of traffic, e.g., a network may have a hierarchical, possibly centralized, structure and thus the traffic flows follow specific patterns. By contrast, traffic can be unstructured or have flat hierarchy depending on the application.

	Patterns of Engineered Traffic	
Patterns of Exogeneous Traffic	Hierarchical	Flat
Homogeneous	SensorNet measuring data generated uniformly over the field	Ad-hoc Network with random traffic pattern
Bursty	SensorNet detecting bursty phenomena	

Table 1.1: Exemplifying scenarios for different cases of spatial traffic patterns.

- **Characteristics of exogenous traffic:** the spatial characteristics of traffic can be either homogeneous or bursty. Irrespective of efforts to engineer traffic loads, the energy performance can vary substantially depending on exogenous factors. Especially bursty nonhomogeneous traffic loads are deemed to have deleterious effect on energy consumption and thus require careful design so as to extend a network's lifetime.

In this thesis we examine several scenarios to see how these factors impact the lifetime of an energy constrained network. First we consider a sensor network in which a node simply acts as a persistent source, gathering data and reporting to a central station at a higher hierarchical layer whose role is to collect and process information. The mission of the network can be to either periodically collect data from every location in the field, or detect asynchronous and spatially bursty events. Second we consider the case where each network node is an active and independent entity so that information is generated and absorbed in an arbitrary manner. The network operates without infrastructure as would be the case in peer-to-peer networks, hence spatial and temporal traffic patterns may be dynamic in nature.

These scenarios are summarized in Table 1 with some examples. Albeit simplified, the scenarios cover a fairly wide range of practical applications of interest for wireless ad-hoc and sensor networks. In each scenario the network exhibits different energy consumption characteristics, thus its architecture must be carefully designed with an understanding of the associated traffic characteristics so as to maximize the network lifetime. The broad objectives of this thesis are as follows:

1. Devise optimal architectures and routing schemes, possibly in the presence of data compression capabilities, which induce good engineered traffic loads given that the pattern of exogenous traffic is spatially homogeneous.

2. Explore the design tradeoffs afforded by *proactive* multipath routing schemes for ad-hoc networks with unstructured traffic.

3. Identify relationships between design choices, e.g., data fusion, energy burden balancing, etc., and the spatio-temporal characteristics of traffic exhibiting bursty pattern.

4. Propose analytical models to study how to design and improve energy performance in such networks.

This thesis is organized as follows. In Chapter 2, we discuss the modelling and analysis of energy optimal routing schemes for large-scale sensor networks using the distributed data compression and hierarchical aggregation. In Chapter 3, we present energy balancing strategies for wireless and ad-hoc networks subject to unstructured loads based on proactive multipath routing, and study the fundamental tradeoffs between the cost and benefits of such schemes. In Chapter 4, we study a model for a large-scale sensor network which performs data fusion via aggregation and load balancing via traffic spreading. Specifically we consider a network detecting spatio-temporally bursty phenomena, and investigate the network lifetime when there are various spatial scales of aggregation, load balancing and dimensions of phenomena. Chapter 5 concludes this thesis and summarizes the key insights/results we have obtained.

Chapter 2

Distributed Data Compression and Hierarchical Aggregation of Traffic

2.1 Introduction

To design systems with hard energy constraints we need to resolve a number of issues, *e.g.,* how to design power-efficient micro-sensors [5], how to organize the aggregation of the sensor data [6], and how to minimize the energy consumption by effective routing/relaying [7, 8]. Broadly speaking, energy consumption can be reduced in two ways. First one can attempt to reduce the total traffic that has to be transmitted. For example the amount of traffic generated by the sensors can be reduced through distributed source coding [9–14] and/or data aggregation [6, 15]. Second one can minimize the energy consumption for each transmission. For example, energy-aware routing [16] and distributed medium access control [17] schemes have been proposed to streamline energy expended when carrying traffic on ad hoc networks.

In this chapter we focus on how structures for aggregation and compression can be combined so as to minimize energy costs for large scale sensor networks. We will focus on two questions:

1. What is an optimal compression and aggregation scheme for gathering spatially distributed sensor readings?

2. How can we design optimal hierarchical architectures for compression and aggregation that can scale to large sensor networks?

As discussed below there has been much closely related work in exploring both the possibility of data compression in the sensor network, and attempting to minimize the energy consumption in transporting data on wireless networks. In particular [18] considered the problem of disseminating sensor data across an ad hoc network

of sensors so that each obtains all the information. They show that asymptotically, as the number of sensors in the network increases, the overall traffic generated by the spatially correlated data sources increases, but at a rate of increase that is much slower than the capacity of an ad hoc wireless networks that supports independent sources. Thus they conclude that it is possible to operate a large sensor network monitoring correlated data sources, despite fundamental limits on the capacity of ad hoc networks[19]. The work in [20] proposes a concrete direction towards realizing a distributed data compression scheme based on tracking the correlation structure of the underlying phenomenon by a linear prediction mechanism and using this information to allocate different quantization levels among sensors. This work focuses on gathering sensor data to a central point, rather than optimizing the aggregation structures in combination with realizing distributed compression. The work of [21] is closest to ours and focuses on clustering of sensors to dedicated nodes, cluster-heads, where clusters are constructed based on proximity, i.e., a Voronoi tessellation of the sensing field. In this work the authors study how to optimize the size of the cluster so that energy consumption of the network might be minimized. In our work we propose a simple cost model for energy consumption and study the optimal hierarchical organization of sensors, aggregation points/compressors, and sinks, so as to minimize the cost of gathering sensor data.

The organization of this chapter can be summarized as follows. In Section 2.2 we formally study a model for optimal distributed compression subject to aggregation, e.g., energy, costs to a single query point or sink. We show that this problem has a simple greedy solution based on assigning minimal bit rates to sensors with high aggregation costs. In Section 2.3 and 2.4 we model and analyze optimal hierarchical structures for a large scale sensor network including multiple sinks, compression/aggregation devices and a large number of sensors. Based on a simplified model capturing the salient features of the problem we determine the optimal hierarchical organization, i.e., that which minimizes the overall energy consumption of the network. Leveraging previous work on stochastic modelling of spatial processes we are able to estimate and compare the aggregation costs associated with various organizational structures. These provide insights on where these structures are likely to be effective. In Section 2.5 we provide numerical and simulation results for the energy savings obtained by using the optimal hierarchical scheme and discuss some design issues associated with the large-scale sensor networks. We conclude this chapter in Section 2.6.
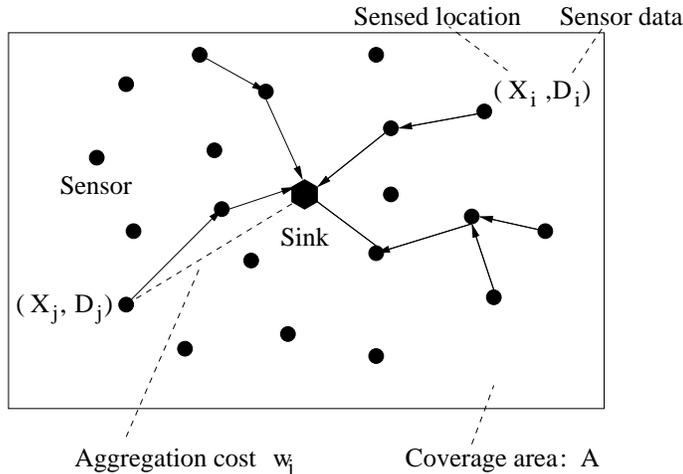
Figure 2.1: An illustration of the sensor reporting.

2.2 Optimal distributed data compression with aggregation costs

Let us first examine the problem of optimal distributed data compression with aggregation costs. As illustrated in Fig. 2.1 we consider a set of sensors $U = \{1, 2, \cdots, n\}$ at some locations $\vec{x} = (x_i, i \in U)$ within a coverage area $A \subset \mathbb{R}^2$. We model information obtained by these sensors as a random vector $\vec{D} = (D_i, i \in U)$, and suppose there is a dedicated node, a sink or query point, which coordinates and aggregates the information from the sensors. Since sensors are located at different positions, they may incur different aggregation costs in forwarding their data to the sink. We use a vector $\vec{w} = (w_i, i \in U)$ to model the costs, where w_i is the cost of moving a unit of data from sensor i to the sink. Thus for example if aggregation costs are modelled by $w_i = |x_i|^\gamma$ then for $\gamma = 1$ the cost for sensor i is proportional to its distance (a rough estimate for the number of hops) to the sink, or if the sensor communicates directly with the sink then for $\gamma \in (2, 4)$ this cost might capture the increased transmit power levels required to overcome path loss on a wireless link.

The information collected by the sensors is likely to be correlated and thus it is possible to jointly compress the data they generate. Let $\vec{r} = (r_i, i \in U)$ denote the number of bits per reading each sensor would generate. Then by Slepian-Wolf's Theorem the sum of the rates for any subset of sensors $S \subset U$ is lower bounded by conditional entropy, $H(D_S | D_{S^c})$, where $D_S = (D_i, i \in S)$ and $D_{S^c} = (D_i, i \in U \setminus S)$ [9]. Our objective is to jointly compress the sensed data while minimizing the overall

6

aggregation cost. We can formally state the problem as follows.

Problem 1. *For a set of devices U sensing an information vector $\vec{D}$, and an associated aggregation cost vector $\vec{w}$ determine the rate vector $\vec{r}^*$ that minimizes the overall aggregation cost subject to joint data compression constraints, i.e.,*

$$\min_{\vec{r}} \left\{ \sum_{i=1}^{n} w_i \cdot r_i \mid \sum_{i \in S} r_i \geq H(D_S | D_{S^c}), \quad \forall S \subseteq U \right\}.$$

Note that the feasible rate region is defined by $2^n - 1$ inequality constraints. Thus a direct application of linear programming is not likely to produce a polynomial time algorithm. Let us however consider a special case which provides some insight. Suppose we have two data sources and $w_1 \leq w_2$. In this case the rate constraints in Problem 1 correspond to

$$r_1 \geq H(D_1 | D_2), \ r_2 \geq H(D_2 | D_1), \ r_1 + r_2 \geq H(D_1, D_2),$$

which in turn translates to two possible solutions:

$$r_1 = H(D_1), \quad r_2 = H(D_2 | D_1),$$

or

$$r_1 = H(D_1 | D_2), \quad r_2 = H(D_1).$$

We can compare the aggregation costs for these two solutions by computing

$$
\begin{aligned}
& [w_1 H(D_1) + w_2 H(D_2 | D_1)] - [w_1 H(D_1 | D_2) + w_2 H(D_2)] \\
= \ & w_1 [H(D_1) - H(D_1 | D_2)] - w_2 [H(D_2) - H(D_2 | D_1)] \\
= \ & w_1 I(D_1; D_2) - w_2 I(D_2; D_1) \\
= \ & (w_1 - w_2) I(D_1; D_2) \leq 0,
\end{aligned}
$$

where $I(D_1; D_2) = I(D_2; D_1) \geq 0$ denotes the mutual information between D_1 and D_2. Thus it is optimum for the sensor which is nearby, i.e., having the smaller weight, to send at a rate $H(D_1)$ while the sensor which is further away sends at a compressed rate $H(D_2 | D_1)$ assuming the side information will be taken care of. This intuitive result is generalized in the following theorem.

7

Theorem 1. Suppose $w_1 \geq w_2 \geq \cdots \geq w_n$, then the optimal solution to Problem 1 is given by

$$r_1^* = H(D_1|D_2, D_3, \cdots, D_n), \ r_2^* = H(D_2|D_3, D_4, \cdots, D_n), \ \cdots, r_n^* = H(D_n).$$

To prove Theorem 1 we will show that the feasible rate region is a contra-polymatroid, and thus the optimal rate vector exhibits a specific structure. A contra-polymatroid is a polyhedron defined as follows:

$$\mathcal{B}(f) = \Big\{ \vec{x} \mid \vec{x} \in \mathbb{R}_+^n, \sum_{i \in S} x_i \geq f(S), \forall S \subseteq U \Big\},$$

where $f : 2^U \to \mathbb{R}_+$ is called the rank function satisfying

(1) $f(\emptyset) = 0$;

(2) $f(S) \leq f(T) \ \forall S \subset T$;   (monotonicity)

(3) $f(S) + f(T) \leq f(S \cup T) + f(S \cap T)$.   (super-modularity)

For such constraint sets the following result has been proved, see e.g., [22].

Lemma 1. A solution $\vec{r}^*$ to the following optimization problem where $\vec{w} \in \mathbb{R}_+^n$,

$$\min_{\vec{r}} \Big\{ \sum_{i=1}^n w_i \cdot r_i \mid \vec{r} \in \mathcal{B}(f) \Big\},$$

is given by, for $i = 2, 3, \ldots, n$,

$$
\begin{aligned}
r_{\pi(1)}^* &= f(\{\pi(1)\}) \\
r_{\pi(i)}^* &= f(\{\pi(1), \pi(2), \cdots, \pi(i)\}) - f(\{\pi(1), \pi(2) \cdots \pi(i-1)\})
\end{aligned}
$$

where π denotes a permutation of $(1, 2, \cdots, n)$ such that $w_{\pi(1)} \geq w_{\pi(2)} \geq \cdots \geq w_{\pi(n)}$.

The key step to proving Theorem 1 is showing that the conditional entropy function $f(S) = H(D_S|D_{S^c})$ is a rank function so the constraints in Problem 1 define a contra-polymatroid. The result then follows immediately by Lemma 1. The details of the proof are deferred to the appendix.

Theorem 1 implies that a data compression strategy that minimizes the aggregation cost can be obtained by a simple ordering of the sources based on their aggregation cost, rather than the actual values of these costs. Perhaps surprisingly, the optimal solution is independent of the correlation structure of the data sources. This greedy strategy turns out to be optimal due to the particular characteristics of the feasible rate region, i.e., a contra-polymatroid, and the linear objective function being used to model the overall aggregation costs.

We note that our result is based on applying Slepian-Wolf's Theorem for sensor sources sampling a spatially correlated field. In addition to spatial correlations one would expect temporal correlations in such data. Thus using the fact that Slepian-Wolf's result generalizes to ergodic data sources [23], Theorem 1 is easily extended to data sources which are spatio-temporally correlated, by replacing conditional entropies with conditional entropy *rates*.

Observe however, that if one increases the coverage area while keeping a fixed homogeneous spatial density of sensors sampling a stationary spatial process the amount of compression that can be realized by such a scheme will be limited. Indeed, when this is the case, an increase in the coverage area will result in increases in the number of sensors n and the overall compressed data forwarded to the sink would eventually grow linearly i.e., $H(D_1, D_2, \cdots D_n) \approx nH(\mathcal{D})$ where $H(\mathcal{D})$ denotes the entropy rate of the vector of sensed data. Thus eventually such an arrangement could achieve at best a compression ratio of $\alpha = H(\mathcal{D})/H(D_1)$. If the sink has a limited capacity, or in the case of a sensor network based on ad hoc wireless communication infrastructure one would expect the sink or ad hoc nodes in its neighborhood to eventually become hot-spots and run out of energy. Hence the proposed optimal compression strategy is not likely to scale for a large networks of sensors.

Let us briefly evaluate the performance gains that might be achieved under optimal compression. Suppose we deploy 10 sensors at a set of randomly selected locations $\vec{x}$ in a square area $A = [-1, 1]^2$, with a sink located at the origin. We shall model the sensed data as a stationary Gaussian field with zero mean and a spatial covariance $R(x_i, x_j) = 10e^{-\theta|x_i-x_j|}$ where θ models the rate of decay in the correlation between sensed data D_i and D_j, as a function of the distance between sensor locations x_i and x_j. This implies a correlation structure for the sensor data $\vec{D}$ that is dependent on the locations of the sensors. To estimate the rates generated by sensors we suppose they are equipped with a 4-bit A/D converter, and thus

9
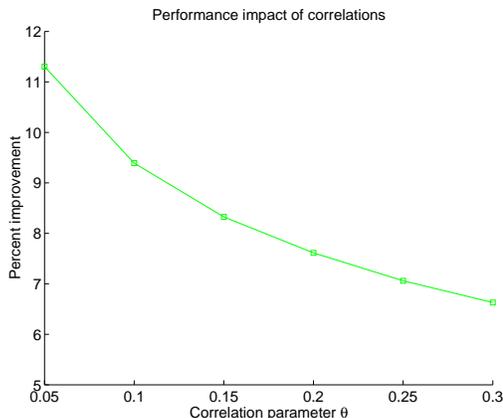
Figure 2.2: Percentage improvement in energy saving as a function of the correlation among data sources.
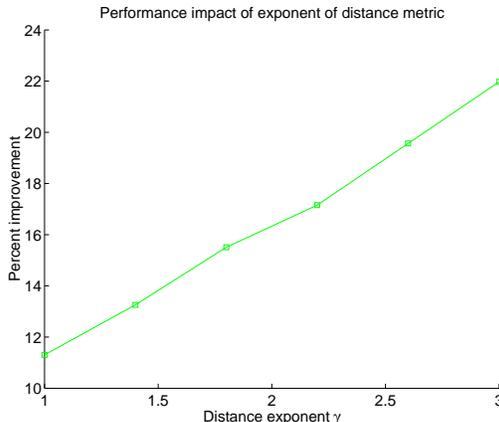
Figure 2.3: Percentage improvement in energy saving as a function of the exponent of distance metric, $\theta = 0.1$

approximately $H(D_i|D_{i+1}, D_{i+2}, \cdots D_n) = 4 + h(D_i|D_{i+1}, D_{i+2}, \cdots D_n)$ where $h(\cdot|\cdot)$ is the conditional differential entropy of the Gaussian source.

We will compare the *percentage reduction* in the aggregation cost for our optimal scheme versus a baseline which achieves optimal compression but allocates bits based on a random ordering of sensors. Fig. 2.2 shows the case where the weight for sensor i depends on its distance to the sink $|x_i|$. As expected the benefits decrease as correlation across sensed data decreases. This is intuitive since eventually with no correlation there will be no benefit for optimal aggregation since sensed data is independent. Fig. 2.3 shows the performance improvement when the weights are given by $|x_i|^{\gamma}$ and $\gamma \in [1, 3]$. Again as expected the larger exponent γ and thus the larger differences in the aggregation costs of sensors the higher the performance gains.

2.3 Optimal hierarchical structures for compression and aggregation sensor data

As argued in the previous section the optimal compression and aggregation scheme is not likely to scale for large sensor networks. As such it is reasonable to view a large scale sensor network in terms of a three level hierarchy. At the top we will suppose there are multiple sinks, e.g. wired nodes. At an intermediate level,

one can consider a class of sensor nodes that serve as traffic aggregators and/or in-network compressors. These nodes will gather and possibly further compress sensed data from a set of sensors, and then forward it to a nearby sink. Our objective in this section is to study the organizational structures that minimize the energy cost across a network of sensors. Given the complexity of such systems, i.e., number of variables and factors, we will follow the methodology proposed in [24]. The idea is to use crude stochastic geometric models to capture the salient features of such systems.

2.3.1   Hierarchical organization and energy models

We shall assume the sensor, compressor and sink locations are determined by homogeneous Poisson point processes Π_0, Π_1 and Π_2, with intensities λ_0, λ_1 and λ_2 respectively. We also will assume that a sink may also serve the role of a compressor, i.e., it may gather data directly from sensors if necessary. We use the location of a node as an index for the device itself. Thus $x_i \in \Pi_0$ denotes a particular sensor at location x_i, $c_j \in \Pi_1$ a compressor at location c_j and $s_k \in \Pi_2$ a sink at location s_k. Our model for a hierarchical organization of these nodes is based on two mappings. Let $c : \Pi_0 \to \Pi_1 \cup \Pi_2$ be a mapping associating a sensor x_i with a compressor or sink $c(x_i)$ and $s : \Pi_1 \cup \Pi_2 \to \Pi_2$ be a mapping associating a compressor or sink y_i with a sink $s(y_i)$. If y_i is a sink it is associated with itself. In the sequel we will consider various organizational structures, i.e., various choices for the functions c and s. In addition, we shall investigate two cases. First we investigate the multi-hop case, in which all the communications among sensors, compressors and sinks are realized on a hop-by-hop basis. Hence if a node wishes to send a packet to another node, the intermediate nodes along the route relay the packet in a hop-by-hop manner to the destination. Second, we explore the two-hop case, in which the sensors do not relay traffic, but instead transmit directly to either a compressor or a sink and in turn compressors transmit directly to a sink. Thus data may be relayed through at most one hop. The cost functions for these two cases will be quite different, but subsequently we combine these cases to express a general model for aggregation.

2.3.1.1   Multi-hop case

Our goal is to roughly model the energy cost associated with aggregating data from sensors to compressors and subsequently to the sink. We will assume that

sensors and compressors communicate over an ad hoc network of wireless links, so packets will be relayed over multiple hops toward their destination. Thus the overall aggregation cost in the system will be proportional to the number of packets that need to be relayed. The energy cost per packet will depend on the overheads to access the channel for transmission, and energy expended during transmission. The latter would in turn depend on the size of the packet and the distances involved. For purposes of our model it is reasonable to assume packets have roughly the same size, and the energy cost per packet relayed in the network is roughly the same. We will further assume that the number of hops required to travel between two locations is proportional to distance between them.[1] Suppose then that sensor x_i generates d_{x_i} packets/sec to its compressor $c_j = c(x_i)$. The total traffic received by c_j would be

$$\sum_{\{x_i : c(x_i) = c_j\}} d_{x_i} \text{ packets/sec,}$$

and the total energy expended in the network to carry this traffic is roughly proportional to

$$\sum_{\{x_i : c(x_i) = c_j\}} d_{x_i} |x_i - c_j| \text{ packets-hops/sec.}$$

Recall that each compressor c_j aggregates incoming data and possibly further compresses it prior to forwarding it to the sink $s(c_j)$. Aggregation may make use of context dependent information, e.g., since the sensors associated with that compressor are known a priori, one need not forward meta-data, e.g., location information, about the data to the sink. Thus the compressor can reduce packet and meta-data overheads, e.g. by header compression, in forwarding information to the sink. Further the sensed data itself may be compressed. Specifically we assume that c_j forwards packets at a rate

$$\alpha_{c_j} \sum_{\{x_i : c(x_i) = c_j\}} d_{x_i} \text{ packets/sec}$$

to its associated sink $s_k = s(c_j)$, incurring an additional energy cost which is proportional to

$$\alpha_{c_j} \sum_{\{x_i : c(x_i) = c_j\}} d_{x_i} |c_j - s_k| \text{ packets-hops/sec}$$

―――――――――――――――――――

[1]This assumes that the spatial distribution of sensor nodes is fairly uniform.

The parameter $\alpha_{c_j} \in [0,1]$ captures the gross benefit of aggregation and compression resulting from c_j's processing and forwarding of sensor data.

Since our purpose is to capture the salient features of such a hierarchy, it makes sense to assume that all sensor nodes offer roughly the same traffic load, i.e., without loss of generality we let $d_{x_i} = 1$ for all $x_i \in \Pi_0$. Note the scheme considered in the previous section can be used to locally minimize the data rates from sensors to compressors. In addition, assuming that the set of sensors associated with a given compressor node is sufficiently large, and variations in α_{c_j} will not be significant, we might assume all compressor/aggregating nodes are equally effective, i.e., $\alpha_{c_j} = \alpha$ for all compressor nodes $c_j \in \Pi_1$. With these two assumptions we obtain an additive energy cost model, with a cost per sensor, $e(x_i)$, which is proportional to

$$e(x_i) = |x_i - c(x_i)| + \alpha|c(x_i) - s(c(x_i))|. \tag{2.1}$$

To motivate our choice of a near-constant compression ratio, let us consider a simple meta-data compression scheme. Suppose all the packets generated by the sensors in the field have a fixed header length of L_h, and a fixed data length of L_d. For a certain compressor (or aggregation point) that takes care of N sensors, the total amount of data aggregated at the compressor would be $N(L_h + L_d)$. Now suppose a compressor removes all the headers and attaches a new header for the aggregated data. Indeed as discussed in [25], the compressor need not include the ordering information for the aggregated data if it is already known to both the compressor and its destination, i.e., the associated sink. Thus we may assume the compressor can send a packet with aggregated data with a header of the same length as that for sensor packets, L_h. The size of the packet routed to the sink is $L_h + NL_d$, assuming the compressor has not applied any additional compression technique. If we assume that $NL_d \gg L_h$, i.e., the compressor's packet contains much more data than header information, then we achieve a compression ratio of

$$\alpha = \frac{NL_d + L_h}{N(L_d + L_h)} \simeq \frac{NL_d}{N(L_d + L_h)} = \frac{L_d}{L_d + L_h},$$

which should be roughly constant over the field. Our objective is, given our three point processes and a fixed gross compression ratio α, to determine the hierarchical organization, i.e. the functions $c(\cdot)$ and $s(\cdot)$, that minimize the overall energy cost in this network.

2.3.1.2 Direct one hop transmission case

In our second model each sensor directly communicates with a sink or a compressor without any relaying. Thus the resultant transmissions would involve two hops at most. We keep a similar setup as that previously considered, thus the sensors have two options: (1) send directly to a compressor which compresses and forwards it to a sink, or (2) send directly to a sink. We assume sensors make the minimum-cost choice among the candidate compressors and sinks. In this context, the cost per sensor might be modelled as

$$e(x_i) = |x_i - c(x_i)|^\gamma + \alpha|c(x_i) - s(c(x_i))|^\gamma \tag{2.2}$$

where $2 \leq \gamma \leq 4$. Here γ reflects the power required to overcome when a sensor transmits directly to a compressor or a sink. In many environments, path-loss exponent is assumed to be between -2 and -4[26]. This means the contribution to the total energy cost between a sensor and a compressor/sink would be proportional to their distance raised to a power between 2 to 4. By contrast, in our previous setup $\gamma = 1$, we have assumed every node spends approximately equal power for transmission. Hence the energy cost is proportional to the number of hops required for sending packets. In the regime when $\gamma \geq 2$, we assume that every node adopts an appropriate power control scheme to compensate for path loss. Hence compressors and sinks should be spatially distributed within a reasonable distances from each sensor, so that a sensor can reach its aggregation point directly.

We can regard (2.2) as the general expression for the cost per sensor since if we let $\gamma = 1$, (2.2) reduces to the multi-hop case. In the following discussions we will maintain the arguments as general as possible assuming γ can be any reasonable positive constant, but we focus on the case where $\gamma = 1$, the multi-hop case. Later we extend our arguments to cover the case where $\gamma \geq 2$.

2.3.2 Optimal sensor hierarchies and spatial tessellations

Prior to considering the optimal organization of sensors and compressors based on the above energy model, let us consider a natural choice. It should be clear that from the perspective of minimizing our cost metric, each compressor should relay packets to the closest sink. Similarly one might think it reasonable to assume each sensor might send its traffic to the closest compressor or, if it is closer,

14

directly to a sink. As discussed in [24] such a hierarchical organization is induced by the *Voronoi tessellation* generated by the locations of the sinks and compressors. We formally define the Voronoi tessellation with respect to a set of points Π as follows.

Definition 1. *The Voronoi tessellation associated with Π, denoted $\mathcal{V}(\Pi)$ is a collection of cells $V_{y_i}(\Pi)$ for $y_i \in \Pi$ such that*

$$V_{y_i}(\Pi) = \{z \in \mathbb{R}^2 \mid |y_i - z| < |y_j - z|, \forall y_j \in \Pi\},$$

i.e., all points in the plane which are closer to y_i than to any other point in Π.

Let $\mathcal{V}(\Pi_1 \cup \Pi_2)$ denote the Voronoi tessellation induced by the union of compressors and sinks $\Pi_1 \cup \Pi_2$ and let us define a hierarchical organization as follows. Let $c_V : \Pi_0 \to \Pi_1 \cup \Pi_2$ be such that for each $x_i \in \Pi_0$,

$$c_V(x_i) = y_j, \text{ iff } x_i \in \overline{V}_{y_j}(\Pi_1 \cup \Pi_2).$$

Here $\overline{V}_{y_j}$ denotes the closure of the set V_{y_i}. Mathematically the probability that a sensor will fall on the boundary of sets V_{y_i} vanishes. However, if a sensor is observed to be located on the boundary due to the finite granularity in the positions of nodes, such ties must be broken arbitrarily. Let $\mathcal{V}(\Pi_2)$ denote the Voronoi tessellation induced by the sinks Π_2, and define a mapping $s_V : \Pi_1 \cup \Pi_2 \to \Pi_2$, such that for each $y_j \in \Pi_1 \cup \Pi_2$,

$$s_V(y_j) = s_k, \text{ iff } y_j \in \overline{V}_{s_k}(\Pi_2).$$

The pair c_V and s_V capture a simple hierarchical organization for sensors, compressors and sinks based purely on proximity rather than achieving our goal, i.e., minimizing the energy cost of aggregation. In particular a sensor may connect to the closest compressor even if the energy cost would be reduced if it connected to another compressor which is closer to its eventual sink and thus results in an energy savings. Below we define the tessellation $\mathcal{T}^\alpha(\Pi_1, \Pi_2)$ that induces an optimal organization for a given overall compression ratio α.

Definition 2. *The tessellation $\mathcal{T}^\alpha(\Pi_1, \Pi_2)$ associated with two sets of points Π_1 and Π_2 and parameter α, consists of a collection of cells $T^\alpha_{y_i}(\Pi_1, \Pi_2)$ for $y_i \in \Pi_1 \cup \Pi_2$ such that*

$$T^\alpha_{y_j}(\Pi_1, \Pi_2) = \{z \in \mathbb{R}^2 \mid |z - y_j| + \alpha |y_j - s_V(y_j)| < |z - y_l| + \alpha |y_l - s_V(y_l)|, \forall y_l \in \Pi_1 \cup \Pi_2\}.$$

15

We refer to $T_{y_j}^{\alpha}(\Pi_1, \Pi_2)$ as the cell with nucleus y_i.

The next theorem shows that this tessellation induces an organization which minimizes the energy cost for each sensor.

Theorem 2. *Under the energy cost model (2.1), the optimal assignment of compressors $c_j \in \Pi_1$ to sinks is $s^*(c_j) = s_V(c_j) \in \Pi_2$, i.e., the closest sink to c_j. The optimal assignment of sensors $x_i \in \Pi_0$ to compressors or sinks is $c^*(x_i)$ such that $x_i \in T_{c^*(x_i)}^{\alpha}(\Pi_1, \Pi_2) \in \mathcal{T}^{\alpha}(\Pi_1, \Pi_2)$, i.e., the compressor/sink which is the nucleus for the cell in $\mathcal{T}^{\alpha}(\Pi_1, \Pi_2)$ containing x_i.*

Proof. To prove that this hierarchy has minimal cost consider a sensor $x_i \in T_{y_j}^{\alpha}(\Pi_1, \Pi_2)$, so $c^*(x_i) = y_j$. Suppose x_i is assigned to some other compressor or sink y_l and then some sink s_k. The energy cost for such an assignment would be

$$
\begin{aligned}
e(x_i) & = |x_i - y_l| + \alpha |y_l - s_k| \\
& \geq |x_i - y_l| + \alpha |y_l - s^*(y_l)| \\
& \geq |x_i - y_j| + \alpha |y_j - s^*(y_j)| \\
& = |x_i - c^*(x_i)| + \alpha |c^*(x_i) - s^*(c^*(x_i))|,
\end{aligned}
$$

where the first inequality follows from the definition of s^* and the second inequality is a natural consequence of the definition of the cells in the tessellation $\mathcal{T}^{\alpha}(\Pi_1, \Pi_2)$. $\quad\square$

The tessellation $\mathcal{T}^{\alpha}(\Pi_1, \Pi_2)$ introduced earlier is a particular case of a Johnson-Mehl tessellation [27]. Let us briefly consider how this comes about by contrasting it with a Voronoi tessellation. One can visualize the Voronoi tessellation as induced by a set of nuclei which grow grains isotropically at unit speed until they come into contact with another nuclei's grain. These contact points become the boundaries of the cells of the tessellation. A Johnson-Mehl tessellation adds an additional degree of freedom. It allows the nuclei to start growing their grains at different times. The tessellation $\mathcal{T}^{\alpha}(\Pi_1, \Pi_2)$ results if all sinks $s_k \in \Pi_2$ start growing grains together isotropically at unit speed. Subsequently each compressor $c_j \in \Pi_1$ starts growing a grain at time proportional to $\alpha |c_j - s^*(c_j)|$ which depends on its distance from the closest sink and α. The resulting boundaries can be shown to be hyperbolic. In this

16

tessellation cells associated with sinks will tend to be large relative to others, since they start growing earlier. We give a formal proof of this argument.

Theorem 3. *Suppose we have two point processes Π_1, $\Pi_2 \subset \mathbb{R}^2$, which correspond to nuclei for the cells of a tessellation. Suppose the nuclei grows isotropically with the identical velocity v, but a nucleus $c_i \in \Pi_1 \cup \Pi_2$ starts to grow at time $\tau(c_i)$ given by*

$$\tau(c_i) = \frac{\alpha |c_i - s_V(c_i)|}{v}$$

where the function $s_V(\cdot) : \Pi_1 \cup \Pi_2 \to \Pi_2$ maps a point in $\Pi_1 \cup \Pi_2$ to the closest point in Π_2 and if $c_i \in \Pi_2$, then $s_V(c_i) = c_i$. Hence all the nuclei in Π_2 starts to grow at time 0. Also, the nuclei keep growing to cover the entire space not been occupied by other nuclei. Then the resulting tessellation induced by Π_1, Π_2 is the same as the optimal tessellation $\mathfrak{T}^\alpha(\Pi_1, \Pi_2)$ given in Definition 2.

Proof. Let us define $T(x, c_i)$ as the time when the growth from $c_i \in \Pi_1 \cup \Pi_2$ reaches some arbitrary point $x \in \mathbb{R}^2$, ignoring all other nuclei for the moment. Then $T(x, c_i)$ can be explicitly written as

$$T(x, c_i) = \frac{|x - c_i|}{v} + \tau(c_i) = \frac{|x - c_i|}{v} + \frac{\alpha |c_i - s_V(c_i)|}{v}.$$

In order for x to be covered by the growth of c_i, x must be reached at the shortest time, i.e., before any nucleus other than c_i occupies x. Specifically,

$$c_i = \operatorname*{argmin}_{c_l \in \Pi_1 \cup \Pi_2} T(x, c_l) = \operatorname*{argmin}_{c_l \in \Pi_1 \cup \Pi_2} \left\{ \frac{|x - c_l|}{v} + \frac{\alpha |c_l - s_V(c_l)|}{v} \right\}.$$

Since v is a constant, the collection of x that satisfies the above constraint, i.e. the region that is occupied by c_i, is the same as $T_{c_i}^\alpha(\Pi_1, \Pi_2)$ given in Definition 2. This is true for every $x \in \mathbb{R}^2$, thus we conclude that the resulting tessellation has the same shape as $\mathfrak{T}^\alpha(\Pi_1, \Pi_2)$. $\qquad\square$

Examples of a Voronoi tessellation and an energy optimal Johnson-Mehl tessellation for $\alpha = 0.5$ are shown in Figure 2.4. The figure on the left shows the optimal Johnson-Mehl tessellation $\mathfrak{T}^\alpha(\Pi_1, \Pi_2)$ for compression ratio $\alpha = 0.5$, and the figure on the right shows the Voronoi tessellation $\mathcal{V}(\Pi_1 \cup \Pi_2)$. The small dots represent compressors in Π_1 and big dots represent sinks in Π_2. We have superposed the

Voronoi tessellation $\mathcal{V}(\Pi_2)$ induced by the sinks and drawn the boundaries with thick straight lines. The organization of sensors to compressors or sinks, and compressors to sinks should be clear based on its location with respect to the two superposed tessellations.

Considering the case of direct transmission to the aggregation points, we can define the optimal tessellation in a similar way. Given a fixed path-loss exponent γ and compression ratio α, the optimal tessellation is given as follows.

Definition 3. *The tessellation* $\mathcal{T}^{\alpha,\gamma}(\Pi_1, \Pi_2)$ *associated with two sets of points* Π_1 *and* Π_2 *and parameter* $\alpha \in [0,1]$ *and* $\gamma \geq 1$, *consists of a collection of cells* $T_{y_i}^{\alpha,\gamma}(\Pi_1, \Pi_2)$ *for* $y_i \in \Pi_1 \cup \Pi_2$ *such that*

$$T_{y_j}^{\alpha,\gamma}(\Pi_1, \Pi_2) = \{z \in \mathbb{R}^2 \mid |z - y_j|^\gamma + \alpha|y_j - s_V(y_j)|^\gamma < |z - y_l|^\gamma + \alpha|y_l - s_V(y_l)|^\gamma,$$
$$\forall y_l \in \Pi_1 \cup \Pi_2\}.$$

Theorem 4. *Under the energy cost model (2.2), the optimal assignment of compressors* $c_j \in \Pi_1$ *to sinks is* $s^*(c_j) = s_V(c_j) \in \Pi_2$, *i.e., the closest sink to* c_j. *The optimal assignment of sensors to compressors or sinks is* $c^*(x_i)$ *such that* $x_i \in T_{c^*(x_i)}^{\alpha,\gamma}(\Pi_1 \cup \Pi_2) \in \mathcal{T}^{\alpha,\gamma}(\Pi_1, \Pi_2)$, *i.e., the compressor/sink which is the nucleus for the cell in* $\mathcal{T}^{\alpha,\gamma}(\Pi_1, \Pi_2)$ *containing* x_i.

Proof. Proof similar to Theorem 2. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

One of the properties of the optimal tessellation when $\gamma \geq 2$, is that there may be some cells *which do not contain their nucleus*. Compared with a Voronoi tessellation of $\Pi_1 \cup \Pi_2$, the cells tend to be stretched and shifted away from each sink, so compressors that are far away from their associated sinks usually are outside of their own cells. Also note that in the optimal tessellation for $\gamma \geq 2$, some compressors have zero-area cells. In other words, from the perspective of our energy model, they would not be energy efficient to route through. These bogus compressors appear around the boundaries of the Voronoi cells of Π_2. The reason is that the compressors that are around the boundary of Voronoi cells of Π_2 tend to be preempted by other compressors nearer to the sinks. If we consider the germ-grain model, they look as if 'sandwiched' by cells expanding faster in opposite directions

18
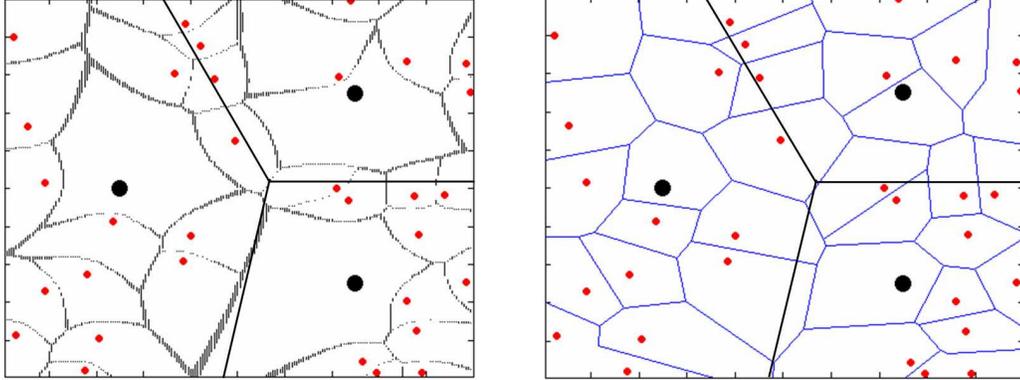
Figure 2.4: The optimal Johnson-Mehl tessellation for $\alpha = 0.5$ and Voronoi tessellation.

from their respective sinks. Thus if we remove these useless compressors, we can achieve the same cost performance with less compressors. Detailed analysis on these properties are left for future study.

We continue with some analytical results for the energy cost and statistics of the Voronoi and Johnson-Mehl tessellations of multiple-hop case, i.e. when $\gamma = 1$.

2.4 Analytical Results

In this section we will compare various sensor,compressor and sink hierarchies in terms of their expected energy costs for aggregation. Our approach draws on the elegant framework introduced by [24], for a rough analysis and modelling of communication hierarchies.

2.4.1 Hierarchy based on Voronoi tessellation

Let us first consider the hierarchy based on Voronoi tessellation whereby each sensor sends data to its closest compressor or sink node, and compressor nodes relay a fraction α of the traffic they receive to the closest sink. Following [24] we can mathematically define an expected energy cost G_V under such an arrangement

as

$$G_V = \mathbf{E}_2^0 \left[\sum_{c_j \in \Pi_1 \cap V_0(\Pi_2)} \{ \alpha |c_j| \mathcal{N}_{c_j} + \sum_{x_i \in \Pi_0 \cap V_{c_j}(\Pi_1 \cup \Pi_2)} |x_i - c_i| \} + \sum_{x_i \in \Pi_0 \cap V_0(\Pi_1 \cup \Pi_2)} |x_i| \right],$$

where we have used the following conventions: $\mathbf{E}_k^0$ denotes the expectation with respect to the Palm distribution $\mathbf{P}_k^0$ of the point process Π_k, see [28]; and $\mathcal{N}_{c_j}$ is the number of sensors, i.e., points in Π_0 which are associated with compressor c_j, that is in the cell $V_{c_j}(\Pi_1)$.

Thus G_V corresponds to the energy cost associated with a typical sink cell $V_0(\Pi_2)$. The first summand consists of two terms: the first corresponds to the energy costs in forwarding compressed data from compressors to sinks, while the second corresponds to the energy cost of forwarding data from sensors to compressors. The second summand in the above expression corresponds to energy costs of sensors that directly send data to the sink. One can directly use the results in [24] to evaluate the expected cost G_V. In our case, G_V is given by

$$
\begin{aligned}
G_V &= \frac{\lambda_0}{2\lambda_2\sqrt{\lambda_1 + \lambda_2}} + \frac{\alpha\lambda_0}{2\lambda_2^{3/2}} - \frac{\alpha\lambda_0}{2(\frac{\lambda_1}{4} + \lambda_2)^{3/2}} \\
&+ 12\pi^{3/2}\lambda_0\lambda_1\alpha \int_0^{\frac{\pi}{2}} \frac{(\pi - \gamma)\sin\gamma\cos^2\gamma\, d\gamma}{(\pi(\gamma_1 + 2\gamma_2) + \gamma_2\sin(2\gamma) + 2\gamma_2(\pi - \gamma)\cos(2\gamma))^{5/2}}.
\end{aligned}
$$

We note for simplicity that the same results can be applied to evaluate the cost G_S of an aggregation hierarchy with no compressor nodes at all, i.e., sensors merely send uncompressed data to the closest sink. Applying the results of [24] one finds the energy cost for a typical sink cell would be given by

$$G_S = \frac{\lambda_0}{\lambda_2} \cdot \frac{1}{2\sqrt{\lambda_2}}.$$

The basic insight here is that the area of a typical sink cell would be $1/\lambda_2$ and so the average number of sensors in that cell is expected to be λ_0/λ_2. Further one can show that $1/2\sqrt{\lambda_2}$ corresponds to the mean distance from a sensor to its closest sink. Thus the average cost should be the product of these two terms.

As we will show in our numerical evaluation and simulations of the energy costs for these hierarchies, it may be the case that $G_V > G_S$ for some α. Specifically when α is close 1 it will typically be more expensive to route through the closest compressor or sink versus going directly to the sink. Thus a greedy aggregation strategy based purely on proximity may be worse than not using one at all.

2.4.2 Optimal hierarchy based on Johnson-Mehl tessellation

Next we evaluate the energy cost G_T for the optimal hierarchy introduced in the previous section. We can mathematically define it as

$$G_T = \mathbf{E}_2^0 \left[\sum_{c_j \in \Pi_1 \cap V_0(\Pi_2)} \{\alpha |c_j| \mathcal{N}_{c_j}^\alpha + \sum_{x_i \in \Pi_0 \cap T_{c_j}^\alpha(\Pi_1,\Pi_2)} |x_i - c_j|\} + \sum_{x_i \in \Pi_0 \cap T_0^\alpha(\Pi_1,\Pi_2)} |x_i| \right],$$

where $\mathcal{N}_{c_j}^\alpha$ denotes the number of sensors, i.e., points in Π_0, which are associated with compressor c_j, that is are within the cell $T_{c_j}^\alpha(\Pi_1,\Pi_2)$. Again G_T corresponds to the energy cost associated with a typical sink cell in the higher level Voronoi tessellation associated with sink cells.

As an intermediate step toward estimating the energy cost let us first consider the area of a typical cell in $\mathcal{T}^\alpha(\Pi_1,\Pi_2)$ which is associated with a sink. Recall that cells associated with sinks will tend to be larger than others since they start growing earlier, see Fig. 2.4. The following theorem, proved in the appendix, and gives an explicit formula for the area of a typical sink.

Theorem 5. *Consider two point processes Π_1 (compressors) and Π_2 (sinks) with densities λ_1 and λ_2 respectively. The expected area of a typical cell associated with a sink in the tessellation $\mathcal{T}^\alpha(\Pi_1,\Pi_2)$ is given by*

$$\mathbf{E}_2^0[\ |T_0^\alpha(\Pi_1,\Pi_2)|\ ] = \frac{\pi}{f(\alpha)\lambda_1 + \pi\lambda_2}$$

where $|T|$ denotes the area of a set T and $f(\alpha)$ is defined as follows. We shall let $O_\beta^\alpha(x_1, x_2)$ denote the Cartesian oval with foci at $x_1, x_2 \in \mathbb{R}^2$, given by

$$O_\beta^\alpha(x_1, x_2) = \{y \in \mathbb{R}^2 \mid \alpha |y - x_1| + |y - x_2| \leq \beta\}.$$

Then $f(\alpha)$ is the area of $O_1^\alpha(0, x)$ when $|x| = 1$ which is given by

$$f(\alpha) = \frac{4}{(1-\alpha^2)^2} \left\{ (\frac{1}{2} + \alpha^2)(\frac{\pi}{2} - \sin^{-1}(\alpha)) + \frac{\sin(2\sin^{-1}(\alpha))}{4} - 2\alpha \cos(\sin^{-1}(\alpha)) \right\}.$$

$$(2.3)$$

Note that the Cartesian oval $O_\beta^\alpha(0, x)$ can be interpreted as the set of possible compressor locations with respect to a sink at the origin and a sensor at x that would
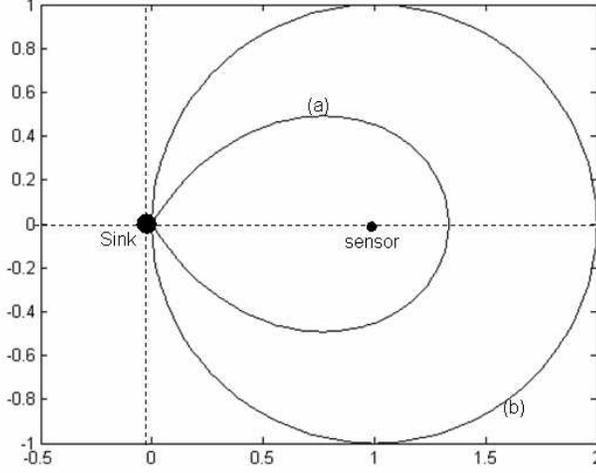
Figure 2.5: The figure shows the ball $B_{|x|}(x)$ and Cartesian oval $O_{|x|}^{\alpha}(0,x)$ associated with a potential (sensor) location at $x = (1,0)$ and sink at the origin.

result in a lower energy cost than β if used as a relay point, see e.g., Fig. 2.5. It visualizes the fact that, in order for the sensor at (1,0) to be associated with the sink at the origin, there must be no compressors in the Cartesian oval (a) and no other sinks in region in the ball (b). One can show that $f(\alpha)$ is a monotonically decreasing function in $\alpha \in [0,1]$ where $f(0) = \pi, f(1) = 0$. Thus when $\alpha = 1$, i.e., there is no compression gain, the area of a typical sink cell becomes $\frac{1}{\lambda_2}$. This is the average area of a typical cell for a Voronoi tessellation with density λ_2, thus connecting to the nearest sink must be optimal. At the other extreme if $\alpha = 0$, and thus the data which will be relayed from compressors to sinks is negligible, then the optimal sink cell's area becomes $\frac{1}{\lambda_1+\lambda_2}$ which is the average area of a typical cell of a Voronoi tessellation with density $\lambda_1 + \lambda_2$. Clearly connecting to the nearest sink or compressor is the optimal policy when $\alpha = 0$. Thus at the two extremes our Johnson-Mehl tessellation reduces to one of two Voronoi cases. However when $\alpha \in (0,1)$ we can expect to get cells with hyperbolic faces associated with an optimal hierarchy.

Next we shall estimate energy costs for the optimal hierarchy. Unfortunately since $\mathcal{T}^{\alpha}(\Pi_1, \Pi_2)$ depends in an inhomogeneous manner on the two point processes

this is quite challenging. As such below we derive an upper bound G_U for the energy cost G_T achieved by our optimal tessellation. To do so we consider yet another alternative hierarchical organization for the sensor network. In this hierarchy the data associated with a sensor x_i is necessarily routed to its closest sink $s_V(x_i)$. This data may however be routed either directly to the sink or via a compressor depending on which option minimizes the contribution to the energy cost of the system. Letting $c_U(x_i)$ denote the energy optimal routing choice for sensor x_i, i.e., to a sink or compressor, with an associated energy cost given by

$$e(x_i) = |x_i - c_U(x_i)| + \alpha |c_U(x_i) - s_V(x_i)|. \qquad (2.4)$$

Although this hierarchy is suboptimal it is quite close to the optimal one. In particular sensors which are not close to the boundary of the Voronoi cells associated with the point process of sinks, are likely to see the same energy costs – consider Fig. 2.4.

As before we can define the cost for this hierarchical arrangement as the energy cost for a typical sink cell

$$G_U = \mathbf{E}_2^0 \left[\sum_{x_j \in \Pi_0 \cap V_0(\Pi_2)} e_U(x_i, \Pi_1, \Pi_2) \right].$$

Under this new organization only sensors within $V_0(\Pi_2)$ will be associated with the origin, thus one can directly express the energy cost of a typical cell as a sum over sensors in this cell. However the cost per sensor depends on the locations of compressor and sink nodes Π_1 and Π_2 – we introduce e_U to make this explicit.

The following theorem, proved in the appendix, gives an explicit upper bound for the energy cost achieved by the optimal aggregation scheme.

Theorem 6. *The energy G_U is an upper bound for the optimal cost G_T and is given by*

$$
\begin{aligned}
G_U \;=\; & \frac{\lambda_0 \pi \Gamma(3/2)}{(\lambda_2 \pi + \lambda_1 f(\alpha))^{3/2}} \\
& + \lambda_0 \lambda_1 \int_{\mathbb{R}^2} \int_{O_{|x|}^\alpha(x,0)} (|y| + \alpha|y - x|) e^{-\lambda_2 \pi |x|^2 - \lambda_1 |O_{|y| + \alpha|y - x|}^\alpha(x,0)|} dy dx
\end{aligned}
$$

$$(2.5)$$

where $f(\alpha)$ and $O_\beta^\alpha(x,0)$ are as defined in Theorem 5.

23

We shall numerically evaluate this result in Section 2.5 to compare different hierarchical organizations for data aggregation. Also we show that G_U is a tight bound on the optimal cost via simulation.

The expression in Theorem 6 contains a numerical integration that cannot be evaluated in closed form. Specifically, it is not possible to evaluate the area of a Cartesian oval to a closed form except for some special cases. We propose a simpler estimate for G_U as follows. Consider the association rule as given in (2.4), i.e., the sensors always select their final destination for their data to be the *closest* sink. Denote the average cost incurred by a typical *sensor* under this rule, as G_{U0}. Then G_{U0} can be related to G_U via the Neveu exchange formula [24] as

$$G_{U0} = \frac{\lambda_2}{\lambda_0} G_U.$$

Thus G_{U0} is equal to G_U scaled by the ratio between the densities of sensors and sinks. Now we shall derive a tight upper bound for G_{U0}, which we denote $\hat{G}_{U0}$.

Theorem 7. *The energy cost per sensor G_{U0} is upper bounded by $\hat{G}_{U0}$ given by:*

$$\hat{G}_{U0} = \frac{\alpha}{2\sqrt{\lambda_2}} + \frac{1-\alpha}{2\sqrt{\lambda_2 + \lambda_{eff}}}, \tag{2.6}$$

where $\lambda_{eff} = \frac{\lambda_1 f(\alpha)}{\pi}$, and $f(\alpha)$ is as defined in Theorem 5.

We observe that $\hat{G}_{U0}$ is consistent with the optimal cost in two extreme cases, $\alpha = 0$('perfect' compression) and $\alpha = 1$(no compression at all). Indeed recall that $\frac{1}{2\sqrt{\lambda_2}}$ is the cost seen by a typical sensor in Voronoi tessellation of Π_2, and $\frac{1}{2\sqrt{\lambda_1+\lambda_2}}$ is the cost seen by a typical sensor in Voronoi tessellation of $\Pi_1 \cup \Pi_2$. When $\alpha = 0$, the optimal tessellation reduces to the Voronoi tessellation of $\Pi_1 \cup \Pi_2$, and $\hat{G}_{U0}$ also reduces to $\frac{1}{2\sqrt{\lambda_1+\lambda_2}}$, since $f(0) = \pi$. When $\alpha = 1$, $\hat{G}_{U0}$ reduces to $\frac{1}{2\sqrt{\lambda_2}}$, which is also the optimal cost for a typical sensor when there is no compression gain. Thus this bound clarifies the dependence of the energy cost on the compression ratio. We later show that $\hat{G}_{U0}$ gives an excellent approximation for G_{U0}. A proof of Theorem 7 is given in the appendix.

2.4.3   Extensions

2.4.3.1   General case with computational costs

The expression for the expectation of overall cost can be generalized to include a path loss exponent γ for transmission, in the two-hop aggregation case. Also one can introduce *computational costs associated with compression*, although this is often considered negligible compared to the cost of transmitting bits [29]. Let us denote the number of sensors that are associated with a compressor at location c_j by $\mathcal{N}_{c_j}^{\alpha,\gamma}$. The computational cost at a compressor can be a function of the number of the associated sensors, say $g(\mathcal{N}_{c_j}^{\alpha,\gamma})$. Then we can write the overall cost denoted by G_T^γ as follows:

$$
G_T^\gamma = \mathbf{E}_2^0 \left[\sum_{c_j \in \Pi_1 \cap V_0(\Pi_2)} \{ \alpha |c_j|^\gamma \mathcal{N}_{c_j}^{\alpha,\gamma} + \right.
$$

$$
\left. g(\mathcal{N}_{c_j}^{\alpha,\gamma}) + \sum_{x_i \in \Pi_0 \cap T_{c_j}^{\alpha,\gamma}(\Pi_1,\Pi_2)} |x_i - c_j|^\gamma \} + \sum_{x_i \in \Pi_0 \cap T_0^{\alpha,\gamma}(\Pi_1,\Pi_2)} |x_i|^\gamma \right].
$$

However, it is not easy to calculate high order moments of $\mathcal{N}_{c_j}^{\alpha,\gamma}$ although its moment generating function can be bounded, see [30]. If the computational overhead is proportional to the number of sensors and is not negligible, we can replace the first summand in the above by $(\alpha |c_j|^\gamma + \phi)\mathcal{N}_{c_j}^{\alpha,\gamma}$, where ϕ represents computational overhead incurred per sensor. Following the same approach as Theorem 6, we can obtain an upper bound on the cost in this general case. The upper bound G_U^γ is given by

$$
\begin{aligned}
G_U^\gamma =& 2\pi\lambda_0 \int_0^\infty r^{1+\gamma} e^{-\lambda_2 \pi r^2} e^{-\lambda_1 |O_r^{\alpha,\gamma}((r,0),0)|} dr \\
&+ \lambda_0 \lambda_1 \int_{\mathbb{R}^2} \int_{\Delta_x} d^{\alpha,\gamma}(x,y) e^{-\pi\lambda_2 |x|^2 - \lambda_1 |O_{d^{\alpha,\gamma}(x,y)}^{\alpha,\gamma}(x,0)|} dy dx, \\
d^{\alpha,\gamma}(x,y) =& |y|^\gamma + \alpha|y-x|^\gamma + \phi, O_\beta^{\alpha,\gamma}(f_1,f_2) = \{x \in \mathbb{R}^2 | d^{\alpha,\gamma}(f_1,f_2) \leq \beta\}, \\
\Delta_x =& \{y \in \mathbb{R}^2 \, |\alpha|x|^\gamma + \phi \leq d^{\alpha,\gamma}(x,y) \leq |x|^\gamma \}.
\end{aligned}
$$

2.4.3.2   A special case when the path-loss exponent $\gamma = 2$

Let us consider the case where the path-loss exponent γ is given by 2. The value of γ corresponds to free-space radiation in an environment with negligible

multi-path components. Let $c(x_i), s(c(x_i))$ be the positions of the cost-minimizing compressor and sink for the sensor at position x_i. We denote the cost function incurred by this sensor as $e(x_i)$ which is given by

$$e(x_i) = |x_i - c(x_i)|^2 + \alpha |c(x_i) - s(c(x_i))|^2.$$

In this case, the boundaries of the tessellation are *straight lines*. The cells have surprisingly similar shapes to those of a Voronoi tessellation. The reason can be deduced from the following example. Let us assume that there is a sink at the origin and a sensor at $z \in \mathbb{R}^2$. Without loss of generality, we assume that z is on the x axis, i.e. $z = (|z|, 0)$. Let us find the set of compressor locations, such that a sensor at z, that achieves the same cost through the compressors at these locations as when the sensor send data directly to the sink. Let $(x, y) \in \mathbb{R}^2$ denote the position of such a compressor, then the cost for a sensor to sending directly to sink is $|z|^2$, where the cost via the compressor at (x, y), is given by

$$e(z) = (x - |z|)^2 + y^2 + \alpha(x^2 + y^2).$$

By equating these, we have the following condition on (x, y) that makes the routing choices even in terms of the cost as follows:

$$(x - \frac{|z|}{1+\alpha})^2 + y^2 \quad = \quad (\frac{|z|}{1+\alpha})^2.$$

For the sensor to directly associate itself with the sink, there should be no compressors within the sphere given by the above expression. This 'void' region plays an important role in determining the shape of the tessellation. In the Voronoi scheme, the region for a sink cell is also a sphere. Thus both tessellations have convex shapes, and the optimal tessellation reduces to the Voronoi tessellation when $\alpha = 0$.

An example of an optimal tessellation when $\gamma = 2$ is shown in Fig. 2.6. We continue with geometrical properties of this tessellation. Let us denote the optimal tessellation for this case as $\mathcal{T}^{\alpha,2}(\Pi_1, \Pi_2)$. Then the following result gives the area of a typical sink cell.

Theorem 8. *The expected area of a typical cell associated with a sink in the tessellation $\mathcal{T}^{\alpha,2}(\Pi_1, \Pi_2)$ is given by*

$$\mathbf{E}_2^0[\ |T_0^{\alpha,2}(\Pi_1, \Pi_2)|\ ] \quad = \quad \frac{1}{\frac{\lambda_1}{(1+\alpha)^2} + \lambda_2}$$

26
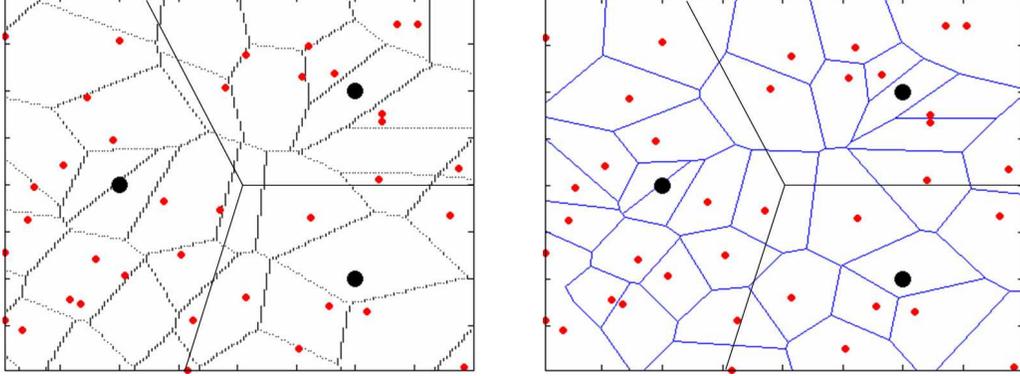
Figure 2.6: The left figure shows optimal tessellation with sinks (big dots) and compressors (small dots) when $\gamma = 2$ and $\alpha = 0.45$. The figure on the right is a Voronoi tessellation with same configuration of sinks and compressors. In the left figure, the tessellation is convex and it has cells that have similar shapes to the Voronoi cells. Especially the cell which does not contact the boundaries of $\mathcal{V}(\Pi_2)$ cells, is expanded to the factor of $(1 + \alpha)^2$ compared to the Voronoi cell with same nucleus.

where $T_0^{\alpha,2}$ denotes the typical cell with the sink at the origin as its nucleus, $|\mathcal{C}|$ denotes the area of the set $\mathcal{C}$ and λ_1, λ_2 denote the density of compressors and sinks respectively.

Using a similar techniques used in deriving Theorem 6, i.e. the data from a sensor eventually goes to the sink which is the closest to that sensor, one can get energy cost estimate. The cost function per sensor is modified to

$$e(x_i) = |x_i - c(x_i)|^2 + \alpha|c(x_i) - s_V(x_i)|^2 \qquad (2.7)$$

where $s_V(\cdot)$ denotes the function that maps a location to the location of its closest sink. Let us denote the typical cost seen by a sensor under this suboptimal scheme as G_{U0}^2. Then we have the following result.

Theorem 9. *The energy cost of a typical sensor, G_{U0}^2, under energy cost (2.7) is an upper bound for the cost seen by a typical sensor in the optimal tessellation. Also*

G_{U0}^2 *is given by*

$$G_{U0}^2 = \frac{\lambda_2}{\pi(\frac{\lambda_1}{(1+\alpha)^2} + \lambda_2)^2} + \frac{\frac{\lambda_1}{(1+\alpha)^2}\left\{\frac{\alpha\lambda_1}{(1+\alpha)^2} + (2\alpha+1)\lambda_2\right\}}{\pi\lambda_2(1+\alpha)\left\{\frac{\lambda_1}{(1+\alpha)^2} + \lambda_2\right\}^2}. \tag{2.8}$$

To check for consistency, note that when $\alpha = 0$, $G_{U0}^2 = \frac{1}{\pi(\lambda_1+\lambda_2)}$ and it is indeed the same as the cost for a typical sensor to associate itself with the closest compressor or sink. It is interesting to note that when $\alpha = 1$, G_{U0}^2 *does not* correspond to the cost for the typical sensor to associate itself with the closest sink, which differs from the case where $\gamma = 1$ (Note that when $\gamma = 1$, if there is no compression gain ($\alpha = 1$), it is always optimal for a sensor to go straight to the closest sink because of the triangle inequality.). Instead, when $\gamma = 2$, the average optimal cost is always less than the average cost for a sensor to directly send to a sink as long as $\lambda_1 \neq 0$. This implies that *the optimal relaying always reduces the average cost, even if the aggregation point does not perform any compression.* Hence if we replace α with 1 in the Eq. (2.8), we obtain a bound on the average cost achieved by the optimal routing alone. We shall revisit this result via simulation. A proof of Theorem 9 is given in the appendix.

In the next section, we give some numerical results on the multi-hop case and simulation results with different values of γ. For the case when $\gamma \neq 2$, a detailed analysis is mathematically intractable so we study energy distribution and tessellation through simulations. Also we explore how the cost function is distributed over the entire field, and some issues on traffic congestion resulting from hierarchical aggregation schemes.

2.5 Numerical and simulation results

2.5.1 Numerical Evaluation

We numerically evaluated the cost functions computed in the previous section. On the top in Fig. 2.7 we compare the energy consumption at a typical sink of three possible hierarchical schemes for different compression ratios α. G_S is the case where sensors send directly to the sink without going through compressors, G_V is the case where sensors send to a compressor or a sink, whichever is closer, and
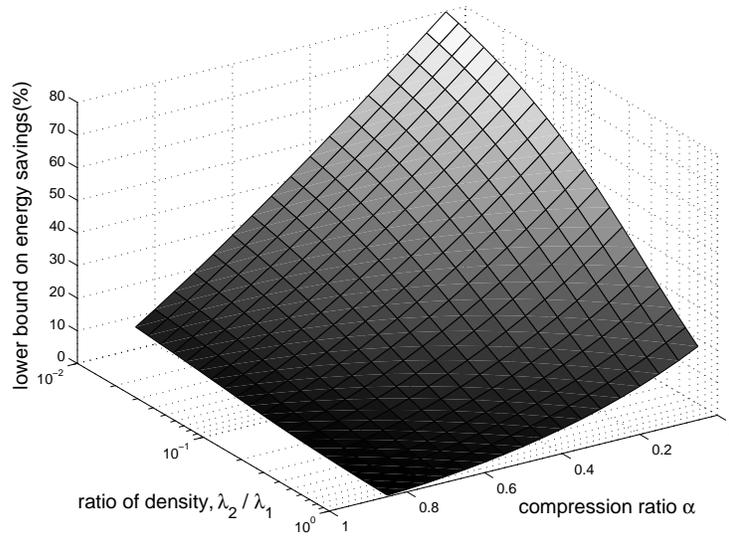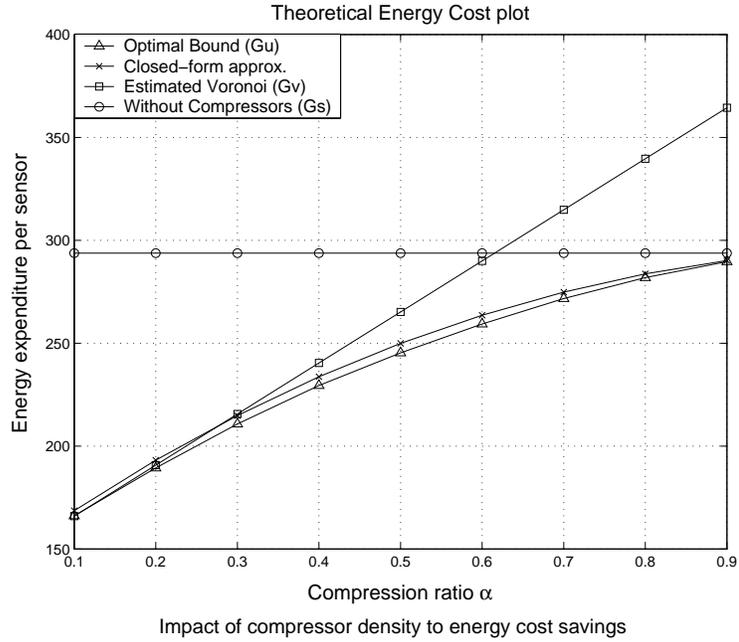
Figure 2.7: The graph at the top exhibits a numerical evaluation of G_S, G_V and our upper bounds G_U and $\hat{G}_U$ for the optimal hierarchy, as a function of α with $\lambda_0 = 1$, $\lambda_1 = 0.0473$ and $\lambda_2 = 0.0142$. The graph at the bottom exhibits energy savings, comparing G_U to G_S, as a function of α and λ_2/λ_1.

then compressors send to sinks. G_U is our upper bound on the cost for the optimal strategy where sensor traffic eventually goes to the closest sink, but may be relayed and compressed by an intervening compressor if that minimizes the cost. We observe that organizations based on an optimal tessellation as represented by upper bound G_U can outperform those based on Voronoi tessellation, i.e. G_S and G_V. With moderate compression performance, e.g. $0.5 \leq \alpha \leq 0.9$, the relative energy savings of the optimal scheme versus the Voronoi scheme is shown to be 6%-20%. Also we can see that the closed-form approximation $\hat{G}_U$ is a tight upper bound of G_U: the error is within 1.4% of G_U. Note that $\hat{G}_U = \frac{\lambda_0}{\lambda_2} \hat{G}_{U0}$ by Neveu exchange formula, where $\hat{G}_{U0}$ is an estimate of the cost for a typical sensor and given by Theorem 7. We observe as α increases the second scheme G_V is inferior to the simpler scheme G_S; this is intuitive since G_V is constrained to connect to the closest compressor if there are no sinks which are closer, thus requiring a "detour" even when no compression gain is available. Note however, that our optimal scheme always performs better than the other two. In particular as soon as the aggregation/compression nodes provide a reasonable compression ratio it is able to significantly reduce the overall energy cost of the network. On the bottom in Fig. 2.7, we have plotted the energy costs of $\hat{G}_U$ (approximation on the optimal bound) relative to G_S (no compressors) as a function of the ratio between the density of compressors, λ_1, and sinks, λ_2. We observe that the sensitivity of the gain using the optimal scheme increases as the density of compressors increases. As λ_1 increases while λ_2 is fixed, the sensor can *adapt* itself to the change in the compression ratio by having more opportunity to select the optimal compressor with varying α. For example, assume that there are a large number of compressors around a sensor. If α changes, the sensor will have more choices for candidate compressors that will provide less cost, thus larger gain. In contrast, if there are few compressors, the sensor might have to remain associated with the same compressor, hence some compression gain might be achieved if α decreases by a small amount, but no gain from the optimal routing. Thus the sensitivity of the energy savings is higher to changes in the compression ratio when the number of compressors is high - this should be clear from Fig. 2.7.
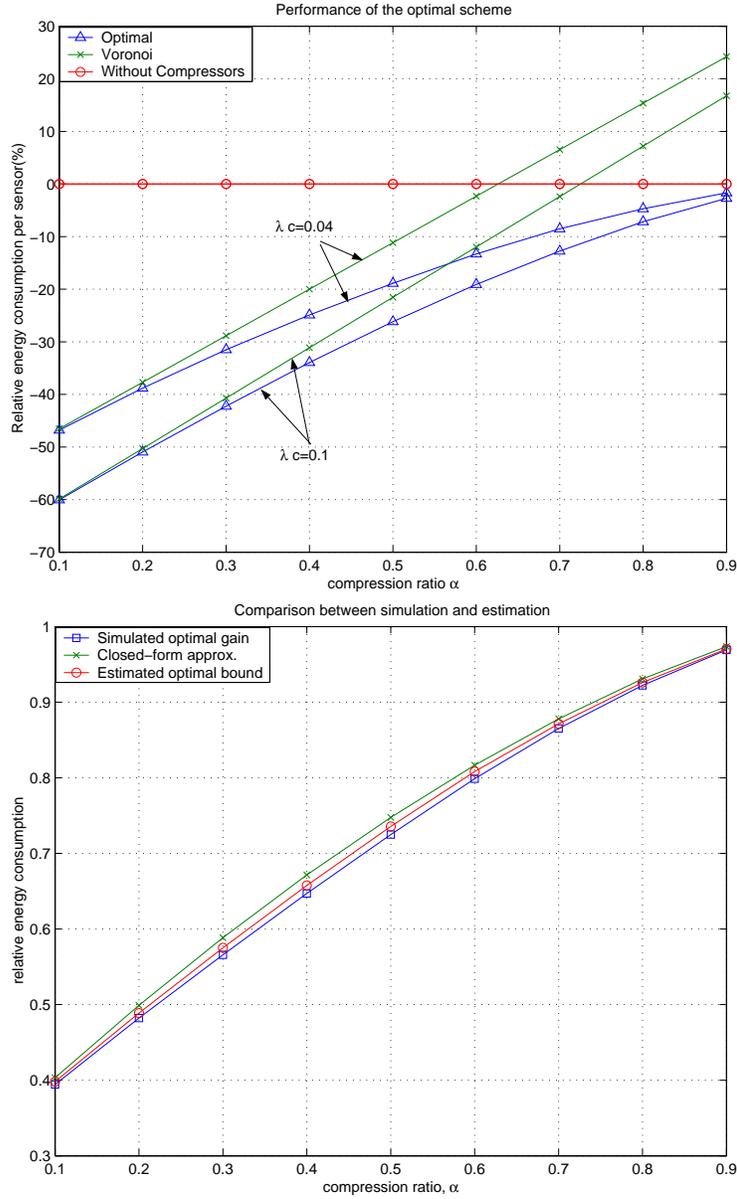
Figure 2.8: The performance comparison between the optimal and Voronoi scheme and comparison between the analytical approximation and simulation results when $\lambda_1 = 0.1$ and $\lambda_2 = 0.01$.

2.5.2 Simulations

2.5.2.1 Cost performance in Multi-hop case

In our simulations, we placed compressors and sinks according to two homogeneous Poisson point processes on a 100×100 unit square area. The numbers of compressors and sinks are Poisson random variables with mean $100^2 \times$ (target density). Once the number of nodes are determined, the nodes are placed uniformly on the square. To facilitate simulation, we placed a typical sensor at the origin and we have varied the density of compressors and sinks to compute the average cost incurred at the typical sensor.

Fig. 2.8 shows the simulated cost savings for the optimal hierarchy and that based on a Voronoi tessellation relative to the baseline scheme without compression i.e., directly sending traffic to the closest sink. On the top in Fig. 2.8, we show the relative energy cost of a typical sensor compared to a scheme without compressors but maintaining the configuration of sinks. Fixing the density of sinks as 0.01 (i.e., on average, a sink takes care of 100 sensors), we have simulated two different setups. The first case is when the compressor density $\lambda_c = 0.1$ and the second case is when $\lambda_c = 0.04$. As can be seen from the plot, the relative gain of the optimal and Voronoi scheme is larger with a higher density of compressors ($\lambda_c = 0.1$). However, if we consider the relative performance of the optimal scheme to the Voronoi scheme, the optimal scheme outperforms Voronoi scheme more when the compressors are sparse ($\lambda_c = 0.04$). The intuition is, as there are more compressors in the field, the probability that the closest compressor becomes the optimal compressor increases, and Voronoi scheme performs well. By contrast, if there are not many compressors available, it becomes more critical for a sensor to pick the "right" compressor, i.e. the one that will provide the minimum energy cost. When the compression performance is high ($\alpha \leq 0.5$), we see that the difference in the energy savings between the optimal and Voronoi strategy is small. Thus the Voronoi scheme is nearly optimal if one can achieve a high compression gain. When the compression performance is moderate ($0.5 \leq \alpha \leq 0.9$), the optimal scheme outperforms the Voronoi scheme by about 8%-28% when $\lambda_c = 0.04$ and 5%-20% when $\lambda_c = 0.1$. Also the optimal scheme is always better than the scheme without compressors, but Voronoi scheme degrades as the compression ratio becomes moderate. On the bottom in Fig. 2.8, we have compared the relative performance of the optimal scheme, suboptimal scheme and the closed form upper bound $\hat{G}_U$, to the baseline scheme (a scheme without compressors). The
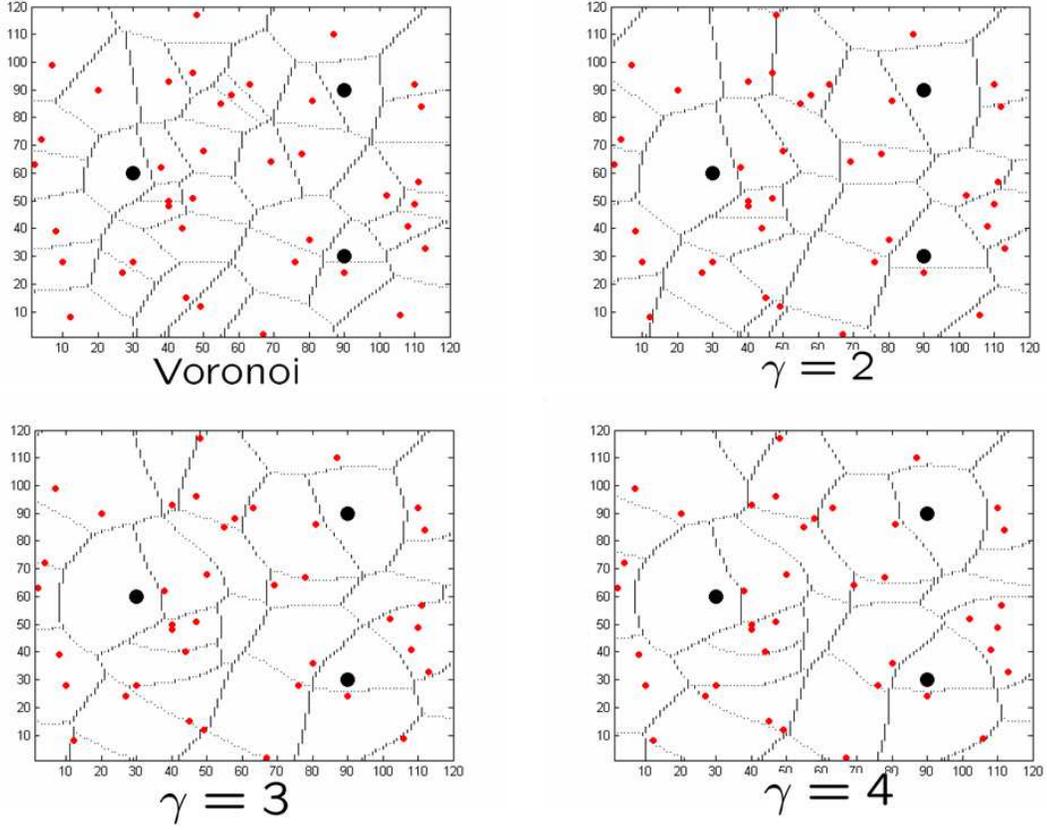
Figure 2.9: Tessellations of two-hop scheme with varying path-loss exponent γ.

densities of compressors and sinks are set to $\lambda_1 = 0.1$ and $\lambda_2 = 0.01$ respectively. One can see that the theoretical estimates are quite close to the simulation results for a range of compression ratios.

2.5.2.2   Two-hop case: tessellations and performance

Examples of optimal tessellations when the path-loss exponent $\gamma = 2, 3$ and 4 are shown in Fig. 2.9 with the Voronoi tessellation as a reference. When $\gamma = 2$, we obtain a convex tessellation as discussed in the previous section. When $\gamma \neq 2$, the tessellations are no longer convex and they look more distorted relative to the original Voronoi tessellation. Note that some cells do not contain their nucleus at higher γ, as discussed in the previous sections. Also the simulation of comparison in
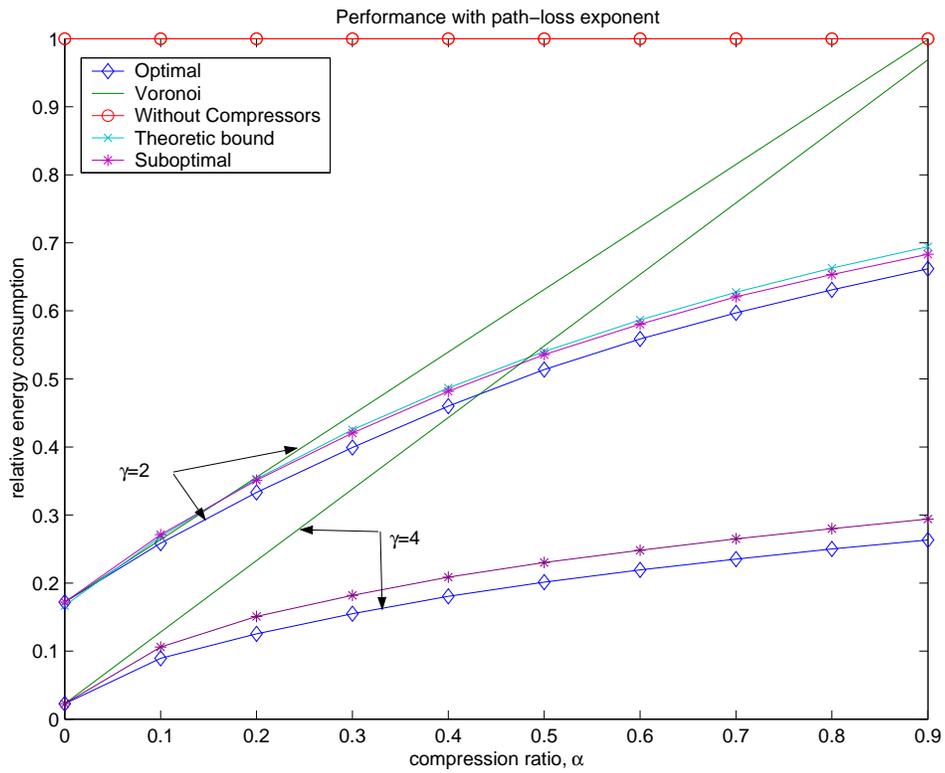
Figure 2.10: The comparison of energy expenditure of the optimal and Voronoi scheme, relative to a scheme without compressors as a function of compression ratio α for the path-loss exponents $\gamma = 2, 4$ when $\lambda_1 = 0.1$ and $\lambda_2 = 0.02$.

relative energy savings in the optimal strategy with the Voronoi strategy, with the path-loss exponents of $\gamma = 2, 4$ is shown in Fig. 2.10. The performance criteria is the energy consumption relative to the baseline scheme without compressors. The relative gain increases dramatically with γ, but as we can see from the case when $\gamma = 4$, the impact of the proper routing choice dominates the performance and the compression ratio α plays a lesser role in energy savings. Also we see that the proposed suboptimal scheme is a good estimate of the optimal scheme, and matches well with the theoretical estimates when $\gamma = 2$, see Eq. (2.8). Furthermore, our proposed scheme outperforms the Voronoi scheme considerably versus the $\gamma = 1$ case. Under moderate compression performance ($0.5 \leq \alpha \leq 0.9$), when $\gamma = 2$, the optimal scheme has an average gain over the Voronoi scheme of 14%-34%. When $\gamma = 4$, the optimal scheme has an energy savings over 45%-70% as compared to the Voronoi scheme. Finally, we observe that even with an inferior compression performance ($\alpha = 0.9$), the optimal scheme still gives a considerable gain, 36% when $\gamma = 2$ and 74% when $\gamma = 4$. This implies that the proper routing always provides large savings in the energy even without any compression gain, as discussed in [8].

2.5.2.3 Energy contribution plots and congestion

In Fig. 2.11, we show the spatial distribution of the energy contribution for an optimal Johnson-Mehl tessellation for $\alpha = 0.5$ and a hierarchical Voronoi scheme. The figure on the top corresponds to the optimal strategy. Red (bright) regions indicate areas where sensors incur a higher energy cost to the overall system, and blue (dark) regions are the areas of less cost. Note the color scale shows relative intensity in each scheme. The absolute value of energy consumption is lower in the left(optimal) scheme. We observe that, in the optimal tessellation, the energy contours vary in a smoother way versus the Voronoi scheme in which every node just connects to its closest compressor and/or sink. The reason is that the optimal tessellation is constructed for each sensor to achieve the minimal cost at its location, and the cost function is a smooth function of the sensor positions in $\mathbb{R}^2$. In contrast, in energy contribution plot for Voronoi scheme, the tessellation boundaries stand out in the contours since the sensors simply associate themselves with closest compressors (or sinks) and do not exploit cost-minimizing routing scheme.

Still it is of practical interest to examine the *traffic congestion* under various aggregation schemes. If the data collected at each sensor is routed to a fixed aggre-
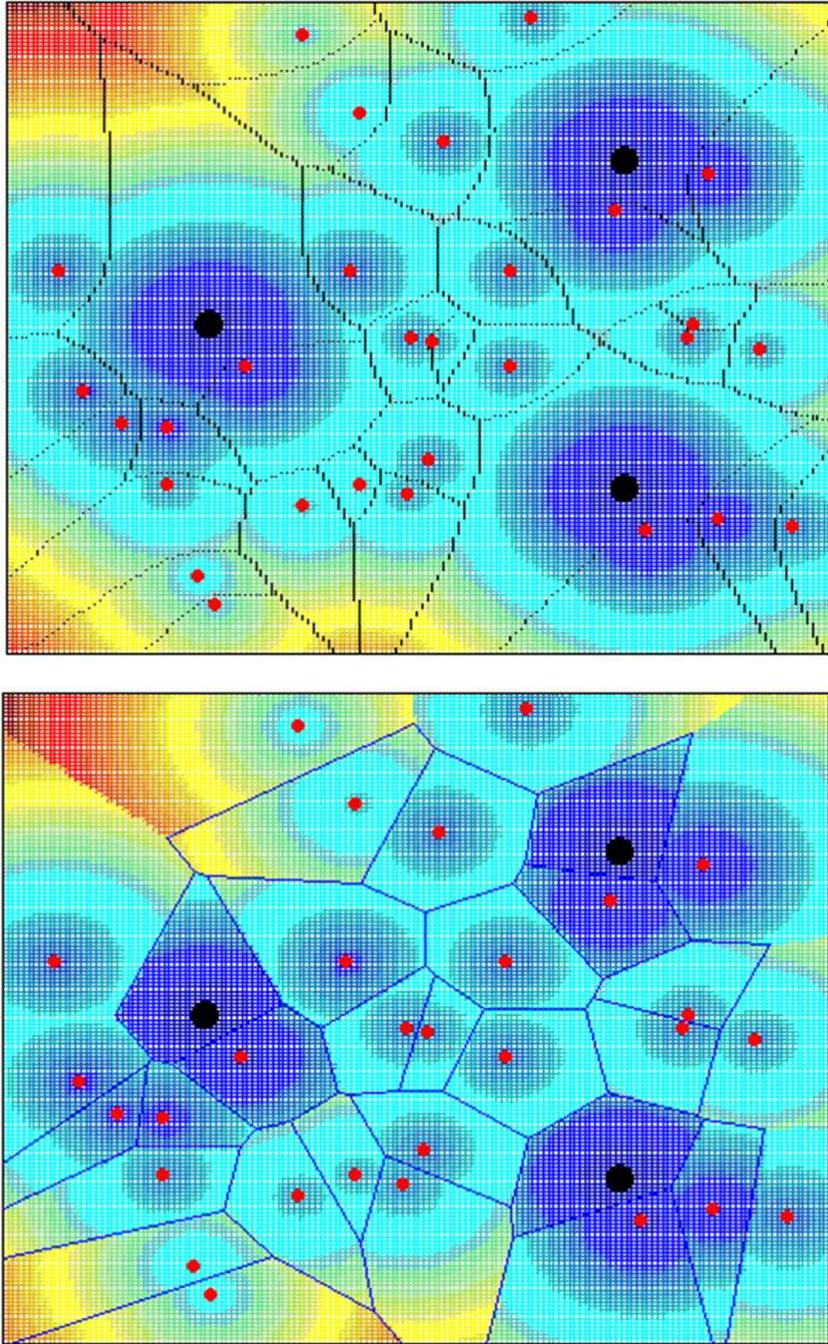
Figure 2.11: Spatial distribution of the energy contribution of sensors to the overall cost.
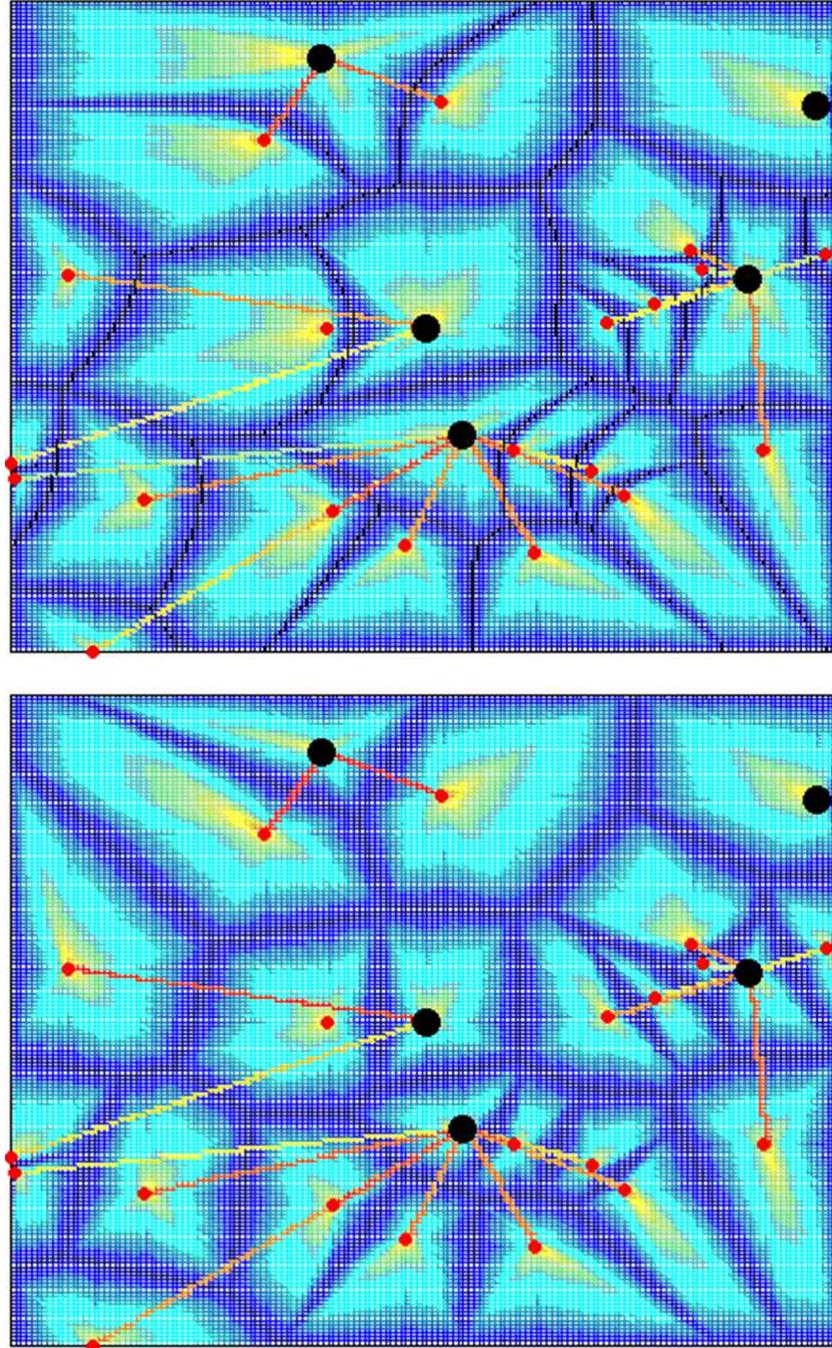
Figure 2.12: The relative energy burden by the number of packet transmissions required for every sensor in the field to send 1 unit of data to respective sink where $\alpha = 0.5$ for the optimal tessellation.

gation point, the sensors next to each aggregation point will have a higher energy burden since they have to relay all the incoming packets generated from sensors within the cell of the aggregation point. Fig. 2.12 exhibits a rough estimate of such energy consumption per unit transmission time at each location in a network, assuming that the density of the sensors is high. The figure on the top corresponds to the case for the optimal tessellation. The unit transmission time is defined as the duration or the cycle of every sensor transmitting 1 unit of data to each sink. We can observe that the energy burden around the aggregation points, i.e. compressors or sinks, is quite high. Also, the region of sensors near the boundaries of the tessellation experiences a relatively low energy burden. Thus sensors that are close to the aggregation points are subject to quickly running out of energy in a time scale much shorter than that of a sensor along the tessellation boundary. Furthermore, it must be ensured that these congested routing paths are able to support such flow, i.e. the transport capacity along the shortest paths should be large enough to accommodate the traffic. These are crucial scalability problems and must be taken into account when designing a large-scale sensor network.

2.6   Conclusion

In this chapter we address two questions involving data compression and traffic aggregation. We have investigated how to optimally compress sensed data when there is a known aggregation cost for each sensor's data, e.g., the energy expended to gather the data to a query point. This provides an efficient solution to gathering information to an aggregation/compression point, but will not scale in a large network of sensors. Thus next we considered how data aggregation might be optimized in large-scale sensor networks. We have introduced a hierarchy of sensors, compressor/aggregation nodes, and sinks and show that the optimal hierarchy is associated with a Johnson-Mehl tessellation induced by the locations of the compressor and sink nodes. Using models from stochastic geometry we study the aggregation costs and properties of such hierarchical organizations, and demonstrate how these might depend on the densities of various types of nodes and the degree of compression one might achieve in a given application. Our results suggest that there are significant gains to be achieved in jointly optimizing compression and aggregation structures for sensor networks.

Chapter 3

Proactive Multipath Routing for Unstructured Traffic

3.1   Introduction

Energy efficient design and operation of ad hoc multi-hop wireless networks is a key problem in the context of mobile and/or distributed sensing applications, where energy storage and availability may be quite limited. There are many levels at which one can address this problem. Advances in silicon technology can realize energy savings through power efficient circuitry, e.g., voltage scaling, while specialized architectures can be devised to allow components to enter 'sleep' modes. At the same time power control and optimized MAC protocols which put nodes to sleep can bring substantial energy savings enabling networks with thousands of sensors. Particularly, in large ad hoc wireless networks the data originated from a source might need to be relayed a *long distance* to a destination or sink wireline node. Relaying through many hops may cause intermediate nodes to consume substantial amounts of energy and thus make energy efficient routing a particularly critical task.

Consider the network shown in Fig. 3.1. Two sources $S1, S2$ send to destinations $D1, D2$ on opposite ends of the network respectively. In the network on the left these sessions are supported along shortest hop routes. If one of these sessions were sustained for a long time, nodes along the route would eventually see depleted energy reserves, roughly 'dividing' the network into two parts. Subsequently if other nodes needed to communicate across this depleted zone they may result in exhaustion of energy along the diagonal, or require selection of routes around this area of the network, which in turn would incur additional energy burdens. This simple example, shows how energy depletion along long routes combined with interactions with future overlapping and/or routing of additional traffic flows might exacerbate the energy problem. A natural solution to this problem is to spread out the energy burden of sustained sessions so as to obtain a spatially balanced energy burden. Specifically, one may split traffic across two disjoint routes as shown on the right in Fig. 3.1. Assuming energy consumption is roughly proportional to the load this
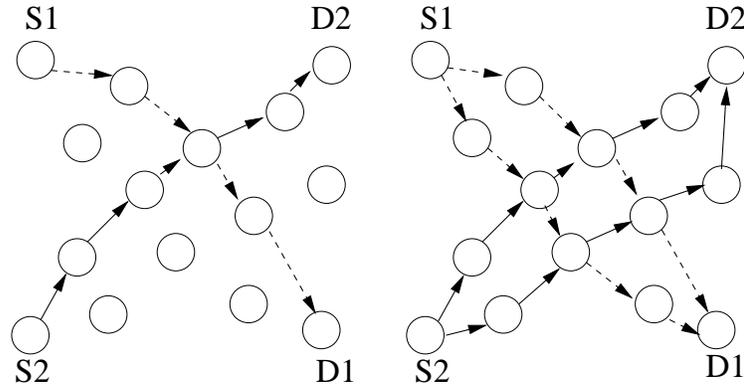
Figure 3.1: Comparison of the shortest path routing scheme (on the left) and a typical load-balancing scheme (on the right). The dotted arrows represent flows for $S1 - D1$ and the dashed arrows represent flows for $S2 - D2$, respectively.

leads to a more balanced energy burden across sets of intermediate nodes. At the same time this scheme may involve a larger number of nodes, e.g., a route with four versus three hops, and thus an increased overall energy burden.

In this chapter, we consider overall system design aspects for such multipath routing strategies – we refer to this as *proactive* balancing of energy burdens over multiple routes. Our primary interest here is not to devise detailed multipath routing algorithms, but rather to investigate the design of, and possible improvements afforded by, such routing mechanisms. The key intuition is that the more we spread the traffic, the more the energy profile of the network will be balanced. However, spreading traffic requires that some packets take long 'detours', which will incur extra energy cost. This tradeoff associated with the degree of spreading is the main topic investigated in this chapter.

To this end, we use a simple, idealized model to characterize and parameterize the spatial energy balancing aspects of proactive multipath routing schemes. Our model provides sharp insights on design choices under various scenarios. For example, one of the key issues studied in this chapter is the degree to which a session's traffic should be spread, depending on the load and the distance it must travel. Not surprisingly, we show that traffic should be spread *more* as the load and the hop count *increase*, and provide a simple scaling rule to proactively adapt the degree of spreading.

This chapter is organized as follows. In Section 3.2 we discuss related work in this area. Section 3.3 introduces a concrete multipath routing and balancing strategy, and presents continuum and grid models. In Section 3.4 we characterize spatial energy burdens using a shot-noise process associated with our continuum model. Section 3.5 uses a grid network to explicitly analyze a parameterized family of energy balancing strategies. In Sections 3.6 and 3.7 we formulate and investigate the design and optimization of such spreading using second order and asymptotic approximations. Section 3.8 includes simulation results and a discussion of various scenarios. Finally Section 3.9 presents our conclusions.

3.2   Related Work

There has been substantial research on the design and implementation of energy conserving routing protocols suitable for ad hoc networking applications. Let us review some of this work. In [31] a characterization and algorithm determining the most energy efficient route between two nodes is proposed. However, it is not clear whether using such routes will extend 'network lifetime,' nor how this would impact network capacity for non-homogenous traffic loads. By contrast, [32] and [33] propose and evaluate routing mechanisms to maximize network lifetime based on nodes' current residual energy reserves. Unfortunately, scalability and the effectiveness of greedy routing to spread energy burdens are a concern. Among recent work, [34] and [35] show that, by properly defining cost metrics as exponential functions of the residual energy at each node, one can achieve competitive optimality for throughput under energy constraints. The work of [36] takes yet another tack – they propose packet-level randomized routing in order to proactively balance energy burdens across the network. A unifying principle emerges from this body of work: the tradeoff between minimizing the energy expended to carry an offered load versus the balancing of energy burdens across the network. To the best of our knowledge, the spatial character of this tradeoff has not been studied. The primary contribution of this work is the use of a stochastic geometric framework to analyze, and then work towards realizing this tradeoff in an 'optimal' manner.
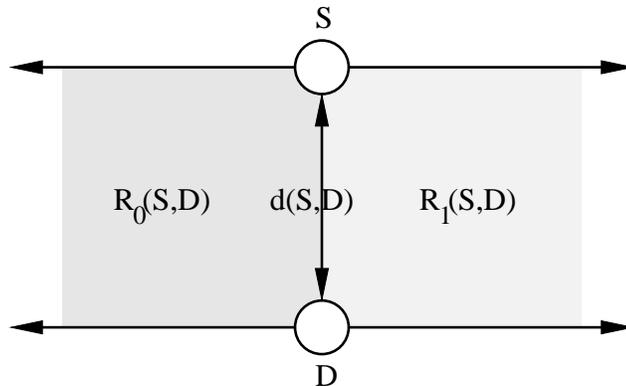
Figure 3.2: Illustration of regions $R_0(S, D)$ and $R_1(S, D)$ for the source-destination pair (S, D).

3.3 Spatial Modelling

This section is divided into three parts. We start by stating our model assumptions. We then introduce a multi-path routing scheme based on nodes' spatial relationships. Finally, we propose a continuum model where we regard the field of the wireless nodes as an infinitesimal 'medium' that carries fluid, i.e., the traffic flow. This leads to a simple shot-noise process model for the spatial field of energy expenditures which is amenable to analysis.

3.3.1 Model assumptions

We will use a simplified model for energy expenditures associated with data transmissions. Nodes are assumed to share a common transmit power level sufficient to guarantee the network is connected. They relay packets towards the destination via *neighboring* nodes in a hop-by-hop manner. Our key assumption is that the energy consumption at each node is *proportional* to the traffic it is carrying, so we use the terms 'traffic load' and 'energy burden' interchangeably. We refer to a flow of traffic between a pair of source-destination nodes as a *session*, and refer to the source and destination nodes as a *session pair*. In reality there are other factors that contribute to energy burdens in wireless networks, e.g., interference and channel contention. Such factors may be critical depending on the traffic offered to the network which may result in energy burdens that are not proportional to traffic loads. Nevertheless a rough model based on proportionality makes sense, specifically

in the long run we may assume that such extraneous factors are averaged out.

We assume traffic is relayed only via 'neighboring' nodes which we shall define based on proximity as follows. We model the locations of the nodes as fixed and following a spatial point process in $\mathbb{R}^2$ plane. A natural notion of proximity can be introduced via the Voronoi tessellation and Delaunay graphs induced by the locations of the nodes. These are discussed below.

Suppose the locations of the nodes constitute a point process Ψ on the $\mathbb{R}^2$ plane . Each point $x_i \in \Psi$ serves as a *seed* for a *cell* $V(x_i)$,

$$V(x_i) = \{y \mid |x_i - y| \le |x_j - y|, \forall x_j \in \Psi\}$$

in the Voronoi tessellation induced by Ψ. If $V(x_i) \cap V(x_j)$ is not an empty set, we refer to $V(x_i)$ and $V(x_j)$ as *neighboring cells* and we say that x_i and x_j are neighbors.

A *Delaunay graph* is a graph whose vertex set is Ψ and whose edges connect nodes that are neighbors. We denote the Delaunay graph by $G(\Psi, E)$ where E is the set of Delaunay edges. Routes considered in the discussion below will be based on the Delaunay graph. We shall assume that the spatial distribution of nodes is fairly uniform and sufficiently dense that each node can in fact reach its neighbors.

3.3.2 Proximity-based multipath routing

Consider a route connecting two nodes $x_i, x_j \in \Psi$, which has a minimal length, i.e., sum of the Euclidean lengths of the edges it traverses. This path is referred to as the *Shortest Delaunay Route* (SDR) and has a length that is within a factor of 2.42 of the Euclidean distance between x_i and x_j, see e.g., [37], [38]. Note that the SDR is based on the Euclidean norm, thus the SDR may not coincide with a route having a minimum number of hops. We will see in the sequel (Section 3.8) that this subtle difference may impact the *spatial distribution* of energy expenditures significantly.

Based on the SDR, we propose the following simple construction for a set of paths between two nodes, say $S, D \in \Psi$. In order to visualize their spatial characteristics we present a geometric view of this construction :

1. Draw a straight line segment $d(S, D)$ between S and D, and draw two additional lines, through S and D and orthogonal to $d(S, D)$. Let $R_0(S, D)$ and
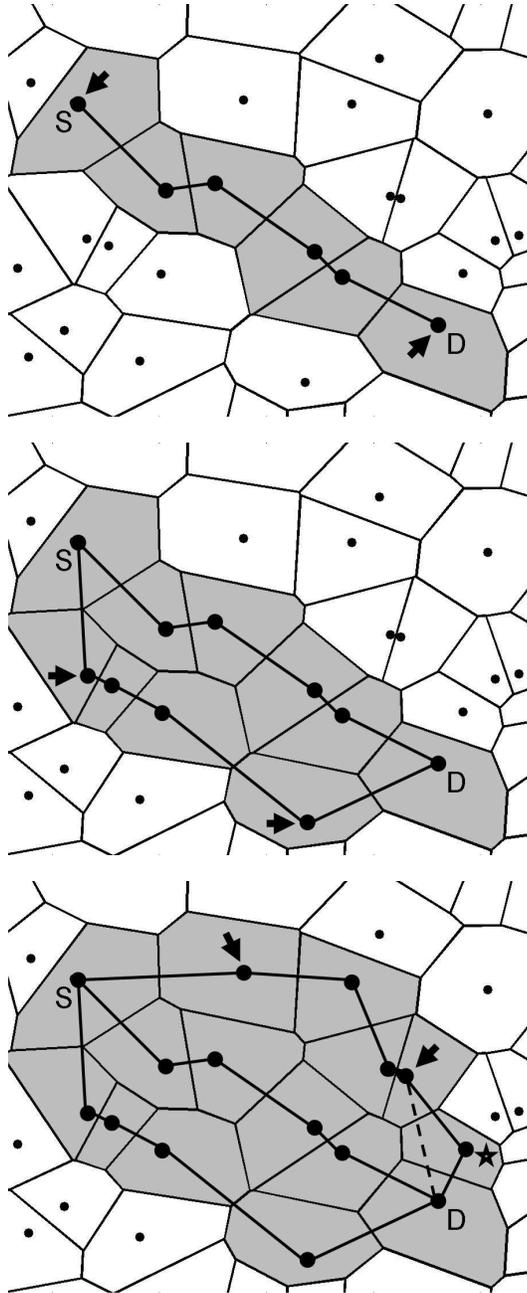
Figure 3.3: The figure shows the construction of Level 1,2 and 3 routes for nodes S and D, from top to bottom. The route at each stage is shown with solid lines and the shaded regions are the cells for the nodes in routes. The nodes marked by an arrow are the *connectors* at each route.

$R_1(S, D)$ denote the open planes with their boundaries being $d(S, D)$ and its orthogonal lines as shown in Fig. 3.2.

2. We let $N_1(S, D)$ denote the set of nodes included in the SDR from S to D, and refer to this route as the Level 1 route. S and D are referred to as the *Level-1 connectors*.

3. Now find the set of nodes $N_2(S, D)$ which are neighbors of $N_1(S, D)$ and fall in $R_0(S, D)$. Create a route that connects the nodes in $N_2(S, D)$ with Delaunay edges. We refer to the two nodes located at each end of this route as *Level-2 connectors*.

4. Construct a SDR for a Level-2 connector to its closest Level-1 connector and repeat the same for the other connector. If this SDR crosses new nodes, update $N_2(S, D)$ by adding these new nodes in $N_2(S, D)$. Now the nodes in $N_2(S, D)$, S and D can be connected via a Delaunay route which is referred to as *Level-2 route*.

5. Next determine the set of nodes $N_3(S, D)$ which are neighbors of $N_1(S, D)$ but falls in $R_1(S, D)$ this time. Following the similar process as above, find the Level-3 connectors, update $N_3(S, D)$ and construct the Level-3 route.

6. For $w \geq 4$, determine the new set of nodes $N_w(S, D)$ that are neighbors of N_{w-2} and fall in $R_{(w \mod 2)}(S, D)$. Following the above steps, the Level-w route is recursively constructed.

The basic idea is to recursively construct higher level routes based on nodes which are neighbors of those included in previous levels but alternating between $R_0(S, D)$ and $R_1(S, D)$ in order to balance the spreading cost as the levels increase. We confine relaying nodes to the regions $R_0(S, D)$ and $R_1(S, D)$ so as to prevent routes from extending backward[1]. As will be clear from the construction, the role of connectors is to ensure connectivity via Delaunay routes among S, D and the routes at different levels. These routes can be constructed by each node if it has the information on the locations of its neighbors, the source and the destination.

---

[1]In step 4) of the construction, some nodes that are not contained in these regions may be included in a route.
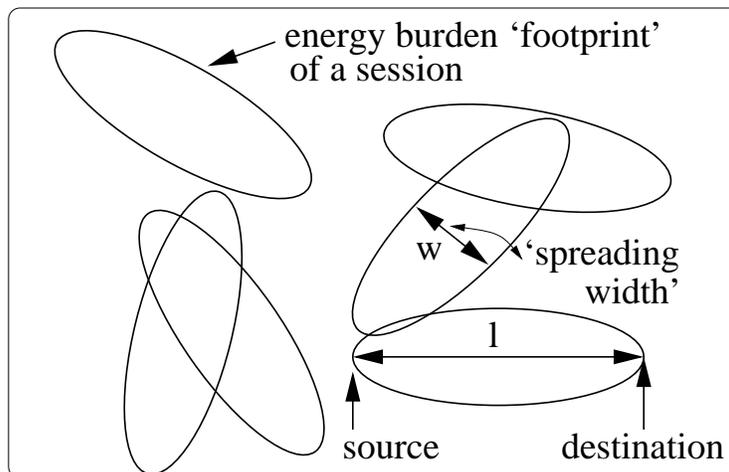
Figure 3.4: A realization of the energy footprints for sessions in ad hoc network. A footprint is assumed to have elliptical shape for the purpose of illustration.

An example of our route construction for a source-destination pair (S, D) is illustrated in Fig. 3.3. At each level, the level connectors are marked with arrows. Note that at level 3, the level connector associated with node D is not its neighbor, thus an SDR route is constructed between them. Also note that this route includes a node, indicated by a star, which is in fact not contained in $R_1(S, D)$.

We shall refer to this construction as *proximity-based multipath* (PBM) routing which gives us a concrete set of paths over which to distribute traffic so as to spread out energy burdens.

As shown in Fig. 3.3, the set of PBM routes associated with a session is *spatially clustered* by construction. We shall refer to the set of nodes in such a cluster as the *spatial footprint* of a session. For example, the nodes within the shaded Voronoi cells in Fig. 3.3 correspond to the spatial footprints of the session as we increase the degree of spreading. Thus the traffic pattern on a network can be viewed as a dynamic set of, possibly overlapping, spatial footprints. This is the basic idea behind the *continuum* model introduced in the next section.

3.4 Characterization via Continuum Model

3.4.1 The continuum model and shot-noise formulation

Consider the case where the density of the nodes in a network is large. We can think of an infinitesimal area in space as corresponding to a node with an initial energy reserve. Each footprint will correspond to a closed set in $\mathbb{R}^2$, and is assumed to have a well-defined, possibly random, shape. Since an infinitesimal area (node) serves as a 'carrier' of a flow, a footprint can be regarded as a 'vessel' which contains the flow from a source to a sink. Session pairs are located at the ends of footprints. We refer to the 'length' of a footprint which corresponds to the distance between a session pair as the *span*, and the maximum 'width' of a footprint as the *spreading width*. We refer to the midpoint of the session pair as the *center* of the associated footprint. We assume that the centers of footprints constitute a point process in $\mathbb{R}^2$. Fig. 3.4 exhibits a realization of the process capturing the energy burdens incurred over a period of time – only sessions' footprints with span l and spreading width w are shown.

Each location within a footprint would, in general, experience a different load/energy burden. To model this, we define a *load distribution* function which is the load per unit area, i.e., the energy burden *density* at each location, in a footprint. The load distribution function depends on the 'strategy' used to spread traffic within a footprint – this will be further quantified in the sequel.

The proposed continuum model can be mathematically formalized as follows. Let Φ_0 denote an isotropic, random closed subset of $\mathbb{R}^2$. We will assume Φ_0 has the distribution of a typical footprint in a given network. The load distribution function $h(\cdot, \Phi_0) : \mathbb{R}^2 \to \mathbb{R}_+$ gives the spatial density of energy burden at each location, for a session with unit load. Specifically, for $y \in \mathbb{R}^2$, if $y \notin \Phi_0$ then $h(y, \Phi_0) = 0$ and otherwise $h(y, \Phi_0)$ corresponds the relative energy burden per unit area at location y of the footprint Φ_0.

We assume that the centers of sessions constitute a homogeneous spatio-temporal Poisson point process Π with intensity λ per unit time per unit area. Let us denote by Π_t a *spatial* point process in $\mathbb{R}^2$ for the centers of the sessions/footprints that have been offered to the network during time $[0, t]$. Thus Π_t is a homogeneous spatial Poisson point process with intensity of λt. Each point $x_i \in \Pi_t$ has an associated footprint denoted by Φ_i. We assume $\{\Phi_i\}$ are i.i.d. copies of Φ_0. The

contribution of the energy burden on location x from a session centered at x_i with footprint Φ_i, is given by $h(x - x_i, \Phi_i)$. Note since we equivocate load and energy burden, $h(\cdot, \Phi_i)$ depends on how a routing mechanism chooses to spread the flow of session i within its footprint Φ_i. For now, assume that the 'strategy' $h(\cdot, \Phi_i)$ does not depend on the amount of load Φ_i is carrying, and denote the offered load, in bits, by U_i – these are assumed to be i.i.d. with the same distribution as U. Thus $h_{U_i}(\cdot, \Phi_i) := U_i h(\cdot, \Phi_i)$ gives the spatial load density of session i.

The total energy burden accumulated at location $x \in \mathbb{R}^2$ during time $[0, t]$ can be represented as a *shot-noise process* as follows[2].

$$G(x, t) = \sum_{x_i \in \Pi_t} h_{U_i}(x - x_i, \Phi_i). \tag{3.1}$$

Next we state several known results from shot-noise theory. Since Π_t is stationary, we can consider a 'typical' location at the origin based on the following result.

Lemma 2. *(See [39].) Let us define $G_0(t) := G(O, t)$, the energy burden at the origin. Also let $\chi^{(n)}(t)$ be the nth order cumulant of $G_0(t)$. Since Π_t is a homogeneous Poisson process with intensity λt, we have that*

$$\chi^{(n)}(t) = \lambda t \ E\left[\int_{\Phi_0} h_U(x, \Phi_0)^n dx\right].$$

Defining the normalized mean μ and variance σ^2 as:

$$\mu := \ E\left[\int_{\Phi_0} h_U(x, \Phi_0)dx\right], \tag{3.2}$$

$$\sigma^2 := \ E\left[\int_{\Phi_0} h_U(x, \Phi_0)^2 dx\right], \tag{3.3}$$

we have that

$$E[G_0(t)] = \lambda t \mu, \ \ and \ \ Var[G_0(t)] = \lambda t \sigma^2.$$

---

[2]Precisely, if the session i has not ended by time t, the offered load by that session will be less than U_i. Thus we assume that each load is offered in an *instantaneous* manner – we are interested in the *cumulative* spatial loads induced by footprints.

We shall for now assume the load U equals 1 with probability 1, i.e., $h_{U_i}(\cdot, \Phi_i) = h(\cdot, \Phi_i)$. We revisit this in Section 3.6.

As mentioned earlier, the function $h(\cdot, \Phi_0)$ captures both the 'shape' and how the flow is spread within a typical footprint – these are the design choices one can make to control the mean and variance of the spatial energy burdens. Although using only two moments to describe the statistical properties of $G_0(t)$ may not be sufficient, the following theorem suggests this might give a good approximation (see [39]).

Theorem 10. *(Asymptotic normality of shot-noise process) Consider $G_0(t)$ defined in Lemma 2. We have that*

$$\frac{G_0(t) - \lambda t \mu}{\sqrt{\lambda t} \sigma} \xrightarrow{d} N(0,1) \quad as \quad t \to \infty$$

where $N(0,1)$ is the standard normal distribution.

From this theorem, we have that, for large t, the probability that the energy burden per unit area exceeds a prescribed level b is given by

$$P(G_0(t) > b) \simeq \phi\left(\frac{b - \lambda t \mu}{\sqrt{\lambda t} \sigma}\right), \tag{3.4}$$

$$\phi(u) := \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-v^2/2} dv. \tag{3.5}$$

In order for this approximation to be useful, we assume a typical node in the network sees a large number of overlapping footprints on average.

3.4.2 Depletion probability and network lifetime

A common criterion for the energy performance of a network is its lifetime, e.g., the time before some fraction of nodes (or any single node) drop(s) below a certain battery level. Our objective lies in the complementary question: given a *desired network operation time*, can we minimize the fraction of the depleted nodes? For example, if one wishes to operate a sensor network for a week, what fraction of nodes might survive the week and what is a good multipath routing strategy to achieve this? We believe this to be a practical objective in engineering such networks. To address this question, we shall use the approximation in Theorem 10.

Let τ be the desired network operation time, and assume $\lambda = 1$. Suppose the critical reserve level b per unit area is specified as a multiple k of $\tau\underline{\mu}$ where $\underline{\mu}$ is defined as the mean energy consumption of the *baseline* scheme, i.e., a scheme without multipath routing. Thus b is specified in terms of a factor k times the mean energy consumption of baseline scheme during τ. Thus by letting $b = k\tau\underline{\mu}$, and by defining $z_k(\tau)$ as

$$z_k(\tau) := \sqrt{\tau}\ \frac{k\underline{\mu} - \mu}{\sigma}, \tag{3.6}$$

we can estimate the fraction of nodes that have not depleted the critical level b by time τ by $\phi(z_k(\tau))$. To reduce the likelihood of depletion we wish to maximize $z_k(\tau)$ for a given τ, i.e., minimize the probability of depletion $\phi(z_k(\tau))$ through admissible choices for μ and σ.

Eq. (3.6) provides us with crucial insights. Certainly we would like to minimize both μ and σ, however as we will see later, there is a *tradeoff* between these parameters, i.e., we can decrease μ at the cost of increasing σ, and vice versa. The optimal tradeoff will depend on k. If $k\underline{\mu}$ is small, one might try to decrease μ. Conversely, if k is relatively large, one might prefer strategies that give smaller σ. This captures the fundamental tradeoff addressed in this work, i.e., that between the benefit of spatial energy balancing by traffic spreading (σ) and the cost associated with such spreading (μ). To this end, in the following section we discuss parameterized energy balancing strategies for which μ and σ can be estimated explicitly.

3.5 Proactive Energy Balancing Strategies

Let us now consider strategies that give desirable 'footprints' and 'flow distributions' over the footprints, $h(\cdot, \Phi_0)$ so as to reduce $\phi(z_k(\tau))$. For simplicity, we assume the spans of footprints are *fixed* to l. Also we let the loads offered by sessions be fixed to 1 and treat the maximum footprint width w as a design parameter. We will deal with issues involving random spans and session loads in later sections.

3.5.1 Balanced flow of continuum flows

We make an analogy of the continuum model to physical systems of fluids, in particular, flow of incompressible ideal fluid. Let us consider a fluid flow in $\mathbb{R}^2$
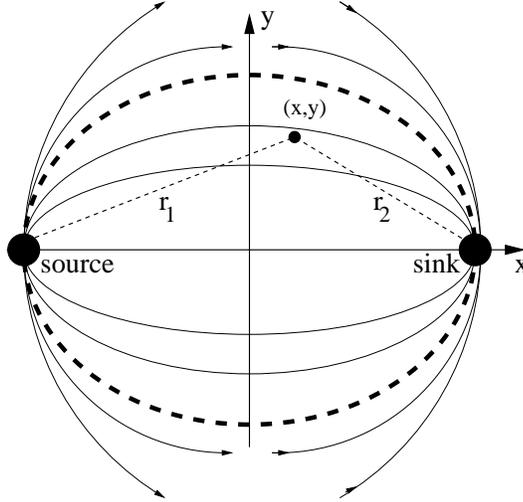
Figure 3.5: A continuum flow analogy. The curves are streamlines, i.e., the trajectories of the flow. The thick dotted curve is the boundary of an example footprint.

plane between a source and sink. Let the motion of the fluids correspond to a flow of packets, and the velocity correspond to the flow rate at each infinitesimal location, i.e., each node. The density of fluid is *constant* over the field, i.e., the packets are not queued or compressed in its size. Also there is no friction or interactions but there is only conservation of flows. Thus the flow behaves as an ideal incompressible fluid.

To be specific, we formulate this as follows. We consider a pair of source and sink on the 3-D plane, where each is a *cylinder* that extends to infinity of the z-axis but with small radius[3] ϵ, thus we only consider streams in $\mathbb{R}^2$. The source emits a flow of such ideal fluid at a constant rate which is absorbed at the sink. Here we denote the density of the fluid by ρ and the velocity of the fluid as a 2-D vector $\mathbf{v}$. In the steady state, due to flow conservation, $\mathbf{v}$ satisfies the following *continuity equation*:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \tag{3.7}$$

Here $\nabla \cdot$ is a 2-dimensional divergence operator defined as $\nabla \cdot \mathbf{v} := \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y}$. Since

---

[3]This is to avoid the singular values of the field solutions in the vicinity of the sources.

ρ is assumed to be constant, this becomes

$$\nabla \cdot \mathbf{v} = 0. \tag{3.8}$$

Thus the velocity field is *solenoidal* or *divergenceless*, i.e., there are no other flow sources present in the field except the source and the sink. Moreover, considering a 'desirable' rate allocation strategy, the rate vectors should be *irrotational*, i.e., be curl-free. Roughly, the flow directions should always move outward from the source and inward towards the sink, otherwise the flow direction is not 'effective' since a rotation would correspond to moving towards the source as well as getting further away from the sink.

Under this setup, it is straightforward to calculate $\mathbf{v}$. Since the field is irrotational, there exists a flow potential, i.e., a scalar potential function ψ such that

$$\mathbf{v} = -\nabla \psi$$

where ∇ is the gradient operator, i.e., *flow always moves in a direction which minimizes the associated potential of the system.* This refers to the 'effective' direction of a flow. It turns out that the solution to (3.8) is equivalent to solving an electrostatic problem with two linearly charged sources[40]. Thus ψ has the form

$$\psi(x, y) = c \, \log \frac{r_2(x, y)}{r_1(x, y)}$$

where c is a constant and r_1 and r_2 are the distances from the source and the sink respectively: see Fig. 3.5. Let us denote the velocity vectors associated with the source and sink by $\mathbf{v}_1$, $\mathbf{v}_2$, respectively. By the superposition principle, $\mathbf{v}$ is given by

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 = \frac{c}{r_1(x, y)} \mathbf{a}_1(x, y) - \frac{c}{r_2(x, y)} \mathbf{a}_2(x, y) \tag{3.9}$$

where $\mathbf{v}_1$ is derived from $-\nabla(\log r_1(x, y))$ and $\mathbf{a}_1(x, y)$ is the unit vector originating from the source and points to (x, y). $\mathbf{v}_2$ and $\mathbf{a}_2(x, y)$ are similarly defined.

Since a footprint is restricted to a finite region, in order to specify flow velocities within the footprint, one has to solve a boundary value problem, which may yield a fairly complicated solution. Instead, one may consider a region such that its boundaries fall along the streamlines that include the source, e.g., see Fig.

3.5. The solution within that region retains the same form with a proper scaling of the source strength such that the total flow emitted by the source is 1. Let us denote this closed region as Φ_0 corresponding to a hypothetical footprint. We can make Φ_0 'wider' by selecting a pair of streamlines that widens the footprint and adjusting the source strength, thus we effectively achieve an increased spreading. We can parameterize the degree of spreading by the mid section separation width of the boundary streamlines.

Note that from the solution (3.9) we see that, *the flow rate is of the order $1/r$ where r is the distance from either the source or the sink.* Let C be one of the equipotential contours. Then the total flow on this contour is given by a linear integral $\int_C ||\mathbf{v}||\, dl$ where $||\cdot||$ is the L^2 norm, since at any point in C, $\mathbf{v}$ is normal to the curve C. Thus integrating $||\mathbf{v}||$ over some closed subset in $\mathbb{R}^2$ results in the total (flow)$\times$(distance) in that region, which motivates us to associate the load distribution function $h(\cdot, \Phi)$ with $||\mathbf{v}||$.

Thus for a given footprint Φ_0, we have that

$$\mu = E[\int_{\Phi_0} ||v||dx], \tag{3.10}$$

$$\sigma^2 = E[\int_{\Phi_0} ||v||^2 dx]. \tag{3.11}$$

In fact Φ_0 has a deterministic shape except its random rotation which does not affect the above integrations. Thus one can integrate the flow velocity over Φ_0 to estimate these parameters.

Based on this we present numerical results where the session span is 20, $\epsilon = 1$ and c is properly adjusted so that the source flow into a footprint remains 1 for different values of w. Fig. 3.6 exhibits the distribution of flow rate $||v||$ over a footprint. Fig. 3.7 shows the plot of estimated μ and σ versus increasing spreading width based on numerical evaluation. As stated earlier, μ increases but σ decreases with w, and these parameters cannot be minimized simultaneously, which exhibits the fundamental tradeoff relationship between μ and σ.

Note within a continuum footprint a flow is well 'balanced' in a spatial manner since there are no obstacles that hinder flow within a footprint. Moreover footprints have nice and symmetric shapes. Surprisingly, we observe the similar scaling $1/r$ of flow rates in the regular grid model and associated optimal strategies which will be studied in the next section.
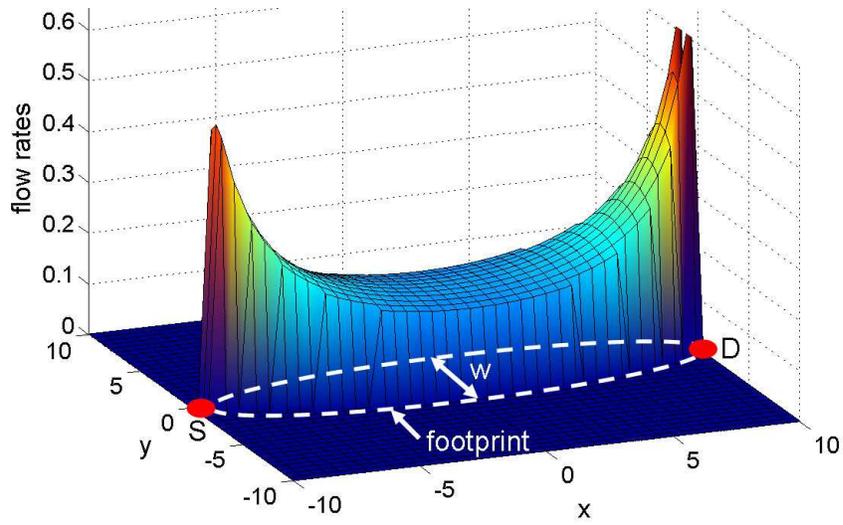
Figure 3.6: 3-D plot of flow rates over a footprint.
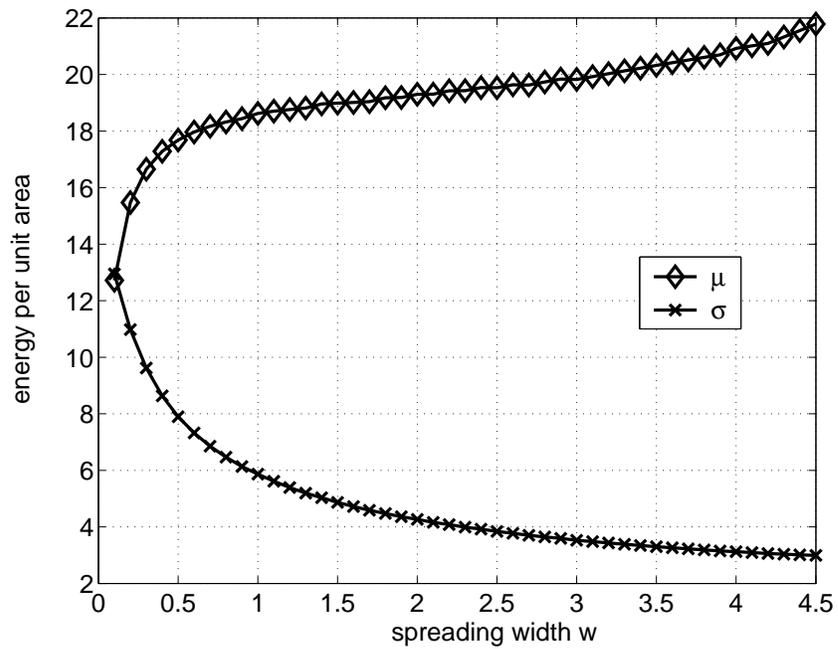
Figure 3.7: The plot of μ and σ over varying spreading width w under the continuum flow approach.
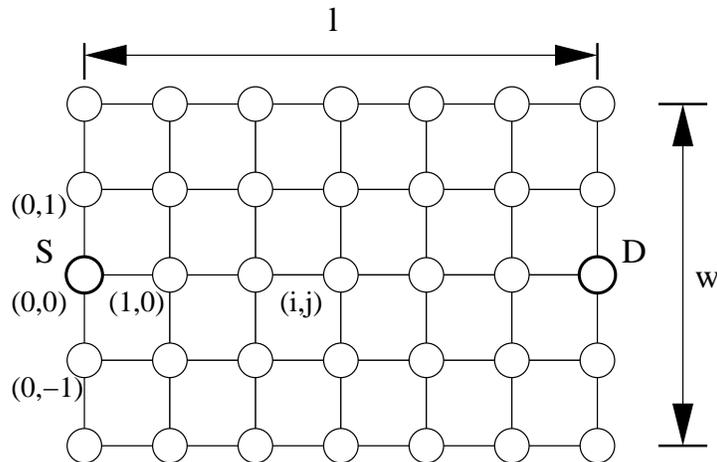
Figure 3.8: Topology of the regular grid footprint. The coordinates of locations are shown for some nodes in their lower left corners. The source and the destination is marked by S and D respectively and the dimensions, l and w, of the grid are shown.

3.5.2 Grid Approach

In this section we consider parameterized spreading strategies over a *regular grid*. Let us consider a session pair in a network. We assume that the intermediate relaying nodes form a regular grid topology[4], as shown in Fig. 3.8. The hop count between the session pair is fixed to l. The source emits 1 unit of flow which is distributed among the intermediate nodes to reach the destination, and a flow can be relayed only among adjacent nodes. A *grid footprint* is defined to be the set of grid nodes which carry nonzero flow for a given session pair. Accordingly, a *spreading strategy* corresponds to assigning the flow rates at the intermediate nodes such that the total flow is conserved. These are analogous to Φ_0 with fixed span and spreading width, and $h(\cdot, \Phi_0)$ respectively.

We refer to the maximum degree, in number of hops, to which the flow is spread as the *spreading width w*. Our main goal is to determine 'good' spreading strategy and grid footprint for given l and the constraint w. In particular, we will consider the *minimum-variance* strategy specified as follows.

---

[4]This serves as a coarse approximation for a dense, uniform network. Also note its similarity with the topology induced by PBM construction discussed in Section 3.3, e.g., those shown in Fig. 3.3.

Let us label each node with an integer coordinate (i, j) placing the source node at the origin as shown in Fig. 3.8. The total *outgoing* flow rate at node (i, j) is denoted by $e_{i,j}$ which will also represents the energy consumed at node (i, j). Define the set $H := \{(i, j) \mid 0 \le i \le l, \ |j| \le (w-1)/2\}$. Under the above construction, we have that

$$\mu = E\left[\int_{\Phi_0} h(x, \Phi_0) dx\right] \approx \sum_{(i,j) \in H} e_{i,j}, \qquad (3.12)$$

$$\sigma^2 = E\left[\int_{\Phi_0} h(x, \Phi_0)^2 dx\right] \approx \sum_{(i,j) \in H} e_{i,j}^2 \qquad (3.13)$$

where the continuum footprint Φ_0 corresponds to a grid footprint within H, i.e., $\{(i, j) \mid e_{i,j} > 0, \ (i, j) \in H\}$. Similarly, the load per unit area $h(\cdot, \Phi_0)$ corresponds to the flow rates $e_{i,j}$ associated with a grid footprint. We wish to minimize (3.13) by properly setting $e_{i,j}$ for given w and l.

The rationale behind the minimum variance strategy is that, in order to maximize $z_k(\tau)$, it is an efficient strategy to minimize σ assuming that k is moderately large, i.e., all the nodes have large energy reserves at the beginning so they would not suffer from depletion when only a few sessions overlap at a given location.

Before we formally define the flow distribution problem for the minimum variance, let us assume that l and w are even and odd integers respectively for symmetry purposes and $w \le l$, i.e., traffic is not spread beyond l. We also add an extra constraint that *the flow rates are equal along the middle abscissa (MA)*. The MA is defined to be the nodes that lie on the middle of the horizontal span of H, i.e., the nodes along $i = l/2$ axis in Fig. 3.8. We refer to this sub-optimal approach as *equi-flow minimum variance* strategy. The implications of the additional constraint will be explained in the sequel.

Problem 2. *Equi-flow minimum variance problem.*

$$\text{Minimize: } \sum_{i,j \in H} e_{i,j}^2, \tag{3.14}$$

$$\text{Constraints:}$$

$$\text{flow is conserved at each node,} \tag{3.15}$$

$$e_{i,j} \geq 0,$$

$$e_{0,0} = 1, \tag{3.16}$$

$$\max\{|j| \mid e_{i,j} > 0, \ (i,j) \in H\} = (w-1)/2, \tag{3.17}$$

$$e_{l/2,k} = \frac{1}{w}, \ (|k| \leq (w-1)/2), \tag{3.18}$$

$$\text{Variables: } e_{i,j}.$$

where (3.16) is the source emitting 1 unit of flow, (3.17) corresponds to the definition of the spreading width and (3.18) is the equal flow constraint at the MA.

The solution can be explicitly obtained by exploiting the convexity of the objective – see Appendix.

Lemma 3. *The solution to Problem 2 is, for all $(i,j) \in H$,*

$$e_{i,j} = \begin{cases} \{2(|i| + |j|) + 1\}^{-1}, & |i| + |j| \leq \frac{w-1}{2}, \\ \{2(|l-i| + |j|) + 1\}^{-1}, & |l-i| + |j| \leq \frac{w-1}{2}, \\ w^{-1}, & \text{otherwise.} \end{cases}$$

An example of such a flow assignment is illustrated in Fig. 3.9. The dotted lines are contours that represent level sets of nodes which have the same total flow. The value of the levels decreases harmonically, i.e., $(1, \frac{1}{3}, \frac{1}{5}, \ldots)$ as the contour expands outward from the source. Note its similarity with the $1/r$ flow rate scaling order in continuum footprint case introduced in the previous section. Note $e_{i,j}$ is the sum of horizontal and vertical outgoing flows at location (i,j) in the figure.

We comment on the additional constraint. Without (3.18), the minimum variance flow allocation can be obtained using standard optimization techniques such as the projected gradient method [41]. However with (3.18), the problem yields a closed-form, simple and intuitive solution. Moreover, it gives an excellent approximation especially for large values of l. In fact, one can show that the solution
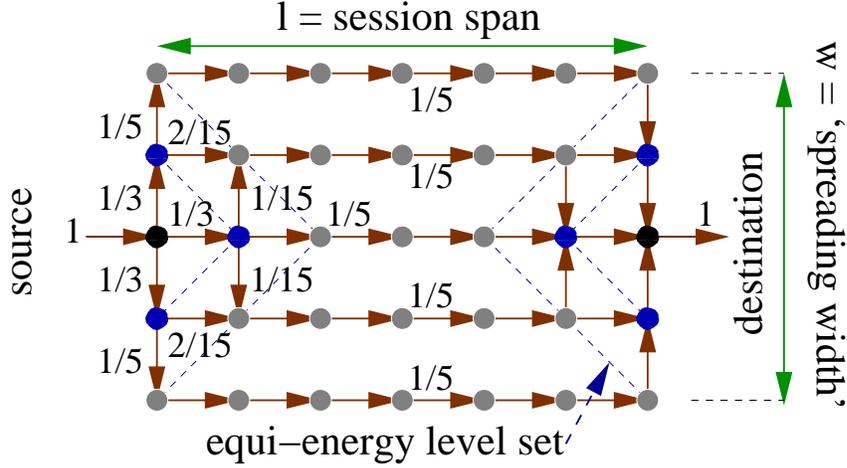
Figure 3.9: The optimal flow allocation when $w = 5$ and $l = 7$.

obtained for the suboptimal scheme converges to the optimal solution when l is large, which will be shown shortly.

Given the flow assignment in Lemma 3, we have that

$$\mu \approx \sum_{(i,j)\in H} e_{i,j} = l + \frac{w}{2} - \frac{1}{2w}, \tag{3.19}$$

$$\sigma^2 \approx \sum_{(i,j)\in H} e_{i,j}^2 = \frac{l}{w} - \frac{1}{2}\left(1 + \frac{1}{w^2}\right) + \sum_{k=1}^{\frac{w+1}{2}} \frac{2}{2k-1}. \tag{3.20}$$

The following observations are in order. The grid footprint suggested by the solution is such that it corresponds to H, i.e., $e_{i,j} > 0$ for all $(i,j) \in H$, which in turn means that the grid footprint for the equi-flow minimum variance strategy has a rectangular shape. Recall that we imposed equal flow constraints on MA parameterized by its width w. However our solution indicates that the flow should be spread as much as possible within H as long as the restrictions on flow directions are not violated.

For $1 \leq w \leq l$, we see that μ increases with w, i.e., the mean energy increases with the spreading width, indeed the flow will travel longer distances. However, σ decreases with w, i.e., the variance decreases with the degree of spreading. Thus, as mentioned earlier, one cannot minimize the mean and the variance simultaneously.
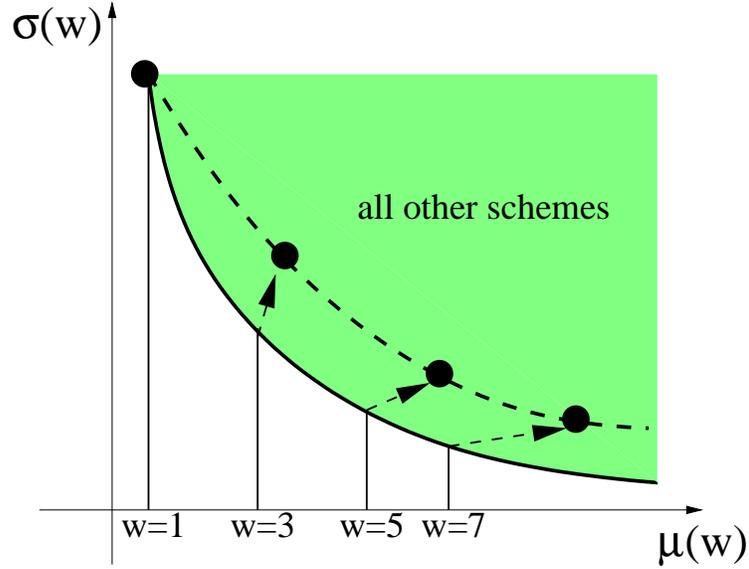
Figure 3.10: Pareto optimality of the parameterized μ and σ.

The interesting case is where l is much larger than w. Then the mean energy is roughly invariant to small changes in w, but the variance is sensitive to w – the dominant term is $\frac{l}{w}$. Thus for a load that traverses a long route, the benefit in terms of reducing variance from spreading is large.

One might ask whether the set of policies parameterized by w are close to Pareto-optimal, that is, given a mean achieved energy cost, they achieve a minimal variance, or vice versa. As shown in Fig. 3.10, our parameterized set of spreading strategies need not be Pareto-optimal. The next result shows that our proposed family of strategies is indeed Pareto-optimal as the session spans l becomes large.

Theorem 11. *Let us denote the sum of flow squared by equi-flow minimum variance scheme as S and that of the minimum variance scheme as S^*. Then we have that*

$$\lim_{l \to \infty} \frac{S^*}{S} = 1.$$

Proof. Let us consider only the left half portion of a footprint. Also let us define the set $A \subset H$ such that $A = \{(i, j) \mid |i| + |j| \leq l - 1\}$. Using the same notations as

above, for given l, S is given by

$$S = \sum_{(i,j) \in A} e_{i,j}^2 + \sum_{(i,j) \in H/A} e_{i,j}^2.$$

Let us denote the flow allocation at (i, j) of the optimal scheme as $e_{i,j}^*$, then

$$S^* = \sum_{(i,j) \in A} (e_{i,j}^*)^2 + \sum_{(i,j) \in H/A} (e_{i,j}^*)^2.$$

However, it is always true that on set A, the proposed scheme has the minimum sum squared, e.g., see the proof of Lemma 3. Thus

$$
\begin{aligned}
S - \sum_{(i,j) \in H/A} e_{i,j}^2 &= \sum_{(i,j) \in A} e_{i,j}^2 \leq \sum_{(i,j) \in A} (e_{i,j}^*)^2 \\
&= S^* - \sum_{(i,j) \in H/A} (e_{i,j}^*)^2.
\end{aligned}
$$

Thus

$$\frac{S^*}{S} \geq 1 + \frac{\sum_{(i,j) \in H/A}(e_{i,j}^*)^2 - \sum_{(i,j) \in H/A} e_{i,j}^2}{S}.$$

As $l \to \infty$, S also tends to infinity, however the numerator of the right term in the above inequality is always finite, thus $\lim_{l \to \infty} \frac{S^*}{S} \geq 1$. However since S^* is the minimum sum of flow-squared, it is always true that $\frac{S^*}{S} \leq 1$. Thus we conclude that $\lim_{l \to \infty} \frac{S^*}{S} = 1$. $\qquad\square$

3.6 Design Tradeoffs: Networks without Energy Replenishing Capability

3.6.1 Depletion probability of the typical node

In this section we numerically evaluate the depletion probability of a typical node, combining the estimates in (3.19) and (3.20) with $z_k(\tau)$. Here τ is assumed to be 1 and l is set to 20. Fig. 3.11 exhibits a plot of $\phi(z_k(\tau))$ for varying k, i.e., varying the initial energy reserves of network nodes versus the spreading factor w. Clearly, there exists a w that minimizes the depletion probability for each k. As expected, for the case where nodes have high initial energy reserves, the optimal w gets larger. The intuition is that, whenever the nodes in the network have large residual reserves, they should cooperate to balance load on the network, i.e., the number of nodes participating in carrying a flow should increase, but up to a degree where the energy cost of load balancing does not overload the network.

60
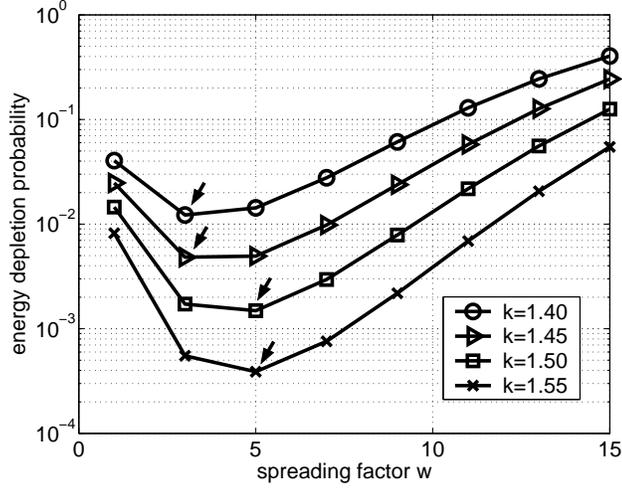
Figure 3.11: A numerical evaluation of optimal design of spreading parameter by the regular grid spreading strategy discussed in Section 3.5. The hop length l is fixed to 20 and the initial energy reserve parameter k is varied. Note the change in the tradeoff points marked by arrows, i.e, the optimal w moves from 3 to 5 with increasing k.

3.6.2 Depletion probability for a network

One can also approximate the spatial energy burden pattern as a stationary, isotropic Gaussian random field in $\mathbb{R}^2$. Consider $G(x,t)$ in (3.1) for $x \in \mathbb{R}^2$. Let us model a network as occupying a 'nice' subset A in $\mathbb{R}^2$, e.g., a rectangle or circle [42]. Consider the probability that the node with the highest energy burden in A exceeds a prescribed level b by some time τ, i.e.,

$$P(\sup_{x \in A} G(x, \tau) > b). \tag{3.21}$$

We can estimate the asymptotic value of this probability as $b \to \infty$ via extreme value theory for homogeneous Gaussian fields, see [43], [42]. Consider the normalized energy burden $Z(x, \tau) := (G(x, \tau) - \tau\mu)/\sqrt{\tau}\sigma$ where $\lambda = 1$. Let us define the normalized spatial covariance function $r_\tau(y)$ at time τ by $r_\tau(y) = E[Z(x + y, \tau)Z(x, \tau)]$ for $y \in \mathbb{R}^2$. Since the field is isotropic, this function depends only on the norm of y, denoted by $|y|$. Suppose that the following holds for some positive constant a,

$$r_\tau(y) \approx 1 - a|y|^\alpha \quad \text{as } |y| \to 0.$$

61
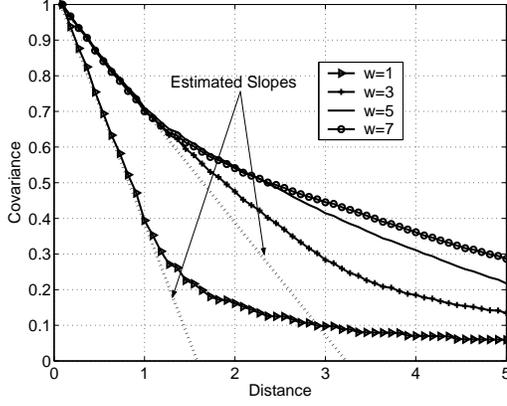
Figure 3.12: A numerical evaluation of $r_1(\cdot)$ as the function of distance. The dotted lines are the linear approximations for distances close to 0 where the slopes corresponds to a for each w.
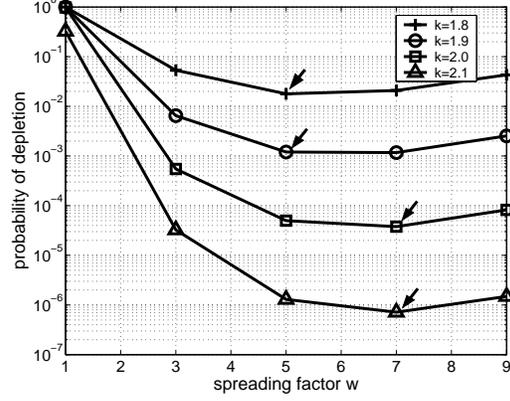
Figure 3.13: Estimate of the probability of depletion of the node with the highest burden varying k over the region A whose area is 100. The optimal choice of w moves from 5 to 7: see the arrows.

Here α denotes the infinitesimal order of decay of the covariance with the magnitude $|y|$. Again assume that b is given by $k\underline{\mu}\tau$ where k is large, then based on the Poisson clumping heuristic [42], we can rewrite and approximate (3.21) as

$$P(\sup_{x \in A} Z(x,\tau) > z_k(\tau))$$
$$\approx H_\alpha |A| a^{2/\alpha} \{z_k(\tau)\}^{4/\alpha} \phi(z_k(\tau)) \qquad (3.22)$$

where $|A|$ is the area of the region A and $H_\alpha > 0$ is the 2-dimensional Pickand's constant which depends only on α. We see that the depletion probability is proportional to the physical area of the network, and is related to the covariance structure of the footprints. Comparing this with the result for the typical node, we note that they share the term $\phi(z_k(\tau))$, but there is the extra term $\{z_k(\tau)\}^{4/\alpha}$ which may increase the depletion probability for large values of $z_k(\tau)$.

In order to estimate a and α, we present a plot of covariance $r_\tau(y)$ as a function of distance $|y|$ in Fig. 3.12 based on numerical evaluations. Here we fix $l = 20$, $\tau = 1$ and let w take the values of $1, 3, 5$ and 7. We observe that, except for the case where $w = 1$, the decay rates for small $|y|$ are virtually the same. We also see for small $|y|$, the covariances decay linearly, i.e., we can fix $\alpha = 1$, which in

turn gives the Pickand's constant by H_1. The slope a remains the approximately same value for $w = 3, 5$ and 7. For $w = 1$, the estimated slope is $a = 0.650$, and for $w = 3, 5$, and 7, $a = 0.324$.

Thus we expect that, by increasing w, we will observe a similar tradeoff curve as obtained for the typical node (Fig. 3.11), but the curve will be *flatter* due to the term $\{z_k(\tau)\}^{4/\alpha}$, which is shown in Fig. 3.13.

3.6.3 Optimal choice of w when sessions have different spans

A natural question arises: is it beneficial to adapt the spreading width w when the distance l between session pairs varies? An intuitive answer is that, if the span of a session is high, it helps to spread more, i.e., choose larger w. The rationale is that the sessions spanning longer distances have an increased chance of overlapping with other sessions. Thus it is more important to reduce the 'hotspots' induced by such sessions. Let L be a random variable whose distribution is that of the span for the typical session pair. Thus one would expect the optimal spreading width w to be a *nondecreasing* function of L. Let us denote this function by $g(\cdot)$. Using our model, we can determine $g(\cdot)$ approximately.

Considering only the dominant terms in (3.19) and (3.20), we will use the following approximations

$$\mu \simeq l + \frac{w}{2}, \ \sigma^2 \simeq \frac{l}{w} \tag{3.23}$$

for the case where l is substantially greater than w. With the above approximations and setting $\tau = 1$, we have that

$$z_k(1) \simeq \frac{kE[L] - E[L + g(L)/2]}{\sqrt{E[L/g(L)]}}.$$

The goal is to find an optimal choice for $g(\cdot)$, i.e., one that maximizes $z_k(1)$, i.e., minimizes the depletion probability of a typical node.

Assume that L is continuously[5] distributed over a finite support $S = [l_{min}, l_{max}]$ and the probability density is nonzero and smooth on that support. Let us partition

---

[5]In fact, L takes integer values. However, our arguments yield more intuitive results and can be trivially extended to such discrete random variable cases.

A into n subintervals $A_i := [l_i, l_{i+1}], 0 \le i \le n-1$, where $l_{min} = l_0 < l_1 < l_2 < \ldots < l_n = l_{max}$. Suppose $g(\cdot)$ is stepwise constant over each interval A_i, i.e., $g(x) = w_i$ for $x \in A_i$. We would like to find an optimal allocation $\{w_i\}$. Now $z_k(1)$ is a function of the vector $\mathbf{w} = (w_0, w_1, \ldots, w_{n-1})$ given by

$$z_k(1) \simeq \frac{kE[L] - E[L + g(L)/2]}{\sqrt{E[L/g(L)]}} = \frac{(k-1)E[L] - \frac{1}{2}\sum_i p_i w_i}{\sqrt{\sum_i p_i E[L|A_i]/w_i}}$$

where $p_i = P(L \in A_i)$ and $E[L|A_i]$ is the expectation of L conditioned on that $L \in A_i$. We first show that $g(\cdot)$ is indeed a nondecreasing function of L.

Lemma 4. $g(\cdot)$ *is a non-decreasing function of L irrespective of its distribution.*

Proof. Let us denote the maximizer of $z_k(1)$ as $\mathbf{w}^* = \{w_i^*\}$. It is easy to see that there exists the unique $\mathbf{w}^*$ which solves $\frac{\partial z_k(1)}{\partial w_i} = 0$ for every i. Using the first order necessary condition for the optimality, after rearrangement, we have that

$$w_i^* = C\sqrt{E[L|A_i]} \tag{3.24}$$

for all i where the constant C is given by

$$C = \frac{\frac{2}{3}(k-1)E[L]}{\sum_i p_i \sqrt{E[L|A_i]}}.$$

However $E[L|A_i] \le E[L|A_{i+1}]$ for all $0 \le i \le n-1$ irrespective of the distribution of L, so $w_i \le w_{i+1}$ for all $0 \le i \le n-1$. Thus $g(\cdot)$ is a non-decreasing function. $\square$

It is now clear how the spreading width should scale with the span of a session. If we decrease the interval lengths of A_i to infinitesimal ones, then $E[L|A_i] \simeq l_i$, and thus from (3.24) we have that

$$w_i^* \sim \sqrt{l_i}.$$

This implies that the optimal spreading width scales as the square root of session spans.

3.6.4 Optimal choice of w for sessions with varying loads

We can also consider the impact of the *variability in traffic load* of sessions. Let the distribution of variable loads be that of a random variable U which has finite support. Thus the spreading width can be a function of both length and load, i.e., $W = g(L, U)$. Using the similar approximations as (3.23), we have that

$$z_k(1) \simeq \frac{kE[L]E[U] - E[U(L + g(L,U)/2)]}{\sqrt{E[U^2 L/g(L,U)]}}$$

where we have assumed that L and U are mutually independent for a typical session. Taking a similar approach as earlier, we let $g(L, U) = w_{i,j}$ if $L \in A_i$ and $U \in B_j$ where $B_j := [u_j, u_{j+1}]$, $0 \leq j \leq m - 1$, are the m partitions of the support of U for some positive integer m in a similar manner to that of A_i. Then we have that

$$z_k(1) = \frac{kE[L]E[U] - \sum_{i,j} p_i q_j E[U|B_j]\{E[L|A_i] + \frac{1}{2}w_{ij}\}}{\sqrt{\sum_{i,j} p_i q_j E[L|A_i]E[U|B_j]w_{ij}^{-1}}}$$

where $q_j = P(U \in B_j)$. If we apply the first order necessary condition for the optimality, we obtain

$$w_{ij}^* = C\sqrt{\frac{E[U^2|B_j]E[L|A_i]}{E[U|B_j]}} \tag{3.25}$$

where C is given by

$$C = \frac{\frac{2}{3}E[L]E[U](k-1)}{\sum_{i,j} p_i q_j \sqrt{E[U^2|B_j]E[U|B_j]E[L|A_i]}}.$$

Again we find that w_{ij}^* is an increasing function of both U and L irrespective of their distributions from (3.25), using the following lemma.

Lemma 5. *Let U have a positive probability density function over the partitions B_j. Then $\frac{E[U^2|B_j]}{E[U|B_j]}$ is a nondecreasing function of j.*

Proof. Let $f_U(\cdot)$ be the density function of U. We have that

$$\frac{E[U^2|B_j]}{E[U|B_j]} = \frac{\int_{B_j} u^2 f_U(u)du}{\int_{B_j} u f_U(u)du}$$

$$\leq \frac{\int_{B_j} \sup\{v : v \in B_j\} u f_U(u)du}{\int_{B_j} u f_U(u)du}$$

$$= \sup\{v : v \in B_j\}.$$

However, for any $i > j$, we have that

$$
\begin{aligned}
\frac{E[U^2|B_i]}{E[U|B_i]} &= \frac{\int_{B_i} u^2 f_U(u) du}{\int_{B_i} u f_U(u) du} \\
&\geq \frac{\int_{B_i} \inf\{v : v \in B_i\} u f_U(u) du}{\int_{B_i} u f_U(u) du} \\
&= \inf\{v : v \in B_i\}.
\end{aligned}
$$

By definition, $\inf\{v : v \in B_i\} \geq \sup\{v : v \in B_j\}$ and the result follows. $\qquad\square$

By taking the limits $p_i, q_j \to 0$ and by (3.25), we have the scaling rule of $w_{ij}^* \sim \sqrt{l_i u_j}$, i.e., *the optimal spreading factor for a given session approximately follows the square root of the bits·meters of its offered load.* We will verify by simulation that indeed such dynamic spreading schemes outperform those with fixed w.

3.7   Design Tradeoffs: Networks with Energy Replenishing Capability

Next we consider the case where nodes have the capability to replenish their energy at constant rate of c units per unit time and their energy storage capacity is b. We model the energy level of a node by a queue where arrivals correspond to new energy burdens to be served, i.e., replenished at rate c. Note that the dynamics of the queue and their physical interpretation are reversed: 'filling' the queue with energy burden corresponds to 'consuming' its energy reserves. Thus, we are interested in the likelihood that the queue length exceeds the level b.

For a typical node which is covered by multiple session footprints over time, the energy load burden for each footprint would be 'buffered' in the node's energy queue which is replenished at rate c. In reality, an energy request fills the queue, i.e., consumes energy at a roughly constant rate, which can be modelled by using a continuous load model, e.g., fluid queues (see [44]). For simplicity we will assume that energy burdens are imposed *instantaneously* on nodes and the offered load at a typical node depends only on its location within the footprints that 'hit' the location. We will again assume the footprint arrival process is a homogeneous Poisson process in time and space. With these assumptions we will use a discrete-time queueing model that approximates the $M/GI/1$ queue corresponding to these dynamics of energy burden at a typical node.

In this regime, we study the *asymptotic decay rate* of the queue content as an indicator of the probability that the energy burdens exceed a large initial energy reserve of b. For a stable, single-server queue, we denote the steady-state workload by W. If the following condition is satisfied for some $\theta^* > 0$:

$$b^{-1}\mathrm{log}P(W > b) \stackrel{b \to \infty}{\Rightarrow} -\theta^*,$$

then we refer to θ^* as its asymptotic decay rate [45]. We will use the results in [46], [45] to describe the behavior of the tail probabilities.

Let us define the problem. The energy burdens of each footprint are assumed to be i.i.d. with a distribution that is not heavy-tailed. We denote the virtual workload for the energy queue associated with a typical location in time slot $(i, i+1]$ for $i \in \mathbb{Z}$ by W_i . Then we have that

$$W_{i+1} = \max[W_i + X_{i+1}, 0] = [W_i + X_{i+1}]^+,$$

where $X_i = S_i - c$ and S_i is the total energy burden per unit time slot, and c is the replenished energy per time slot. These dynamics correspond to a Lindley process, and since $\{X_i\}$ are i.i.d., we can readily apply the following results on the decay rate function.

Theorem 12. *[45] Let us assume $\{W_i\}$ is stationary and thus stable under condition $E[X_i] < 0$, i.e., $E[S_i] - c < 0$. If $\{X_i\}$ are i.i.d., then θ^* satisfies*

$$\rho(\theta^*) = 0, \ \frac{d}{d\theta}\rho(\theta^*) > 0$$

where $\rho(\theta) = logE[e^{\theta X_i}] = logE[e^{\theta S_i}] - c\theta$.

We can readily obtain the required cumulant generating function of energy burden per time slot as follows.

Theorem 13. *The cumulant generating function $C(\theta)$ of S_i is given by*

$$C(\theta) = logE[e^{\theta S_i}] = \lambda E[\int_{\Phi_0} \{e^{\theta h(x, \Phi_0)} - 1\}dx].$$

Hence we have that the rate decay function $\rho(\theta)$ is given by

$$\rho(\theta) = \lambda E[\int_{\Phi_0} \{e^{\theta h(x, \Phi_0)} - 1\}dx] - c\theta$$

67

under the stability condition

$$\lambda E[\int_{\Phi_0} h(x, \Phi_0)dx] < c. \tag{3.26}$$

Proof. See appendix. □

The stability condition relates the replenishing rate c, and the rate of new energy burden requests per unit area, λ, times the average total energy per footprint. The root θ^* of $\rho(\theta) = 0$ may be found numerically. Using (3.19) and (3.20), several decay rates with varying spreading widths are given in Table 3.1. Here $l = 8$ and $\lambda = 1$, and let us denote the critical replenishing rate to satisfy the stability condition when $w = 7$ as c^*. The replenishing rate c is set to βc^* where $\beta = 1.2$ and 2.0.

Again we observe tradeoffs associated with different replenishing rates. When $\beta = 1.2$, the optimal spreading width is 3, but with a higher replenishing rate $\beta = 2.0$, the optimal w increases to 5. The intuition is that, with higher replenishing rates, one can spread traffic further to get more benefits from spatial balancing. However, if the replenishing rate decreases close to the critical value, the mean energy cost to spread is no longer negligible so that a smaller spreading width is preferred.

Table 3.1: Decay rates with varying spreading widths

Spreading width	Decay rate θ^*	
w	$\beta = 1.2$	$\beta = 2.0$
1	0.8673	1.7125
3	**1.2506**	2.7080
5	1.0965	**2.7593**
7	0.7965	2.6831

3.8   Simulations

3.8.1   Basic setup

In this section we simulate several scenarios to further explore the benefits of proactive spreading. The performance metric will be the probability that

a randomly selected node is depleted after a fixed time given a maximum energy reserve (MER). This metric is of fundamental interest from an engineering perspective, when given a network operation time and a MER, we wish to minimize the probability of the typical node is depleted, or equivalently, the fraction of depleted nodes in the network.

A total of 400 node locations were generated according to an uniform distribution on a 20×20 unit square area. Session arrivals are homogeneous in space, and a total of 200 sessions are generated at each simulation run. This is repeated for 500 times to obtain an averaged energy profile. Unless otherwise specified, each session offers 1 unit of load per unit time with a holding time of 1 unit time. We simulate session arrivals by picking two nodes at random, which corresponds to a session pair and then setting up a unidirectional flow. We set up multipath routes based on the PBM route construction introduced in Section 3.3, and the flow is equally divided on each path in order to approximate the scheme in Section 3.5. In our simulations, the 'shortest path routing' (SPR) is a routing that takes the minimum number of hops on the Delaunay graph of nodes. This must be distinguished from the shortest Delaunay routing (SDR) which is a PBM routing with a spreading width w of 1.

3.8.2   Scenarios

3.8.2.1   Nodes without replenishing capability

Fig. 3.14 shows the average energy depletion probability for several values of the spreading width w and SPR. A point (x, y) in this plot should be interpreted as follows: 'the probability that the energy expenditure of a typical node *will exceed* x is y'. If x is the MER then y is the probability that a typical node is depleted.

Let us consider only proactive routing first. When the MER is less than 20 units, routing with a minimal spreading width ($w = 1$) performs best. However, as the MER increases to more than 25 units, proactive multipath routing with the largest spreading width ($w = 5$) outperforms the others. These results are consistent with previous discussions, since with a high MER, a scheme with a lower variance in the energy expenditure ($w = 5$) is preferable at the cost of higher mean energy expenditure. These tradeoffs occur when the maximum reserve is between 20 and 25 units in our simulations. SPR has a lowest mean energy expenditure but the highest variance, and suffers from the worst performance in 'tail behavior', i.e., the
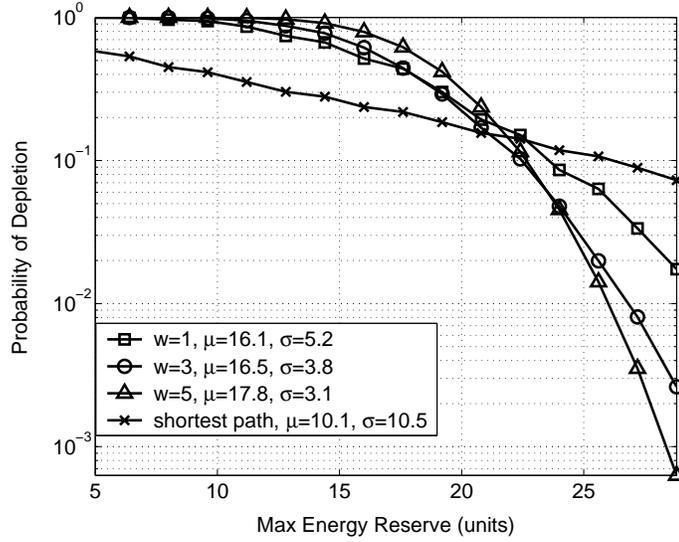
Figure 3.14: Energy depletion probability for nodes without energy replenishing capability. μ and σ represents the mean and the standard variation of energy expenditure of each scheme.
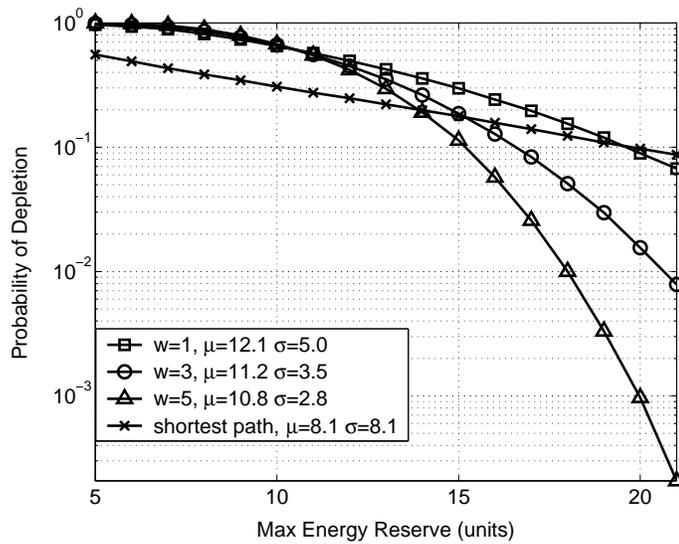
Figure 3.15: Energy depletion probability for nodes with energy replenishing capability.

lowest slope in the decay for the probability of depletion with the MER. Also note that it has different performance as compared to the SDR ($w = 1$) case: the SDR performs better due to its steeper slope in the tail probability. We verified that, for SDR, the shape of the empirical histogram of energy burden indeed resembles the Gaussian p.d.f., while that for the SPR is monotonically decreasing with a heavy tail – this will be discussed in detail later in this section.

3.8.2.2   Nodes with replenishing capability

Fig. 3.15 shows the energy depletion probabilities when the nodes have the capability of replenishing their energy reserves. At each simulation run a total of 200 sessions arrive uniformly on the time interval $[0, 200]$. Nodes have replenishing rate of 0.125 energy units per unit time. The benefit from proactive spreading is greater than that seen for the non-replenishing case. The intuition here is that, for larger w, the average number of nodes that participate in a session is greater than that of a scheme with less w. Thus more nodes have a chance to replenish their energy reserves, which results in a reduced mean *and* less variance in the energy expenditure (see μ and σ in the figure) with the largest spreading width, $w = 5$.

3.8.2.3   Dynamic spreading

Using the results from Section 3.6, we have simulated a scheme with dynamic spreading widths depending on the random load D and random session span L according to our scaling rule $\sqrt{LD}$ where nodes do not have energy replenishing capability. Fig. 3.16 shows the simulation results of such an dynamic spreading scheme. Here each session carries i.i.d. exponentially distributed load of mean 1. As shown in the figure, for small MER region (< 15), the dynamic scheme performs reasonably well but not best perhaps due to the error in rounding w off to an integer. However, it is superior to other schemes with fixed spreading widths as the MER increases.

3.8.2.4   Routing based on residual energy reserves

Next we consider a class of dynamic routing schemes and study how it benefits from proactive load balancing. Specifically we consider a routing scheme which exploits knowledge of the residual energy reserve at each node, i.e., routing with state
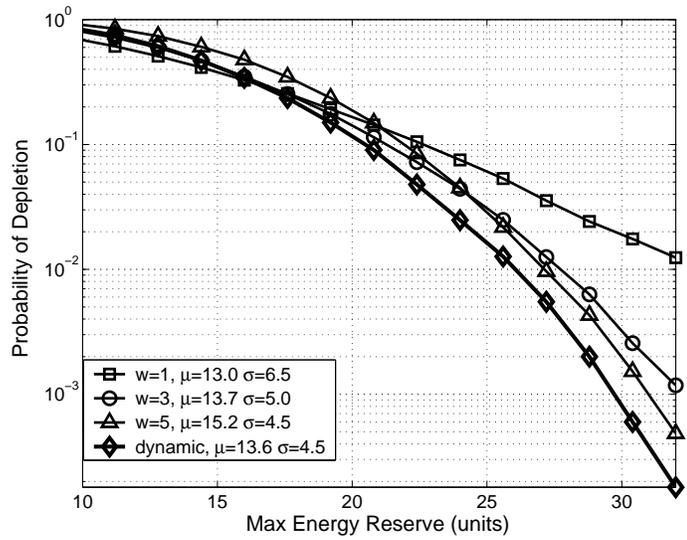
Figure 3.16: Energy depletion probability for the dynamic spreading scheme adjusted to session load and hop length.
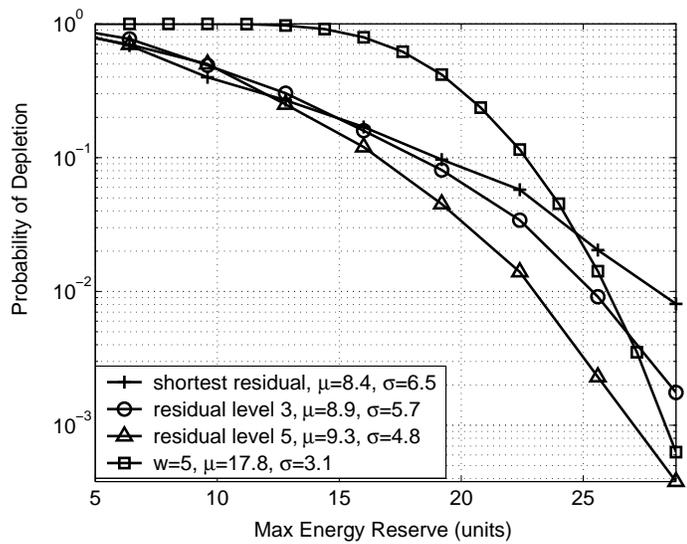
Figure 3.17: Comparison of the shortest residual routing, the level-3 and the level-5 residual routing.

information. We use Bellman-Ford algorithm for minimum cost routing, where the cost is a decreasing function of the fraction of the residual energy to its full reserve. In this way, a route through nodes having relatively high energy reserves might be preferred even if that route involves a higher number of hops. Specifically, if the residual energy of ith node is $b_i(t)$ at time t, the cost of routing traffic to that node is $(b_i(t)/b_{max})^{-\gamma}$ where γ is some positive constant and b_{max} is the maximum energy reserve at a node. Concerning the choice of γ, a related study [47] shows that a value within the range of $0.5 \sim 2.5$ is preferable, so we chose $\gamma = 1$. Also we have assumed that the routes do not change once created since such changes can incur a severe scalability problem due to the large number of nodes.

For these simulations, we define a level-w residual routing to be such that, the w best disjoint routes are chosen. Fig. 3.17 shows one of such comparison. We see that proactive spreading reduces the tail probability although the performance of state-dependent routing schemes is sensitive to the variability of traffic. In particular Fig. 3.17 exhibits the performance of proactive multipath routing with $w = 5$. The results show that, although it does not use dynamic state information, it may be adequate for a network whose nodes have high energy reserves.

3.8.2.5 Comparison of SDR and SPR

As shown in Fig. 3.14, SDR performs better than SPR – consider the steeper slope of the tail probability. Recall that SDR is related to the shortest Euclidean path between nodes while SPR uses the path with shortest hop count. Since the spatial distribution of node is homogeneous, one would expect that the performance would be similar, however the empirical distributions of energy burdens exhibit substantially different behavior: see Fig. 3.18. The distribution of energy burdens achieved by SPR is shown to have heavy tail similar to pareto-type distributions, while that of SDR resembles the Gaussian p.d.f. as our previous analysis has predicted. A possible explanation is that these differences are related to the area of the Voronoi cell associated with each node. The top of Fig. 3.19 shows the energy profile of the nodes in the first simulation of this section sorted by the amount of energy burden. Meanwhile, the figure on the bottom shows the cell area also sorted by the energy burden. Both figures indeed indicate the positive correlation between the energy burden and cell areas as might be expected for both SDR and SPR. However by computing correlation coefficient ρ of the area profile and the energy
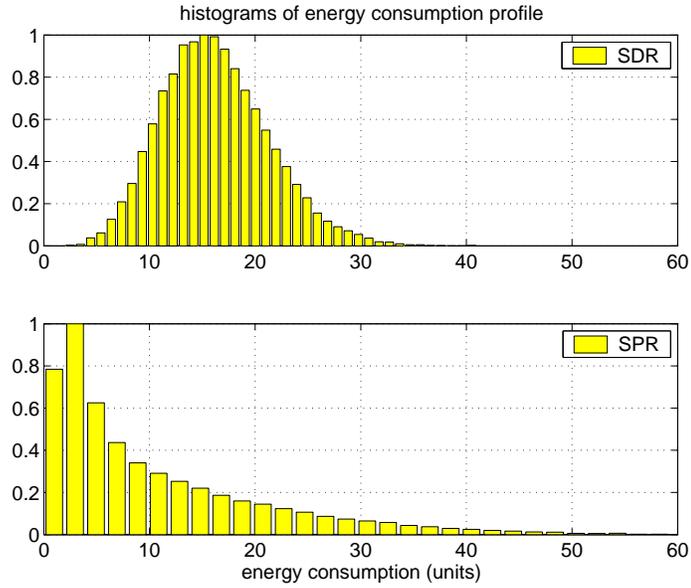
Figure 3.18: The empirical distributions of the energy burden profiles of SDR and SPR.
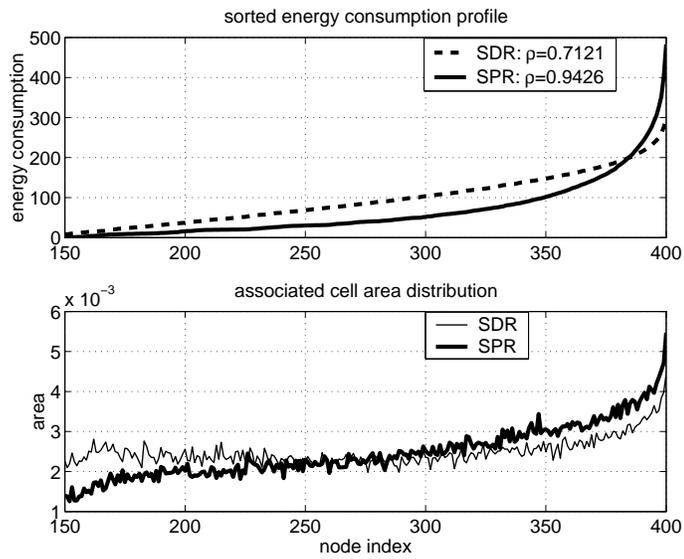
Figure 3.19: Nodes sorted by energy burden profile and their corresponding Voronoi cell area.

74

burden profile, we find that in the case of SPR one finds the stronger correlation $\rho = 0.9426$ compared to SDR where $\rho = 0.7121$. Considering that the load on the network is spatially homogeneous, we conclude that SDR routes are more likely to 'hit' the cells with larger areas. However this phenomenon is more prominent in SPR case, since big cells tend to reduce hop count dramatically, resulting in higher spatial congestion of load on those larger cells. Thus the large cells suffer from highly unbalanced energy burden, which may be the main cause of the heavy tail in the energy burden distribution. This is one of the aspects of the hop-count based strategy that may aggravate the imbalance of the system which can be mitigated by the routing strategy based on the spatial information, e.g., SDR.

3.9   Conclusion

In this chapter we propose a simple model for the spatial distribution of energy burdens in a multihop ad hoc wireless network. Our primary contribution is to use these models to investigate the design and potential benefits of proactive energy balancing multi-path routing schemes. To do so we develop a simple second order approximation permitting one to investigate tradeoffs of several types, e.g., for ad-hoc networks with or without replenishing and with energy storage capabilities. The essential tradeoff is between the mean and variance of a spatial energy (flow) balancing scheme. For our proposed models one might attempt to identify Pareto optimal energy balancing strategies, e.g., one minimizing the variance subject to a mean energy constraint, or conversely one minimizing the mean energy burden subject to a variance constraint. To simplify matters we consider flow/energy balancing on regular grid model for a simple parameterized family of spreading schemes. This permits us to concretely evaluate how this tradeoff should be optimized for the various network types and possible design criteria. The results are insightful but perhaps not unexpected. For networks with increased energy storage and/or replenishing capabilities it pays to be more aggressive in spreading traffic so as to reduce the variance in the energy burden since the additional energy burden can be smoothed by energy reserves or new energy sources – one must however ensure that the energy burden does not exceed the replenishing capability. For the most part our simulations confirm our analytical results and permitted us to evaluate more general regimes of interest.

We note however that the traffic patterns and network geometry used in

75

our simulations are fairly benign in that they are fairly homogeneous in time and space. In practice, one would expect to see irregular topologies and imbalances and variability in traffic loads. These in turn would lead to additional variability in the energy burdens on the network. We expect that the benefits of proactive load balancing to be more prominent and sensitive to design in the presence of the aforementioned fluctuations. The degree of spreading, e.g., w, might advantageously be exploited to adaptively smooth out such spatial variabilities and achieve improved balancing of energy burdens coupled with improved performance on network lifetime. Indeed one of the key results in this thesis shows that w scales roughly as the square root of load times distance, which is well expected by the intuition that the more one should spread traffic when the spatial burden imposed on the network increases.

Chapter 4

Energy Burden Balancing and Data Fusion in Sensor Networks Detecting Bursty Events

4.1 Introduction

In the previous chapters we have shown that local data compression and proactive load balancing provide a substantial gain in the network lifetime. In this chapter we integrate these strategies and study sensor network scenarios, where such a combined approach is applicable. Particularly we endeavor to identify the impact of the spatial character of sensing event/phenomena on engineering network traffic. We begin by motivating the problem via discussion of some drawbacks of previous approaches and proposed potential solutions.

Consider a hierarchical organization of network resources based on local aggregation of sensor data followed by forwarding to a set of information sinks. The rationale is to have nodes in close proximity elect an aggregation node whose role might be to compress, filter, or perform data fusion on spatially correlated data prior to transport to an information sink.

Unfortunately, such a hierarchy faces an intrinsic problem from an energy perspective: the nodes close to sinks will see a disproportionate energy burden as they will see higher loads of traffic that flow to the sinks. One solution is to increase the density of sinks, however this may end up being too costly. In this chapter we will assume that only a small number of sinks are available relative to the size of network, and thus by large-scale we mean that on average there are a large number of sensors associated with each sink.

Another possibility is to make sinks *mobile*, i.e., have sinks change their locations to balance the energy burdens incurred across the network nodes [48]. As shown in Fig. 4.1, the network can operate in two stages. In Stage 1 local aggregation nodes (AGNs) may aggregate information from sensors in close proximity while sinks move around. Then, in Stage 2, the sinks may 'probe' the network at high
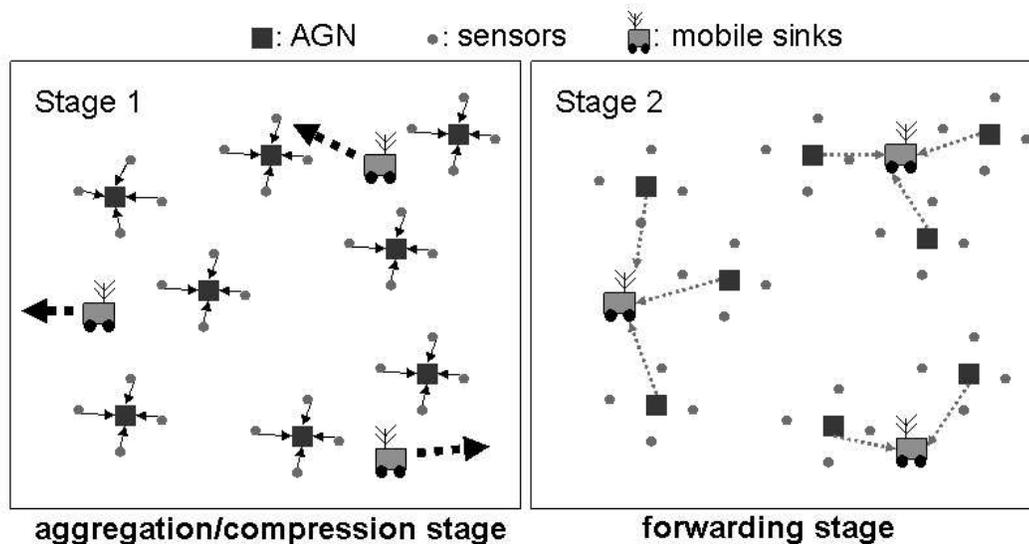
Figure 4.1: Stages of operation for sensor network with mobile sinks.

power, and collect information from AGNs. Not surprisingly the effectiveness of this scheme depends on a number of factors including the very nature of the information being gathered. Among those which we consider to be critical are: the timescale of sink mobility, the spatio-temporal periodicity of sensed data, delay sensitivity for data collection, and the character of the sensed phenomena.

Let us consider a few extreme examples. If *every* sensor generates data periodically on relatively short timescales versus that on which sinks move, and if this data must be relayed to sinks immediately, then the network will scale very poorly for two reasons: concentration of energy burdens and throughput collapse around sinks. In this case the only reasonable solution is to put more sinks. However if data delivery is delay insensitive AGNs may forward data only when a sink is close by i.e., in an *opportunistic* sense the sink mobility increases energy efficiency and throughput capacity [49].

By contrast let us consider application scenarios where the 'events' being sensed correspond to spatio-temporal bursts of information and can tolerate delays on the order of the timescales of sink mobility. We believe a number of interesting applications fall in this category, including surveillance and monitoring of bursty unknown, or poorly characterized, events. Fig.4.2 shows an example, where only
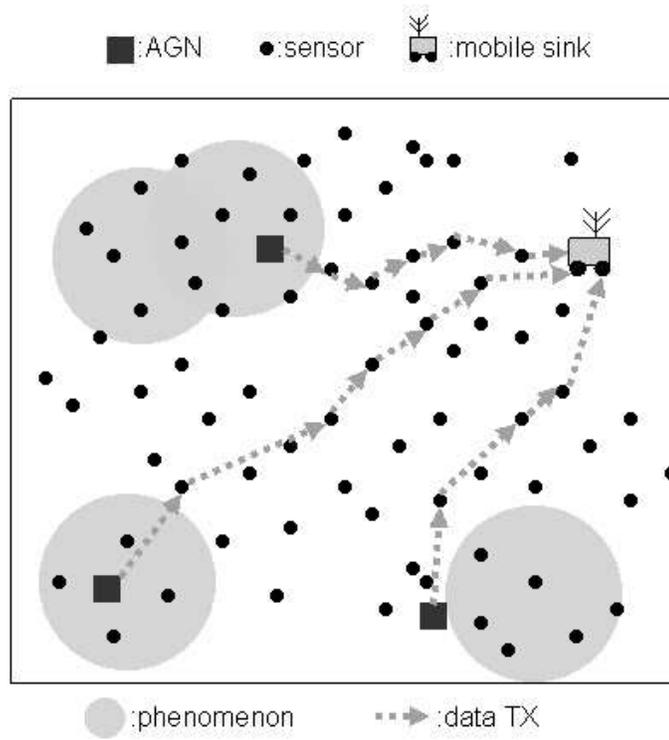
Figure 4.2: Bursty phenomena and traffic concentration.

sensors covered by the spatial phenomena generate data reports to nearby AGNs. In this case the network can significantly benefit from the load and energy balancing resulting from sink mobility. However the following problem remains: aggregated data, which has been successfully 'compressed', can still be substantial, and when forwarded on 'narrow' paths to the sinks, may incur substantial energy burdens on the associated nodes. Furthermore in a system with scarce sinks, these paths may involve a large number of nodes. As such, this may result in substantial inhomogeneities in the energy burdens the network will see eventually shortening its lifetime.

To avoid such concentrations of energy burdens it is reasonable to *spread* aggregated traffic forwarded to the sink in a *proactive* manner. In other words, have nodes cooperate to create multiple paths over which traffic is spread. The focus of this chapter is on modelling and analyzing stochastic geometric models permitting one to evaluate the tradeoffs associated with aggregation for purposes of 'compression' and then spreading of traffic forwarded to mobile sinks.

In particular we will explore the following tradeoffs. First by spreading traffic over several paths, one can certainly balance energy burdens but at the expense of having traffic traverse longer distances and thus larger average energy costs. Second by aggregating information from nodes one can reduce the traffic load through compression, filtering and/or data fusion, but this may lead to undesirable inhomogeneities of load and thus energy burdens from AGNs to sinks. The fundamental design questions are:

- *How much spreading is beneficial for traffic from the AGNs to the sinks?*

- *When is the benefit of aggregation (compression) of traffic counteracted by the resulting concentration of energy burdens in the network ?*

To capture the characteristics of the problem in this chapter we devise a spatial model for traffic aggregation and spreading. It is without a doubt a very simple caricature, based on a first-order model for energy and compression at AGNs, yet it allows us to study how the network lifetime is affected by a number of design parameters, including the effectiveness of aggregation/compression, the density of sinks, battery capacity of sensors, etc. We will show how one can further jointly optimize the spatial scales for aggregation and spreading so as to maximize the network lifetime and provide a numerical study of the results.

This chapter is organized as follows: in Section 4.2 we discuss related work. In Section 4.3 we briefly present our model and assumptions. In Section 4.4 we derive mean and variance of energy burdens, and based on those we discuss numerical results on optimizing operation to maximize the network's lifetime in Section 4.5. Finally we conclude with Section 4.6.

4.2 Related Work

The use of mobile relays (or base stations) in large sensor networks was proposed and studied in [48, 50, 51]. The idea was to deploy automated robots to solve scalability problems in sensor networks. Particularly in [48] the authors show how the energy concentration problem around sinks can be mitigated by using a single mobile sink, to smooth out fluctuations in energy burdens. In this chapter we tackle a more general problem including hierarchical aggregation towards multiple mobile sinks, and see how the performance can be optimized. In particular we develop and analyze a model which captures the impact that aggregation/compression, traffic spreading and moving sinks will have on the network's lifetime.

In this respect the models and analysis presented in this chapter are closely related to Chapter 3 where we modelled random, unstructured traffic in a homogeneous network as inducing *spatial traces* of energy burdens and showed how proactive multipath routing could be used to extend the network lifetime. We will leverage these results in this work, in modelling the spatial patterns of structured, i.e., hierarchical aggregated traffic.

4.3 Models and Assumptions

4.3.1 Reducing Traffic Loads Via Local Compression, Filtering and/or Data Fusion

Our model is geared at capturing the sensing of physical phenomena which are bursty, unpredictable but spatially clustered. In this context large scale distributed compression is not likely to be effective, yet local aggregation combined with compression, filtering or data fusion can still be carried out to reduce traffic loads.

We will adopt a simple model for such reductions in traffic. Assume a sensor

generates a packet of unit size and let $f(n)$ denote the amount of data after the aggregation of n packets. We assume $f(n)$ has the following properties:

1. $f(n) \leq n$,

2. $\lim_{n \to \infty} \frac{f(n)}{n} = \alpha$, $0 < \alpha < 1$,

3. $\frac{f(n)}{n}$ is a monotonically nonincreasing function of n.

The rationale for the first assumption is clear. The second assumption captures an asymptotic ratio for the possible reductions in traffic. The parameter α represents the efficiency of the aggregator in reducing traffic, where a lower α corresponds to better performance. The third assumption states that if an AGN acquires data from more sensors, the efficiency does not get worse. These assumption may apply to generic classes of sensing applications which use compression, filtering and/or data fusion.

Specifically consider the case where an AGN simply performs local 'compression' without any side information. The AGN might for example simply compress the sensed data based on a universal compression scheme such as Lempel-Ziv coding, assuming no a priori knowledge, versus a distributed compression such as Slepian-Wolf coding [52]. For example, suppose each sensor i generates discrete (possibly correlated) random variable X_i (with entropy at most 1) corresponding to a samples of a stationary process then $f(n) = H(X_1, X_2, \ldots, X_n)$, will satisfy the above assumptions; specifically the second one which is equivalent to the existence of an entropy rate for stationary processes

$$\lim_{n \to \infty} \frac{H(X_1, X_2, \ldots, X_n)}{n} = \alpha.$$

Also it is known that $\frac{H(X_1, X_2, \ldots, X_n)}{n}$ is a nonincreasing function of n, i.e., monotonicity of entropy per element [9] for stationary stochastic process $\{X_i\}$, which is our third assumption.

For simplicity we will define $f(n)$ as follows:

$$f(n) = \begin{cases} 0 & n = 0, \\ \alpha(n-1) + 1, & n \geq 1. \end{cases} \qquad (4.1)$$

Note we let $f(1) = 1$, i.e., there is no fusion gain for a single sample, then for $n \geq 2$, $f(n)$ grows linearly in n with slope α. Thus this functional form satisfies all the above properties. Also for simplicity we assume that aggregated data whose initial size was n is *deterministically* reduced to $f(n)$.

Our intent is to capture a fairly general concept of 'compression', whether it results from data compression, filtering or data fusion, in the sequel we refer to such reductions as compression. For example consider a sensor network in which 5 monitoring sensors see the same event, and report it to their associated AGN. In this case the AGN may choose to filter the data, and send a single warning to a sink, achieving a 5:1 compression ratio.

In summary, our model includes local aggregation, where compression filtering and/or data fusion can take place, with subsequent forwarding to mobile sinks. Overall this is a complex, highly simplified model, yet the key insights we develop appear to be telling of which elements one should consider in engineering such sensor networks.

4.3.2 Depletion Probability

We will use a simplified first order model for energy expenditures associated with data transmissions where energy burdens are proportional to the traffic incident on a node. As mentioned earlier, we assume that mobile sinks periodically coordinate to 'probe' the network, i.e., broadcast their locations and announce they are ready to serve as data sinks. If some AGNs have data to send they forward the data to the nearest sink as depicted in Fig. 4.1 - we refer to this two-stage operation as a *round*. The sinks navigate space at random and we assume the spatio-temporal load on the network is relatively light so that at each round AGNs see sinks at a set of random locations which are effectively independent of past rounds. We also assume that AGNs are selected based on proximity, i.e., once selected the sensors associate themselves with the closest AGN.

At each round the energy burden at a typical sensor is incurred by data relaying/forwarding from AGNs to the nearest sink locations which are independent across rounds. Also we assume the burdens are stochastically homogeneous in space. If these independent burdens are accumulated over a sufficient number of rounds, we model such burdens as a *shot-noise process* and we may apply central limit theorem (CLT) [39] for the energy depletion probability.

Under a given strategy S suppose one is able to determine the mean and the standard deviation of energy burdens *per round* at the typical node. Let us denote them by $\mu(S)$ and $\sigma(S)$ respectively. Suppose one would like to operate the network for m rounds, i.e., specifies the required operational lifetime of the network. Also suppose every node has the identical maximum energy reserve denoted by b which is parameterized by m, specifically b is some multiple k of m, i.e., $b = km$ for convenience. If we denote the energy burdens at a typical node by Z_m at mth round, for m sufficiently large, the depletion probability can be approximated as follows:

$$P(Z_m > b) \approx \phi\left(z_k(S)\right)$$

where we define the following:

$$z_k(S) := \sqrt{m}\left\{\frac{k - \mu(S)}{\sigma(S)}\right\}, \tag{4.2}$$

$$\phi(u) := \frac{1}{\sqrt{2\pi}}\int_u^\infty e^{-v^2/2}dv. \tag{4.3}$$

The objective is to minimize the depletion probability of a typical node after m rounds, i.e., we would like to maximize $z_k(S)$. There are a number of constraints imposed on the network such as the scales of spatial events, the density of mobile sinks, compression performance, etc. Our goal is to determine an 'optimal' strategy S under these constraints. In order to quantify these, in the next subsection we introduce a spatial model for both phenomena being sensed and the network.

4.3.3 Spatial Models for the Network and Sensed Phenomena

In this thesis we will assume the following hierarchical network organization. We assume there is a high density of possibly redundant nodes in the region of interest. For simplicity sensing activities are assumed to occur in rounds. On each round, a random fraction of nodes act as sensors and relays for forwarded traffic. A random fraction of the remaining nodes become AGNs performing sensing, aggregation and forwarding of 'compressed' information to the nearest sink. The remaining nodes, are put to 'sleep' until the next round. Since the original 'pool' of nodes is large the set of locations for sensors and AGNs can be roughly assumed to be identically and independently distributed on each round. To provide a consistent
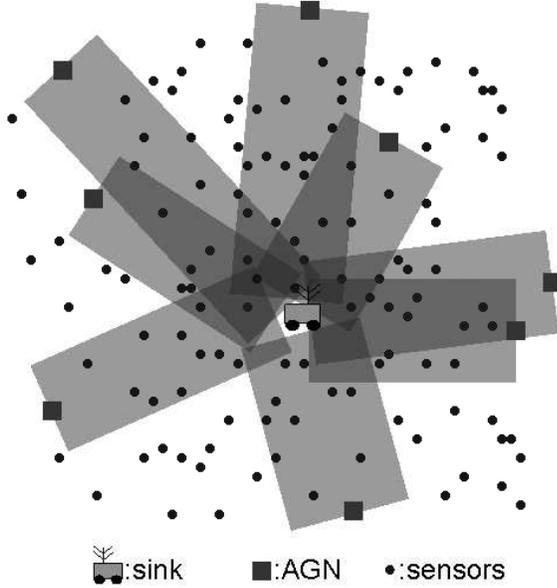
Figure 4.3: Strip model.

coverage for the phenomena of interest as well as relaying capacity the sum of densities of sensors and AGNs is constant at each round.

Finally we will assume the locations of sensors constitute a homogeneous Poisson Point Process (PPP), and the locations of sinks at each round also constitute a PPP with the same intensity but such that they can be assumed *independent* across the rounds. The density of sinks is assumed to be much smaller than that of sensors, i.e., the cost for placing a large number of sinks is high, thus at each round a sink is on average associated with hundreds or thousands of sensors.

We model the spatio-temporal phenomena being sensed by a *Boolean model* [53]. In our model the locations where sensor data would be generated are captured by a random set Φ generated by a homogeneous PPP Π_e with density λ_e and a primary grain Φ_0 given by

$$\Phi_0 := \{x \in \mathbb{R}^2 \mid |x| < r_e\},$$

i.e., a circular disc with the *radius of events* r_e. Specifically Φ is given by

$$\Phi := \left\{ \bigcup_{n=1}^{\infty} (\Phi_0 + x_n) \mid x_n \in \Pi_e \right\}, \qquad (4.4)$$

i.e., this model captures spatially clustered events with radius r_e occurring at random locations. We assume at various rounds the network sees a new realization of Φ, i.e., a new set of such events. A sensor at location x 'detects' an event, and generates a packet of size 1 and sends it to the nearest AGN if $x \in \Phi$ on that round. Finally we assume that these events are 'rare' in the sense that λ_e is much smaller than the density of sensors – we discuss this in more detail in Section 4.4.

Our routing model for spatially spreading traffic is as follows. As in Chapter 3 we will capture the energy burdens induced by traffic routed from AGNs to sinks as two dimensional functions capturing the energy burden per unit area and whose support set captures a 'footprint' of spatially clustered multi-path routes between the AGNs and sinks and corresponds to a closed set in $\mathbb{R}^2$. More specifically we shall assume traffic is spread over rectangular 'strips' of certain width which represents the spatial scale of spreading. Fig. 4.3 depicts the strip model. The strip originating from each AGN represents a spatial footprint for energy burdens towards a sink. For example, in a region where multiple strips overlap, sensors would see energy burdens proportional to the sum of energy burdens contributed by overlapping strips. Intuitively if the width of strips becomes larger, one would see larger regions with overlaps, however this may not translate into higher energy burdens in those regions since each unit of area in a larger strip would carry less traffic due to spreading.

In fact energy burdens are not homogenous over a strip. We incorporate the energy burden associated with traffic spreading in the strip model as follows. We associate each strip with a function capturing the energy burdens at different locations as shown by Fig. 4.4. The proposed function is motivated by 'continuum' analogy for flows in high-density networks [54, 55]. For example, the flow density corresponds to field strength within a homogeneous medium in electrostatic problems. Note that as a solution to such problems one obtains a field strength exhibiting a decay that is harmonic with the distance from the source. We will make a similar analogy such that a 'well-balanced' spreading of traffic leads to a harmonic distribution of flow strengths around source and destination.

Let the set $S_w(s, d)$ be a rectangular strip of width w with its ends located at s and d. We assume $w \geq 1$ where the case $w = 1$ corresponds to the 'baseline'
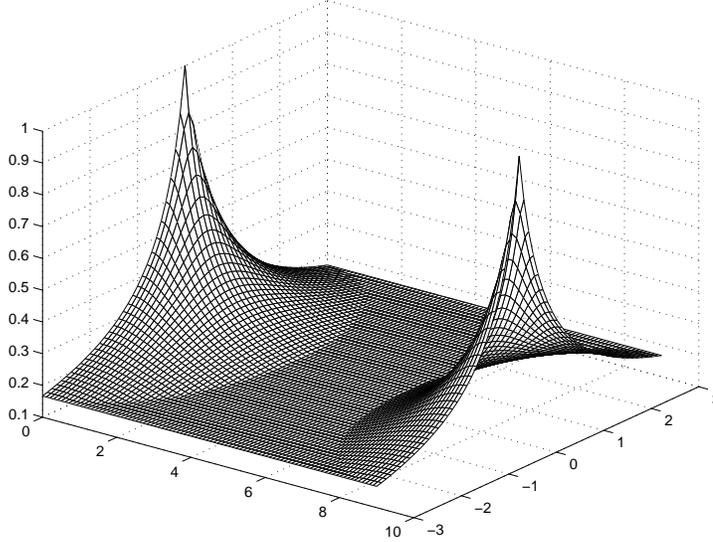
Figure 4.4: Energy burden density on a strip.

scheme, i.e., a scheme without traffic spreading. The density of energy burdens, i.e., burden per unit area, is given by a function $h_w : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ such that, for $x \in S_w(s,d)$ and $|s - d| \geq w$,

$$h_w(x,s,d) = \begin{cases} \frac{w}{2(w-1)|z-x|+w} & \text{if } |z - x| < \frac{w}{2} \text{ for } z = s, d, \\ w^{-1} & \text{otherwise,} \end{cases}$$

and when $|s - d| < w$, $h_w(x,s,d) := 1$, i.e., traffic is not spread for AGNs located within a distance w from the associated sink. If $x \notin S_w(s,d)$, h is defined to be 0. An example of h for $w = 6$ and $|s - d| = 9$ is shown in Fig. 4.4. The function has a peak value 1 at s and d, harmonically decreases to w^{-1} between s and d and assumes constant value in the remaining area of the strip. We use this function to capture the energy burden density associated with change of traffic load when the traffic diverges and converges at endpoints, and is spread by a factor of w.

4.4 Derivation of Energy Burdens under the Strip Model

As stated in the previous section, for the optimal design of the system we need to estimate the mean and variance of the spatial energy burdens, i.e., $\mu(S)$ and

87

$\sigma(S)$. For the sake of simplicity we will estimate the mean and variance of energy burdens *at a typical location* in the network.

4.4.1 Mean Energy Burden

We introduce our notation:

- Π_0, Π_1, Π_2 denote PPPs of sensors, AGNs and mobile sinks during a probing round respectively.

- λ_k is the intensity of point process Π_k for $k = 0, 1$ and 2.

- $V_z(\Pi)$ is the Voronoi cell with the nucleus $z \in \Pi$ induced by the point process Π.

- $\mathbf{E}_k^x$ is the Palm expectation given that a point in Π_k is located at x.

- $S_w(y, z)$ is a rectangular strip of width w with the midpoints of its ends located at y and z.

- $B_r(x)$ is a ball of radius r with its center at x.

- Φ denotes a Boolean process with primary grain $\Phi_0 := B_{r_e}(O)$ generated by a PPP Π_e with intensity λ_e.

In the sequel we assume that $\lambda_2 \ll \lambda_0, \lambda_1$ and $(\sqrt{\lambda_2})^{-1} \gg w$, i.e., the spatial scale of the typical Voronoi cell $V_0(\Pi_2)$ is much larger than that of sensors, AGNs and the spreading width with high probability. We will refer to this as the *large-cell assumption* which accounts for the relative sparsity of mobile sinks versus the density of sensors.

The system design parameters are λ_1 and w, i.e., the density for aggregation and the scale for traffic spreading. Thus we denote the mean and variance of energy burdens at a typical location by $\mu(\lambda_1, w)$ and $\sigma^2(\lambda_1, w)$ respectively. We start by approximating $\mu(\lambda_1, w)$.

We define the following:

$$\mathcal{M}_y := \mathbf{1}(y \in \Phi) + \sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_y(\Pi_1)),$$

$$\mathcal{N}_y := f(\mathcal{M}_y).$$

For an AGN cell with nucleus at y, $\mathcal{M}_y$ accounts for the number of sensors in Π_0, including AGN itself, which are covered by Φ. Thus $\mathcal{M}_y$ is the total amount of data generated from the AGN at y, and $\mathcal{N}_y$ is the output from AGN after 'compression'. We define $h_w(x, y, z)$ as the energy burden density experienced by a sensor at location x incurred by a strip originating from y towards the sink located at z. By multiplying $\mathcal{N}_y$ by $h_w(O, y, z)$ we obtain the energy burden density experienced at the origin when the aggregated traffic at y is forwarded to a sink at z.

We are interested in the overall contribution of strips originating from each AGN to its closest sink that overlaps at the typical sensor from Π_0, e.g., see Fig. 4.3. Since this contribution is stationary, one can write the expectation of the energy burdens from the perspective of a random location in $\mathbb{R}^2$:

$$\mu(\lambda_1, w) := \mathbf{E}\left[\sum_{y_j \in V_{z_0}(\Pi_2) \cap \Pi_1} h_w(O, y_j, z_0)\mathcal{N}_{y_j} \right] \tag{4.5}$$

By z_0 we denote a random point denoting the location of the closest sink to the origin: note this is an approximation which ignores the edge effect on the boundaries of Voronoi cells induced by Π_2: this will be explained with Fig.4.5 in the sequel.

Note that (4.5) captures the energy burden seen at a randomly chosen 'location' on the plane. A better approach would be to estimate the energy burden seen at a 'typical' sensor since this provides an *unbiased* view of the measure of interest. However this requires excessively complicated calculations, hence we will work with an approximation.

With this model we have the following proposition.

Proposition 1. *Under the proposed strip model $\mu(\lambda_1, w)$ is upper bounded by*

$$\lambda_1 \mathbf{E}_1^0[\mathcal{N}_0] \cdot \tilde{m}(w, \lambda_2) \tag{4.6}$$

where we define $\tilde{m}(w, \lambda_2)$ as

$$\tilde{m}(w, \lambda_2) := m(w)\exp(-\pi\lambda_2 w^2) + \frac{1}{2\sqrt{\lambda_2}}.$$

and $m(w)$ as

$$m(w) := \frac{\pi w^2}{(w-1)^2}(w - 1 - log(w)) - \frac{\pi w}{2}, \ w \geq 1. \tag{4.7}$$

89

Proof. From refined Campbell's theorem [53] one can rewrite (4.5) as

$$\lambda_2 \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} \left[\sum_{y_j \in V_0(\Pi_2) \cap \Pi_1} h_w(x, y_j, O) \mathcal{N}_{y_j} \right] dx \right]$$

$$= \lambda_2 \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} \mathbf{E}_{0,1} \left[\sum_{y_j \in V_0(\Pi_2) \cap \Pi_1} h_w(x, y_j, O) \mathcal{N}_{y_j} \right] dx \right]$$

where we have moved the expectation with respect to Π_0 and Π_1, denoted by $\mathbf{E}_{0,1}$, inside the integral from independence among Π_0, Π_1 and Π_2. Applying Campbell's formula [53] for the term inside the integral, we can write

$$\lambda_2 \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} \int_{V_0(\Pi_2)} \mathbf{E}_1^y \left[h_w(x, y, O) \mathcal{N}_y \right] \Pi_1(dy) dx \right] \tag{4.8}$$

$$= \lambda_2 \lambda_1 \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} \int_{V_0(\Pi_2)} h_w(x, y, O) \mathbf{E}_1^0 \left[\mathcal{N}_0 \right] dy dx \right] \tag{4.9}$$

By changing the order of integration in (4.9), i.e., integrating with respect to x first, the following holds for the inner integral:

$$\int_{V_0(\Pi_2)} h_w(x, y, O) dx = \int_{V_0(\Pi_2) \cap S_w(y,O)} h_w(x, y, O) dx \tag{4.10}$$

$$\leq \int_{S_w(y,O)} h_w(x, y, O) dx \tag{4.11}$$

Thus by defining

$$g_w(y, z) := \int_{S_w(y,z)} h_w(x, y, z) dx$$

we have an upper bound for (4.9) given by

$$\lambda_1 \lambda_2 \mathbf{E}_1^0 [\mathcal{N}_0] \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} g_w(y, O) dy \right], \tag{4.12}$$

however one can write

$$\mathbf{E}_2^0\left[\int_{V_0(\Pi_2)} g_w(y,O)dy\right]$$

$$= \mathbf{E}_2^0\left[\int_{\mathbb{R}^2} g_w(y,O)\mathbf{1}(\Pi_2 \cap B_{|y|}(y) = \emptyset)dy\right]$$

$$= \int_{\mathbb{R}^2} g_w(y,O)\mathbf{E}_2^0[\mathbf{1}(\Pi_2 \cap B_{|y|}(y) = \emptyset)]dy$$

$$= \int_{\mathbb{R}^2} g_w(y,O)\exp(-\pi\lambda_2|y|^2)dy. \tag{4.13}$$

since Π_2 is a homogeneous PPP. By the definition of $h_w(x,y,z)$ we have that

$$g_w(y,O) = \begin{cases} |y| & |y| < w, \\ m(w) + |y| & |y| \geq w \end{cases}$$

where $m(w)$ is given by (4.7). We have that

$$\int_{\mathbb{R}^2} g_w(y,O)\exp(-\pi\lambda_2|y|^2)dy$$

$$= \int_{\mathbb{R}^2} |y|e^{-\pi\lambda_2|y|^2}dy + m(w)\int_{\mathbb{R}^2\setminus B_w(O)} e^{-\pi\lambda_2|y|^2}dy$$

$$= \int_{\mathbb{R}^2} |y|e^{-\pi\lambda_2|y|^2}dy + \frac{m(w)\exp(-\pi\lambda_2 w^2)}{\lambda_2}. \tag{4.14}$$

We integrate (4.14) by switching to polar coordinates, defining $\rho := |y|$ and using

$$\int_0^\infty \rho^n \exp(-\pi\lambda_2\rho^2)d\rho = \frac{\Gamma(\frac{n+1}{2})}{2(\pi\lambda_2)^{\frac{n+1}{2}}},$$

and from (4.12) and (4.14) we obtain the proposed upper bound. $\qquad\square$

Although (4.6) serves as an upper bound for the desired expression, we would like to use it as an approximation to the mean cost throughout the thesis considering the inequalities involving (4.11) and (4.14) as follows.

Indeed (4.11) is an overestimate of (4.10) since it includes the extra contribution incurred by some strips that 'protrude' the cell boundary of a typical sink, e.g., see the shaded region A in Fig. 4.5. However we have also ignored the contribution of strips that 'overlap' the cell boundary as is the case for A' in Fig. 4.5, i.e., our
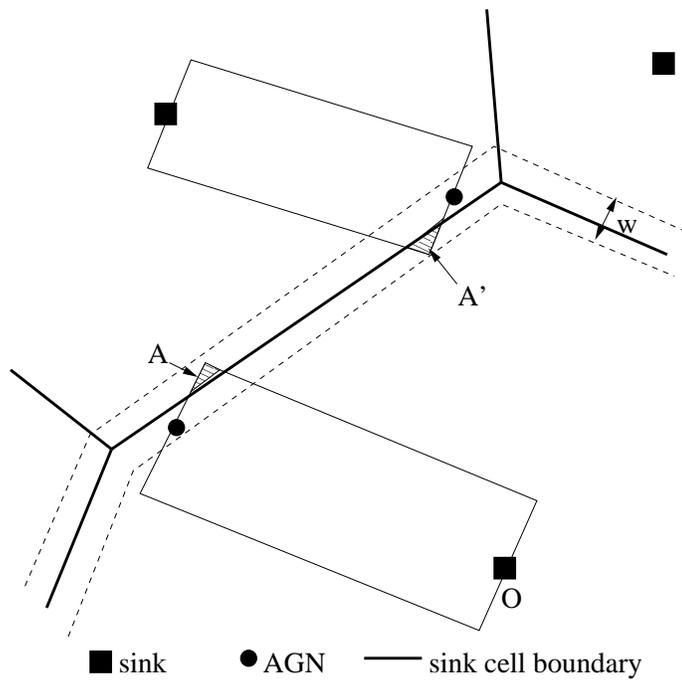
Figure 4.5: Edge effects in the typical sink cell $V_0(\Pi_2)$. Note the dimension of the size of strips are exaggerated in the figure for clarity.

expression is already an underestimate of the actual mean cost. Thus considering the homogeneity of Π_0 and Π_1, such errors in the cost roughly 'cancel out'. We see that these edge effects occur when the strips originate from AGNs which are located within a narrow region along the boundary of the typical sink cell. This region is a subset of a 'band' along the boundary as depicted in dotted line in Fig. 4.5, and we see that its width is at most w. Considering the large size of a typical sink cell with respect to the dimension of strips, we note that the error associated with such band region is relatively small. Thus we conclude that the majority of strips are contained within $V_0(\Pi_2)$ considering the relatively large density of AGNs.

Finally note that $m(w)$ which is an increasing function of w with $m(1) = 0$, can be viewed as a static cost incurred from spreading traffic over a strip of width w.

The following proposition gives an estimate for $\mathbf{E}_1^0[\mathcal{N}_0]$.

Proposition 2. *Under the proposed network model we have that*

$$\mathbf{E}_1^0[\mathcal{M}_0] = (1 + \frac{\lambda_0}{\lambda_1})\{1 - \exp(-\lambda_e \pi r_e^2)\}, \tag{4.15}$$

$$\mathbf{E}_1^0[\mathcal{N}_0] \leq \alpha \mathbf{E}_1^0[\mathcal{M}_0] \tag{4.16}$$

$$+ (1 - \alpha)(1 - \exp\{-\lambda_e(\frac{1}{\lambda_1} + \frac{4r_e}{\sqrt{\lambda_1}} + \pi r_e^2)\}). \tag{4.17}$$

Proof. By definition of $\mathcal{M}_0$, we have that

$$\mathbf{E}_1^0[\mathcal{M}_0] := \mathbf{E}\left[\mathbf{1}(O \in \Phi)\right] + \mathbf{E}\left[\sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_0(\Pi_1))\right].$$

Considering the first term, since Φ is stationary, let us denote the contact probability of a stationary Boolean model [53] which we denote as p and is given by

$$p := 1 - \exp(-\lambda_e \mathbf{E}[|\Phi_0|]) = 1 - \exp(-\lambda_e \pi r_e^2),$$

i.e., the probability that a randomly chosen point on $\mathbb{R}^2$ will intersect with Φ, which

gives the first term. We have that

$$\mathbf{E}_1^0\left[\sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_0(\Pi_1))\right]$$

$$= \lambda_0 \mathbf{E}_1^0\left[\int_{\mathbb{R}^2} \mathbf{1}(x \in \Phi \cap V_0(\Pi_1))dx\right],$$

$$= \lambda_0 \int_{\mathbb{R}^2} \mathbf{E}_1^0[\mathbf{1}(x \in V_0(\Pi_1))]\mathbf{E}[\mathbf{1}(x \in \Phi)]dx$$

$$= \lambda_0 \int_{\mathbb{R}^2} \mathbf{P}_1^0(\Pi_1 \cap B_{|x|}(x) = \emptyset)\mathbf{P}(x \in \Phi)dx$$

$$= \lambda_0\{1 - \exp(-\lambda_e \pi r^2)\}\int_{\mathbb{R}^2} \exp(-\lambda_1 \pi |x|^2)dx$$

$$= \frac{\lambda_0}{\lambda_1}\{1 - \exp(-\lambda_e \pi r^2)\},$$

which proves (4.15).

By definition of $f(\cdot)$ we have that

$$\mathcal{N}_0 = \alpha\mathcal{M}_0 + (1 - \alpha)\mathbf{1}(\mathcal{M}_0 > 0),$$
$$\mathbf{E}_1^0[\mathcal{N}_0] = \alpha\mathbf{E}_1^0[\mathcal{M}_0] + (1 - \alpha)\mathbf{P}_1^0(\mathcal{M}_0 > 0)$$

It turns out that it is difficult to evaluate $\mathbf{P}_1^0(\mathcal{M}_0 > 0)$, thus we will once again resort to a bound. Note that $\mathbf{P}_1^0(\mathcal{M}_0 > 0)$ is the probability that there exists either an AGN or a sensor which lies within Φ within a typical Voronoi cell associated with an AGN at the origin. A necessary condition for that event to occur is that Φ have a nonempty intersection with the Voronoi cell. In other words the following relation between the events needs to hold:

$$\{\mathcal{M}_0 > 0\} \subseteq \{V_0(\Pi_1) \cap \Phi \neq \emptyset\},$$

which implies $\mathbf{P}_1^0(\mathcal{M}_0 > 0) \leq \mathbf{P}_1^0(V_0(\Pi_1) \cap \Phi \neq \emptyset)$.

The 'capacity functional', i.e., probability that a stationary Boolean model intersects with a compact set K, is given by

$$1 - \exp(-\lambda_e|B_{r_e}(O) \oplus K|)$$

where $\oplus$ denotes Minkowski addition [53]. Thus we have that

$$\mathbf{P}_1^0(V_0(\Pi_1) \cap \Phi \neq \emptyset)$$
$$= 1 - \mathbf{E}_1^0 \exp(-\lambda_e |B_{r_e}(O) \oplus V_0(\Pi_1)|)$$
$$\leq 1 - \exp(-\lambda_e \mathbf{E}_1^0[|B_{r_e}(O) \oplus V_0(\Pi_1)|]) \tag{4.18}$$
$$= 1 - \exp(-\lambda_e(\mathbf{E}_1^0 |V_0(\Pi_1)| + \mathbf{E}_1^0[\partial V_0(\Pi_1)]r_e + \pi r_e^2)) \tag{4.19}$$
$$= 1 - \exp\{-\lambda_e(\frac{1}{\lambda_1} + \frac{4r_e}{\sqrt{\lambda_1}} + \pi r_e^2)\}. \tag{4.20}$$

In (4.18) we have used Jensen's inequality. In (4.19) we have used Steiner's formula for Minkowski addition since the primary grain Φ_0 is a disc [53], where $\partial V_0(\Pi_1)$ denotes the perimeter of a typical Voronoi cell. In (4.20) we have used the mean area and perimeter results for a typical cell of a stationary Voronoi tessellation. This proves (4.17). $\qquad\square$

We introduce the following notation:

$$p_b := (1 - \exp\{-\lambda_e(\frac{1}{\lambda_1} + \frac{4r_e}{\sqrt{\lambda_1}} + \pi r_e^2)\})$$

To summarize the mean energy burden (4.5) is upper bounded by

$$\tilde{m}(w, \lambda_2) \left[(\lambda_0 + \lambda_1)\alpha p + \lambda_1 (1 - \alpha)p_b \right].$$

We will use this bound as an approximation to the actual mean energy burden at the typical sensor.

4.4.2 Variance of Energy Burdens

Next we evaluate $\sigma^2(\lambda_1, w)$ which is the variance of the energy burdens at a typical location. Since $\mu(\lambda_1, w)$ has been estimated we need to only consider the

non-centered second moment which is given by the following:

$$
\mathbf{E}\left[\left\{\sum_{y_j \in V_{z_0}(\Pi_2) \cap \Pi_1} h_w(O, y_j, z_0) \mathcal{N}_{y_j}\right\}^2\right]
$$

$$
= \lambda_2 \mathbf{E}_2^0\left[\int_{V_0(\Pi_2)}\left[\sum_{y_j \in V_0(\Pi_2) \cap \Pi_1} h_w(x, y_j, O) \mathcal{N}_{y_j}\right]^2 dx\right]
$$

$$
= \lambda_2 \mathbf{E}_2^0\left[\int_{V_0(\Pi_2)}\left[\sum_{y_j \in V_0(\Pi_2) \cap \Pi_1} h_w^2(x, y_j, O)(\mathcal{N}_{y_j})^2\right.\right. \tag{4.21}
$$

$$
\left.\left. + \sum_{\substack{y_j, y_k \in V_0(\Pi_2) \cap \Pi_1 \\ y_j \neq y_k}} h_w(x, y_j, O) h_w(x, y_k, O) \mathcal{N}_{y_j} \mathcal{N}_{y_k}\right] dx\right]. \tag{4.22}
$$

It turns out that this second moment is difficult to evaluate due to the second term of the above, i.e., (4.22). Thus we will develop an approximation using the first term (4.21) of the above as follows:

1. Estimate (4.21) as done in Proposition 3.

2. In order to do so, estimate $\mathbf{E}[\mathcal{N}_0^2]$ up to the first order in p.

3. Using the estimate for (4.21), we propose an approximation to $\sigma^2(\lambda_1, w)$.

It is difficult to directly evaluate $\mathbf{E}[\mathcal{N}_0^2]$. Thus for simplicity we will only concern ourselves with terms that are $O(p)$. We assume that the sensing phenomena are 'rare' events, i.e., $p \ll 1$. Small p means either the density of events λ_e or the spatial scale of events r_e, or both, are small. In our model we assume λ_e is a small number compared to λ_0 and λ_1 however r_e is such that the area of events πr_e^2 is large relative to the density of sensors, e.g., an event disc covers several sensors with high probability. Overall $\lambda_e \pi r_e^2$ is assumed to be a small number compared to 1 such that $p = 1 - \exp(-\lambda_e \pi r_e^2) \approx \lambda_e \pi r_e^2$ is a good approximation.

We start with the following proposition:

Proposition 3. *One can upper bound (4.21) by*

$$\lambda_1 \mathbf{E}[\mathcal{N}_0^2] \cdot \tilde{v}(w, \lambda_2)$$

where we define $\tilde{v}(w, \lambda_2)$ *as*

$$v(w) \exp(-\pi \lambda_2 w^2) + \frac{1}{2w\sqrt{\lambda_2}} \qquad (4.23)$$

and $v(w)$ *is defined as*

$$v(w) = \frac{\pi w^2}{(w-1)^2} \left\{ \frac{1}{w} - 1 + \log(w) \right\} - \frac{\pi}{2}, \ w \geq 1.$$

Proof. The proof is similar to that of Proposition 1. $\qquad\qquad\square$

Next we bound $\mathbf{E}[\mathcal{N}_0^2]$ up to the first order in p. By definition of $\mathcal{N}_0$ we have that

$$\mathbf{E}[\mathcal{N}_0^2] = \alpha^2 \mathbf{E}[\mathcal{M}_0^2] + 2\alpha(1-\alpha)\mathbf{E}[\mathcal{M}_0]$$
$$+ (1-\alpha)^2 \mathbf{P}(\mathcal{M}_0 > 0).$$

We need to evaluate $\mathbf{E}[\mathcal{M}_0^2]$ which can be written as

$$\mathbf{E}[\mathcal{M}_0^2] = \mathbf{E}\left[\mathbf{1}(O \in \Phi)\right] + \mathbf{E}\left\{ \sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_0(\Pi_1)) \right\}^2$$

$$+ 2\mathbf{E}\left[\mathbf{1}(O \in \Phi) \sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_0(\Pi_1))\right]. \qquad (4.24)$$

The first term is simply the contact probability p. We have the following proposition for the second term of (4.24).

Proposition 4. *We have that, up to the first order in* p,

$$\mathbf{E}_1^0\left[\sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_0(\Pi_1)) \right]^2$$

is bounded above by

$$\frac{\lambda_0}{\lambda_1} p + 2.56 p \frac{\lambda_0^2 \sqrt{\lambda_1} r_e}{\lambda_1^2 (c + 2\sqrt{\lambda_1} r_e)}$$

where the constant c *is by the expression (C.7) in the appendix.*

97

Proof. See appendix. □

Proposition 5. *The third term in (4.24) is, up to the first order in p, bounded above by*

$$\frac{2\lambda_0 p}{\lambda_1}\{1 - \frac{erf(2\pi r_e\sqrt{\lambda_1})}{4r_e\sqrt{\lambda_1}}\}$$

where erf(·) is the standard error function, i.e.,

$$erf(x) := \frac{2}{\sqrt{\pi}}\int_0^x e^{-z^2}dz.$$

Proof. See appendix. □

Combining the propositions we have that $\mathbf{E}[\mathcal{N}_0^2]$ is bounded above by

$$\alpha^2 p\left[1 + \frac{3\lambda_0}{\lambda_1} - \frac{\lambda_0 erf(2\pi r_e\sqrt{\lambda_1})}{2r_e\lambda_1\sqrt{\lambda_1}} + 2.56\frac{\lambda_0^2\sqrt{\lambda_1}r_e}{\lambda_1^2(c + 2\sqrt{\lambda_1}r_e)}\right]$$
$$+ 2\alpha(1-\alpha)(1 + \frac{\lambda_0}{\lambda_1})p + (1-\alpha)^2 p_b + o(p) \tag{4.25}$$

where we will ignore $o(p)$ term. Using this expression for $\mathbf{E}[\mathcal{N}_0^2]$ we get an estimate for (4.21) as stated by Proposition 3. Then using the estimate for (4.21) we resort to simulation of our model and propose an approximation of the variance as follows.

Our simulation results show that, the variance of energy burdens $\sigma^2(\lambda_1, w)$ is well approximated by (4.21) alone provided that the sizes Voronoi cells induced by AGNs are large relative to the spatial scales of phenomena. This is expected since the cross terms $\mathcal{N}_{y_j}\mathcal{N}_{y_k}$ in (4.22) become less and less correlated as the number of Π_1 cells intersected by a *single* event become smaller. Due to the 'scarcity' of phenomena and the large sink cell size, (4.22) will be close to the square of mean burdens thus they effectively cancel out. One can visualize this easily, if the radius of events tends to 0 then $\mathcal{N}_{y_j}$ and $\mathcal{N}_{y_k}$ will be become completely uncorrelated.

However if the spatial scale of events are comparable to AGN cell sizes, the correlation term plays a major role. Since this is related to the relative scales between event discs and the density of AGNs, we have developed the following approximation for $\sigma^2(\lambda_1, w)$ based on extensive simulations:

$$\sigma^2(\lambda_1, w) \approx \min(1, \beta(\pi\lambda_1 r_e^2)^\nu) \cdot \lambda_1 \mathbf{E}[\mathcal{N}_0^2]\tilde{v}(w, \lambda_2) \tag{4.26}$$

where β and ν are some positive constants. In other words we include an extra factor as a function of $\pi\lambda_1 r_e^2$ multiplied by (4.21). Roughly speaking, $\pi\lambda_1 r_e^2$ is the average number of λ_1 points in an event radius, thus it captures the relative scales of phenomena and AGNs. In addition to (4.21) this is a crude approximation to $\sigma^2(\lambda_1, w)$ as an effort to capture the degree of 'clustering effect' of events across AGN cells.

Our simulation results show that $\beta \approx 0.89$ and $\nu \approx 0.5$ provide a good approximation. Thus we will use (4.26) as the expression for variance of energy burdens with the above β and ν in the numerical results in the next section.

Finally, we take the cost associated with aggregation into account. The aggregation of traffic is modelled by, strips of width 1 generated at the sensors towards their AGNs, and each strip carries traffic of size 1. Then the mean energy burden caused by aggregation can be computed by simply adding $\frac{1}{2\sqrt{\lambda_1}}$, i.e., the traffic seen from random location inside the typical AGN cell, to the original expression for the mean, i.e., (4.6).

As for variance we make the assumption such that, due to the large cell size, most of the imbalance in energy burdens are caused by traffic concentration around sinks and accumulative burdens associated with forwarding from AGNs, thus we ignore the contribution to variance from aggregation. A more refined treatment of the model for aggregation is a subject to future study.

4.5 Numerical Results for the Optimal Design

In this section we investigate optimizing the degree of aggregation and traffic spreading for a given compression performance, the density of mobile sinks and the spatial scales of phenomena via numerical methods. As introduced in Section 4.3.2, our objective is to maximize the lifetime of the network. Since we have estimated $\mu(\lambda_1, w)$ and $\sigma(\lambda_1, w)$ the optimization problem can be written as

$$\text{maximize: } z_k(\lambda_1, w) := \sqrt{m} \left\{ \frac{k - \mu(\lambda_1, w)}{\sigma(\lambda_1, w)} \right\},$$

$$\text{subject to: } \lambda_2 < \lambda_1 \leq 1, \ w \geq 1$$

$$\text{variables: } w, \lambda_1,$$

$$\text{given: } \lambda_2, \alpha, k, m, p, r_e, \lambda_e.$$

Note that we assume the sum of densities of sensors and AGNs are fixed to 1, thus the density of sensors $\lambda_0 = 1 - \lambda_1$ with the proper adjustments made to the previous analytical results. We numerically find the optimal scales for traffic spreading and aggregation, i.e., w and λ_1 and it turns out that with the above constraints $z_k(\lambda_1, w)$ admits an unique optimal value.

First we investigate the impact of the sink density λ_2 and the compression ratio α on the optimal spatial scales for traffic spreading and aggregation. In these results we have assumed the network aims to operate a total of $m = 200$ rounds. The maximum energy reserve k is appropriately scaled to obtain each point in the plot such that the optimal spatial scales yield a depletion probability of 10^{-4}. For the plots Fig. 4.6 and Fig. 4.7 the radius of event disc r_e is fixed to 4 which implies there are about 50 sensors observing a single event disc on average. The values of r_e varies in plots Fig. 4.8 and Fig. 4.9. In any case the contact probability p is fixed to 0.1 and thus the density of phenomena λ_e is adjusted with r_e.

Fig. 4.6 (Fig. 4.7) show the optimal spatial scale w^* (resp. λ_1^*) for traffic spreading (resp. aggregation density) with varying λ_2 and α, Note that in these plots the axis for the density of sinks is shown in terms of the average number of sensors associated with a sink: for example 1000 nodes per sink would correspond to $\lambda_2 = 10^{-3}$. The number of sensors per sink ranges from 500 to 2000, and α takes values 0.2, 0.4, 0.6 and 0.8 which represent excellent, good, moderate and poor compression performance respectively. The values of λ_1^* in Fig. 4.7 can be interpreted as follows: $\lambda_1^* = 0.2$ means that the density of sensors are 0.8, thus an AGN aggregates traffic from 4 sensors on average. In the following we list our main observations.

Longer routes require more spreading of traffic. By inspecting Fig. 4.6 we can make the following observation. When we fix the compression ratio
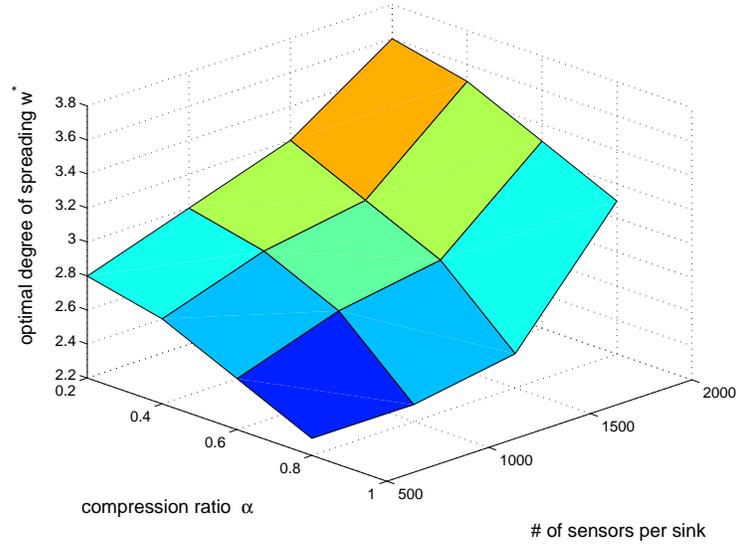
Figure 4.6: Optimal degree of traffic spreading as a function of compression ratio and sink density.
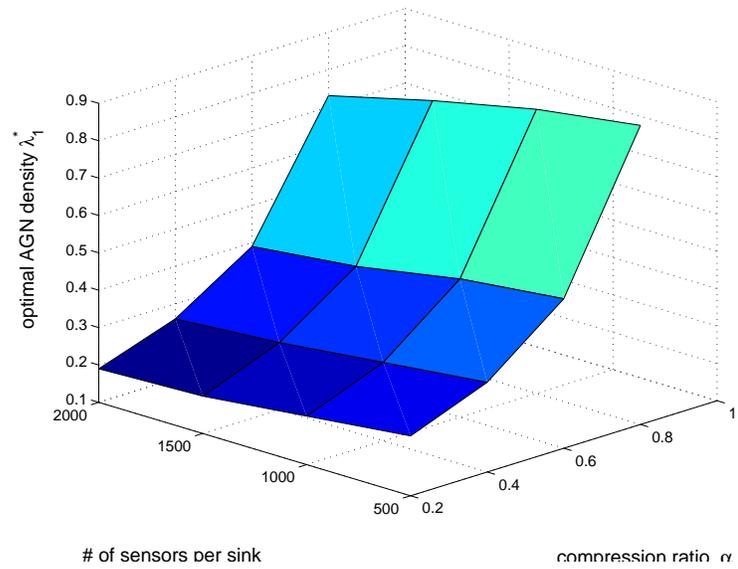
Figure 4.7: Optimal aggregation density as a function of compression ratio and sink density.
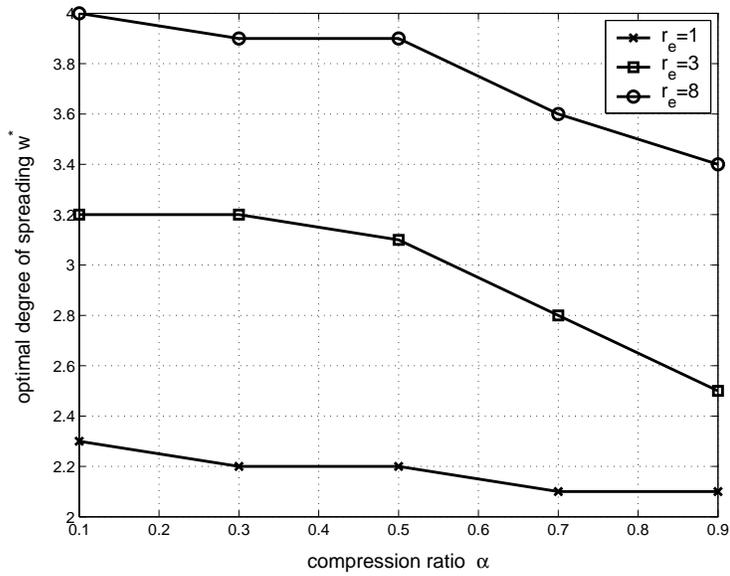
Figure 4.8: Optimal degree of traffic spreading as a function of compression ratio and spatial scale of events.
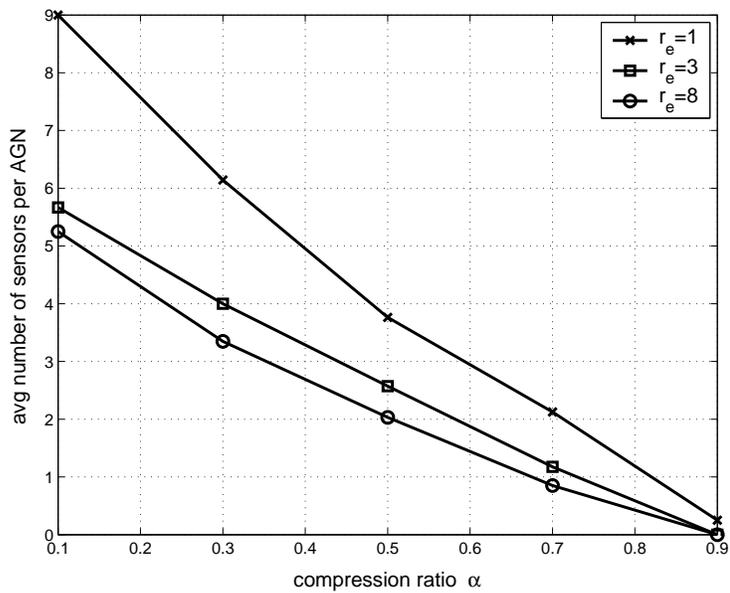
Figure 4.9: Optimal average number of sensors per AGN as a function of compression ratio and spatial scale of events.

the optimal spreading width tends to increase as the number of sensors per sink increases. We see that in Fig. 4.7 the change in λ_1^* is negligible with the variations in λ_2 when α is fixed, i.e., the traffic generated per AGN is roughly constant. This implies that as λ_2 decreases the traffic travels longer distances, in which case the result in Fig 4.6 indicates that traffic should spread *more*. This is intuitive since we would see a larger number of overlapping strips for smaller sink densities thus reducing the variability by spreading improves the performance.

However note that the spreading width ranges between 2 and 4, which implies that the degree of spreading of 2 or 3 achieves most of the gains associated with load balancing in this network. Further spreading would incur extra energy costs to counterbalance these benefits.

Higher loads require additional spreading of traffic. Let us consider how w^* varies when λ_2 is fixed and we change α. Fig. 4.6 shows that we should spread more conservatively when the compression worsens. This may be interpreted as a tension between mean and variance of the energy burdens, i.e., the mean plays larger role per sensor in network lifetime with deteriorating compression ratio when the size of sink cells is fixed. However if we consider Fig. 4.7 at the same time the overall traffic per strip *increases* when the compression performance improves. For example when the average number of sensors per sink is 2000, the optimal pair (w^*, λ_1^*) is given by $(3.1, 0.4)$, i.e., roughly 1.5 sensors per AGN, when $\alpha = 0.8$ and $(3.6, 0.2)$, i.e., roughly 4 sensors per AGN, when $\alpha = 0.2$. Thus each strip offers roughly 0.6 and 2 units of load on average when α is given by 0.8 and 0.2 respectively, where w^* is higher for $\alpha = 0.2$ case at that operating point. Since the sink cell size is fixed the average distance each strip spans is identical, this also implies strips with larger loads should spread *more* given that the average length of strips is fixed. Finally in Fig. 4.7 we see that the degree of aggregation increases, i.e., the densities for AGN decrease with improving compression performance.

The above two observations are intuitive: if one should traverse certain distance and has more traffic, then one should spread the traffic more. Also if one has some traffic which has to traverse longer distances, then one should spread the traffic more. This is considering 'future' balance of the network and an effort to make the energy profile of the network benign such that the upcoming routes over the network see more balanced energy burden profiles along their paths.

When the phenomena are more 'bursty' it requires further spreading of traffic: Fig. 4.8 represents the optimal degree of spreading versus varying compression performance where each curve represents different radius of event discs. The values of r_e are given by 1,3 and 8 each of which corresponds to there being roughly 3, 28 and 200 sensors per event disc on average. Since the contact probability p is fixed, a growing r_e represents how 'clustered' events are, i.e., the burstiness of the spatial scale of events.

As one can see in the figure, it is clear that for larger event radii one should spread more. The intuition is that, as the spatial scale of an individual event grows the concentration of the traffic will be more severe, since the sensors nearby will detect the phenomena and concentrate traffic to their AGN. Further the traffic strips from AGN will overlap with each other with higher probability, as compared to that for smaller event discs. Thus the spreading of traffic becomes more critical if the phenomena of interest have a high degree of burstiness in its spatial scale.

When physical phenomena are less bursty the system admits further aggregation of traffic: Fig. 4.9 represents the optimal degree of aggregation versus varying compression ratio where each curve is associated with different values of r_e. Specifically y axis represents the optimal number of sensors per AGN on average. Clearly one can see that if events are less 'clustered' one can aggregate more traffic thus enjoy the benefits of compression further. If the events are highly bursty then more aggregation only would add the concentration of energy burdens on AGNs that see nearby events.

The aggregation of traffic on a large scale is not beneficial even for excellent compression performance: The final observation is that, rather surprisingly even for a very good compression performance, the degree of aggregation should not be too aggressive. For example see Fig. 4.9 for the case where $r_e = 1$ and $\alpha = 0.1$, i.e., even though one can reduce 90% of traffic via compression, the optimal degree of AGN organization is such that only about 1 out of 10 sensors becomes an AGN. Furthermore this corresponds to the case where the spatial events are highly 'scattered', i.e., $r_e = 1$ implies there are only about 3 sensors per event disc. Our results show that a only a few number of sensors per AGN turns out to be the optimal scale of aggregation in most cases.

The reason can be explained as follows: if one aggregates more traffic, it incurs more concentration of energy burdens between AGNs and sinks. Although

the problem can be mitigated by spreading traffic at a proper scale, it also incurs energy costs to do so. By contrast our model is such that the further one aggregates the traffic, the per node benefit of compression increases. Thus there is a tension between the effort to balance by spreading, and not to incur concentration in the first place by avoiding an excessive degree of aggregation, which sacrifices the benefit of compression.

Thus if the compression performance is substantial then it will be preferable to aggregate traffic, however one should not do this too aggressively – it depends on the spatial burstiness of phenomena. Overall the conclusion is that, the harm from the traffic concentration and burstiness of spatial scales of events have a synergistic effect on the imbalance of energy burdens, thus one should be very conservative on aggregating traffic on a large scale.

4.6   Conclusion

In this chapter we addressed a fundamental scalability problem for energy-constrained large scale sensor networks based on wireless relaying: sensors in the vicinity of sinks incur a much higher energy burden. To mitigate this problem we consider jointly exploiting several ideas: reducing the traffic on the network through local aggregation/compression and making the sinks mobile. However in most regimes where such schemes are applicable, we argued that network operation can still be optimized to enhance the networks lifetime. The key idea is to spread out aggregated traffic when it is forwarded from AGNs to sinks in order to smooth inhomogeneities in energy burdens the network will incur.

In particular we introduced a model for bursty phenomena in space, which serves as a caricature of some practical sensing applications. The burstiness the phenomena have a deleterious effect on the balancing of energy burdens. Interestingly under our model, the degree to which traffic should be spread is interwoven in a subtle way with the spatial scales on which traffic is aggregated. The tension lies in the following key tradeoffs. Increasing the degree of traffic spreading results in smoother energy burdens but incurs additional overall energy burdens. By contrast, increasing the degree of aggregation reduces the per sensor traffic and thus overall energy burdens, but increases the spatial variability of energy burdens. Moreover these tradeoffs are not 'orthogonal' to each other. In this chapter we provide

a stochastic geometric model to investigate this interaction. By adopting simple models for compression and energy burdens we were able to show how the mean and variance of energy burdens for a typical node relate to the scales for spreading and aggregation. We consider the joint optimization of these scales via numerical methods, and find that while spreading reduces the variance induced by overlaps of 'long' routes towards sinks this only helps to the extent that the increased mean energy costs do not counteract the benefit. Similarly the more aggregation the better, unless this adversely affects the variability of energy burdens on the network. From the numerical results we can concretely see how these aspects counterbalance each other.

Our key contributions lie in observing the effect of the spatial burstiness of sensing phenomena to the energy depletion of large-scale network. From our results we see that if the achievable compression is high one should aggregate more traffic. However if the events are spatially bursty this would require more spreading of traffic to counterbalance the concentration of energy burdens. Also if the events are less bursty one can aggregate more traffic to enjoy the benefit of compression. Overall, perhaps surprisingly, a high degree of aggregation for even moderate spatial scales of phenomena turns out to be a harmful idea even for good compression performance: the synergistic effects of clustered events and concentration of burdens at AGNs are typically hard to overcome by simple proactive spreading and gains in reducing traffic via compression. Thus one should avoid this in the first place: at most 10 sensors per AGN seems to be the reasonable way of organizing networks. However if the model incorporates more elegant way of providing compression/fusion gains, e.g., distributed compression or network coding, it may provide better ways overcome such burstiness with or without the traffic aggregation.

Chapter 5

Conclusions and Future Work

In this chapter we summarize the conclusions and insights we have discussed throughout the dissertation and present future research directions.

5.1 Summary of Results

In the first part of the thesis we have studied the engineering of ad-hoc/sensor networks with data compression capability. In a network model with distributed data compression capabilities we have proposed how to minimize energy expenditures associated with gathering data to a query point. We found that the optimal compression strategy is to order nodes based on an associated weight, e.g., distances to gathering point, and compress using the side information for nodes higher up in the ordered sequence.

In large-scale sensor networks with hierarchical aggregation and compression, where the energy cost is modelled by a linear function in the distance between source and destination which approximates the hop count, the network structure which minimizes the mean energy is associated with the Johnson-Mehl tessellation induced by the locations of nodes and depends on the compression performance. For models where any destination can be reached in one hop, however with possibly varying path loss exponents, one can further achieve energy savings by identifying the new optimal organizations of network aggregation hierarchies.

In the second part of the thesis we explored how proactive multi-path routing in multihop ad hoc wireless networks can help extending network lifetime. Specifically, improvements are achieved by exploiting the tradeoff between the increased energy burden associated with traffic spreading and the resulting benefit of balancing energy burdens in the network – we study Pareto optimal operational point for the mean and the variance of the energy burdens seen at a typical node in the network.

We have shown that the spatial scale for load balancing, i.e., width of traffic spreading can be optimized for different spans and loads offered by sessions on the network. Notably when the offered traffic is spatially homogeneous, the optimal degree of spreading is proportional to the square root of the $bits{\times}meters$ of the offered load. This result has an important practical implication regarding the extent to which one should be 'aggressive' in load balancing if the goal is to maximize network lifetime.

In the final part of the thesis we proposed a model for large sensor networks detecting spatio-temporally bursty phenomena. We develop a model and analysis to support designs which avoid excessive concentration of energy burdens in the neighborhood of sinks: deployment of mobile sinks, compression/fusion of local traffic and proactive load balancing routing. Most importantly we find that when phenomena are highly bursty in space, one should spread traffic more in order to counterbalance the concentration of energy burdens in the vicinity of the phenomena.

Finally we found a, perhaps, surprising principle for the design of networks with aggregation and data fusion geared at detecting bursty events. Most of the energy gains can be achieved by performing fusion on several data samples and limiting the degree of spreading to 2 or 3 disjoint nodes. These results demonstrate a crucial guideline for network design, i.e., that design methodologies should be realized with careful consideration of the counterbalancing effects associated with the nature of the sensed phenomena and resource constraints in the network.

5.2   Future Work

In networks using hierarchical aggregation and compression we have observed that by optimally exploiting the spatial organization of the hierarchy one can reap significant energy savings. Thus it may be worthwhile to develop simple network protocols permitting the realization of optimal compression and aggregation structures for gathering data from large networks of sensors.

In our work on proactive multipath routing, we note that our focus was on a preliminary analysis of proactive energy balancing. As such we have used a simplified energy model which is appropriate to the study of a routing scheme. Yet overheads associated with setting up multi-path routes, or other sources of energy expenditure or savings, e.g., putting nodes to sleep, will play a critical role in such

networks. For example, in our model we have for the most part ignored the MAC layer. In practice the temporal granularity on which load balancing is performed might be critical. For example, fine grain spreading of traffic might cause contention for transmission among neighboring paths lessening the energy saving benefits. Such interactions need to be studied carefully, and might be lessened by increasing the granularity of spreading. These and additional aspects of the proposed routing strategies are part of a more detailed analysis of the problem.

Finally possible extensions of our work on networks detecting bursty events include a study of the following conjecture: if the spreading scales are determined non-uniformly depending on load or distance to sinks, we expect additional improvement in the network lifetime. Note that the setup in our work is fairly *benign*, i.e., the traffic loads are spatially homogeneous. If there is additional inhomogeneity in the placement of nodes sensing events, variability in data size, etc., we conjecture that reducing variability in energy burdens will become a more critical issue.

Appendices

Appendix A

Appendix for Chapter 2

Proof of Theorem 1: We will show the feasible rate region is a contra-polymatroid, with rank function $f(S) = H(D_S | D_{S^c})$, $\forall S \subseteq U$. First note that $H(\emptyset | D_U) = 0$. To show that the monotonicity condition (2) holds, we note $\forall\ S_1 \subset S_2 \subseteq U$, we have

$$H(D_{S_1} | D_{S_1^c}) \leq H(D_{S_1} | D_{S_2^c}) \leq H(D_{S_2} | D_{S_2^c}).$$

The first inequality follows from the fact that $S_1^c \supseteq S_2^c$, and that conditioning reduces the entropy. The second inequality holds since $S_1 \subseteq S_2$, and thus $H(D_{S_2} | D_{S_2^c}) = H(D_{S_1} | D_{S_2^c}) + H(D_{S_2 \setminus S_1} | D_{S_1 \cup S_2^c}) \geq H(D_{S_1} | D_{S_2^c})$.

To check if the conditional entropy function is super-modular, i.e.,

$$H(D_{S_2} | D_{S_2^c}) + H(D_{S_1} | D_{S_1^c}) \leq H(D_{S_1 \cup S_2} | D_{(S_1 \cup S_2)^c}) + H(D_{S_1 \cap S_2} | D_{(S_1 \cap S_2)^c}),$$

we first establish the following identity,

$$
\begin{aligned}
H(D_{S_1 \cup S_2} | D_{(S_1 \cup S_2)^c}) &= H(D_{S_2 \setminus (S_1 \cap S_2)} | D_{(S_1 \cup S_2)^c}) \\
&\quad + H(D_{S_1} | D_{(S_2 \setminus (S_1 \cap S_2)) \cup (S_1 \cup S_2)^c}) \\
&= H(D_{S_2 \setminus (S_1 \cap S_2)} | D_{(S_1 \cup S_2)^c}) + H(D_{S_1} | D_{S_1^c}), \quad \text{(A.1)}
\end{aligned}
$$

since $(S_2 \setminus (S_1 \cap S_2)) \cup (S_1 \cup S_2)^c = S_1^c$. Now consider

$$
\begin{aligned}
& H(D_{S_1} | D_{S_1^c}) + H(D_{S_2} | D_{S_2^c}) \\
&= H(D_{S_1} | D_{S_1^c}) + H(D_{S_2 \setminus (S_1 \cap S_2)} | D_{S_2^c}) + H(D_{S_1 \cap S_2} | D_{(S_2 \setminus (S_1 \cap S_2)) \cup S_2^c}) \\
&= H(D_{S_1} | D_{S_1^c}) + H(D_{S_2 \setminus (S_1 \cap S_2)} | D_{S_2^c}) + H(D_{S_1 \cap S_2} | D_{(S_1 \cap S_2)^c}) \\
&\quad [\text{since } (S_2 \setminus (S_1 \cap S_2) \cup S_2^c = (S_1 \cap S_2)^c] \\
&\leq H(D_{S_1} | D_{S_1^c}) + H(D_{S_2 \setminus (S_1 \cap S_2)} | D_{(S_1 \cup S_2)^c}) + H(D_{S_1 \cap S_2} | D_{(S_1 \cap S_2)^c}) \\
&\quad [\text{since } S_2^c \supseteq (S_1 \cup S_2)^c] \\
&= H(D_{S_1 \cup S_2} | D_{(S_1 \cup S_2)^c}) + H(D_{S_1 \cap S_2} | D_{(S_1 \cap S_2)^c}).
\end{aligned}
$$

The inequality again follows from the fact that conditioning reduces entropy, and we use the intermediate result (A.1) in the last step.

Thus the feasible rate region constitutes a contra-polymatroid. From Lemma 1, we have that since $w_1 \geq w_2 \cdots \geq w_n$, the optimizing rates $(r_1^*, r_2^*, \cdots, r_n^*)$ satisfy:

$$
\begin{aligned}
r_1^* &= H(D_1|D_{\{1\}^c}) = H(D_1|D_2, D_3, \cdots, D_n), \\
r_i^* &= H(D_{\{1,2,\cdots,i\}}|D_{\{1,2,\cdots,i\}^c}) - H(D_{\{1,2,\cdots,i-1\}}|D_{\{1,2,\cdots,i-1\}^c}) \\
&= H(D_i|D_{\{1,2,\cdots,i\}^c}) + H(D_{\{1,2,\cdots,i-1\}}|D_{\{1,2,\cdots,i\}^c \cup \{i\}}) \\
&\quad - H(D_{\{1,2,\cdots,i-1\}}|D_{\{1,2,\cdots,i-1\}^c}) \\
&= H(D_i|D_{\{1,2,\cdots,i\}^c}) = H(D_i|D_{i+1}, D_{i+2}, \cdots, D_n), \quad \forall i = 2, 3, \cdots, n-1.
\end{aligned}
$$

$\blacksquare$

Proof of Theorem 5: We will use techniques similar to those in [56]. The following sequence of equalities capture the key steps which we will explain below:

$$
\mathbf{E_2^0}[|T_0^\alpha(\Pi_1, \Pi_2)|] = \mathbf{E_2^0}[\int_{\mathbb{R}^2} \mathbf{1}(x \in T_0^\alpha(\Pi_1, \Pi_2))dx] \tag{A.2}
$$

$$
= \mathbf{E_2^0}[\int_{\mathbb{R}^2} \mathbf{1}(B_{|x|}(x) \cap \Pi_2 = 1)\mathbf{1}(O_{|x|}^\alpha(0, x) \cap \Pi_1 = \emptyset)dx] \tag{A.3}
$$

$$
= \int_{\mathbb{R}^2} \mathbf{P}(B_{|x|}(x) \cap \Pi_2 = \emptyset)\mathbf{P}(O_{|x|}^\alpha(0, x) \cap \Pi_1 = \emptyset)dx \tag{A.4}
$$

$$
= \int_{\mathbb{R}^2} e^{-\pi\lambda_2|x|^2} e^{-\lambda_1 f(\alpha)|x|^2} dx \tag{A.5}
$$

$$
= \frac{\pi}{f(\alpha)\lambda_1 + \pi\lambda_2}. \tag{A.6}
$$

Eq. (A.2) is straightforward, since the area of a cell is the integral of the indicator function of the region. Eq. (A.3) simply says that for location x to belong to the cell two conditions must be in effect. First there must be no additional sinks (in addition to the one at the origin) within ball $B_{|x|}(x)$ of radius $|x|$ centered at x. Second there must be no compressors within the Cartesian oval $O_{|x|}(0, x)$, otherwise x would belong to the cell associated with that compressor rather than that corresponding to the sink at the origin. This second condition should be clear given our definition of the Cartesian oval and our cost function. These two requirements are shown in

Fig. 2.5. Eq. (A.4) uses the fact that the Palm probability $\mathbf{P_2^0}$ for a Poisson process Π_2 is simply that of the stationary Poisson process with same intensity but with an additional point at the origin - the so called Slivnyak's Theorem [53, 57]. In addition on Eq. (A.4) we use the fact that Π_1 is independent of Π_2. Eq. (A.5) explicitly replaces the void probabilities in Eq. (A.4). Recall that for a Poisson process Π with intensity λ and set A, the probability of no points arising in A is easily expressed as

$$\mathbb{P}(\Pi \cap A = \emptyset) = e^{-\lambda|A|}.$$

Also note that the area of a Cartesian oval $O_\beta^\alpha(0, x)$ in the special case where $\beta = |x|$ can be computed explicitly and is given by

$$|O_{|x|}^\alpha(0, x)| = |x|^2 f(\alpha)$$

where $f(\alpha)$ is given by (2.3). Finally Eq. (A.6) is a direct integration giving the desired result.

$\blacksquare$

Proof of Theorem 6: First by the Neveu exchange formula [24] we have that

$$\mathbf{E_2^0}[\sum_{x_j \in \Pi_0 \cap V_0(\Pi_2)} e_U(x_j, \Pi_1, \Pi_2)] = \frac{\lambda_0}{\lambda_2} \mathbf{E_0^0}[e_U(0, \Pi_1, \Pi_2)]$$

where $e_U(0, \Pi_1, \Pi_2)$ denotes the energy cost when the typical sensor at the origin eventually connects to the closest sink, directly or via a compressor. Now note that the energy cost $e_U(0, \Pi_1, \Pi_2)$ depends on whether our sensor chooses to connect directly to the sink or connects through a compressor. We shall consider these two cases separately. In order to send directly to a sink $s_k \in \Pi_2$ it must be the case that s_k is the closest to the sensor at the origin, i.e., $B_{|s_k|}(0) \cap \Pi_2 = \emptyset$ and there is no eligible compressor in the Cartesian oval $O_{|s_k|}^\alpha(s_k, 0)$, i.e., $O_{|s_k|}^\alpha(s_k, 0) \cap \Pi_1 = \emptyset$. Indeed any compressor within this oval would offer a cost which is less than $|s_k|$ and thus preclude direct connection to the sink s_k. We can formally compute the energy

113

cost associated with this first case as follows:

$$\mathbf{E_0^0}[\sum_{s_k \in \Pi_2} |s_k| \mathbf{1}(B_{|s_k|}(0) \cap \Pi_2 = \emptyset) \mathbf{1}(O_{|s_k|}^\alpha(s_k, 0) \cap \Pi_1 = \emptyset)]$$

$$= \mathbf{E}[\sum_{s_k \in \Pi_2} |s_k| \mathbf{1}(B_{|s_k|}(0) \cap \Pi_2 = \emptyset) \mathbf{1}(O_{|s_k|}^\alpha(s_k, 0) \cap \Pi_1 = \emptyset)] \qquad (A.7)$$

$$= \lambda_2 \int_{\mathbb{R}^2} |x| \mathbf{P_2^0}(B_{|x|}(-x) \cap \Pi_2 = \emptyset) \mathbf{P}(O_{|x|}^\alpha(-x, 0) \cap \Pi_1 = \emptyset) dx \qquad (A.8)$$

$$= \lambda_2 \int_{\mathbb{R}^2} |x| e^{-\lambda_2 \pi |x|^2 - \lambda_1 |O_{|x|}^\alpha(-x, 0)|} dx \qquad (A.9)$$

$$= 2\pi \lambda_2 \int_0^\infty r^2 e^{-(\lambda_2 \pi + \lambda_1 f(\alpha))r^2} dr \qquad (A.10)$$

$$= \frac{\lambda_2 \pi \Gamma(3/2)}{(\lambda_2 \pi + \lambda_1 f(\alpha))^{3/2}}. \qquad (A.11)$$

Here Eq. (A.7) follows from the independence of Π_1 Π_2 and Π_0. Eq. (A.8) follows by the refined Campbell's Theorem [53], and the independence of Π_0 and Π_1. In Eq. (A.9) we have explicitly replaced the void probabilities for the associated events. Eq. (A.10) corresponds to a change of variables, and finally Eq. (A.11) an explicit computation of the integral.

The second case is more complex. In this case the sensor at the origin would choose to transmit to a compressor, say c_j, which would relay the data to the sink s_k. Note under the proposed hierarchy the sink is that which is closest to the sensor, so $B_{|s_k|}(0) \cap \Pi_2 = \emptyset$. Note that in this case the energy cost will be $|c_j| + \alpha|c_j - s_k|$ and the compressor leading to the minimum energy cost will be selected. Thus it must be the case that $O_{|c_j| + \alpha|c_j - s_k|}^\alpha(s_k, 0) \cap \Pi_1 = \emptyset$ i.e., there is no alternative compressor that can achieve a lower cost. Also one must ensure that $|c_j| + \alpha|c_j - s_k| \leq |s_k|$ otherwise the sensor would minimize its energy cost by going directly to the sink. For simplicity let $d(c_j, s_k) = |c_j| + \alpha|c_j - s_k|$. We can estimate the average costs

incurred by sensors that choose to relay through compressors as follows.

$$\mathbf{E_0^0}[\sum_{s_k \in \Pi_2} \sum_{c_j \in \Pi_1} d(c_j, s_k)\mathbf{1}(B_{|s_k|}(0) \cap \Pi_2 = \emptyset)\mathbf{1}(O^\alpha_{d(c_j,s_k)}(s_k,0) \cap \Pi_1 = \emptyset)$$

$$\times \mathbf{1}(d(c_j, s_k) \leq |s_k|)]$$

$$= \mathbf{E}[\sum_{s_k \in \Pi_2} \sum_{c_j \in \Pi_1} d(c_j, s_k)\mathbf{1}(B_{|s_k|}(0) \cap \Pi_2 = \emptyset)\mathbf{1}(O^\alpha_{d(c_j,s_k)}(s_k,0) \cap \Pi_1 = \emptyset)$$

$$\times \mathbf{1}(d(c_j, s_k) \leq |s_k|)] \tag{A.12}$$

$$= \lambda_2 \int_{\mathbb{R}^2} \mathbf{E_2^0}[\sum_{c_j \in \Pi_1} d(c_j, x)\mathbf{1}(B_{|x|}(-x) \cap \Pi_2 = \emptyset)\mathbf{1}(O^\alpha_{d(c_j,x)}(x,0) \cap \Pi_1 = \emptyset)$$

$$\times \mathbf{1}(d(c_j, x) \leq |x|)]dx \tag{A.13}$$

$$= \lambda_2 \int_{\mathbb{R}^2} e^{-\lambda_2\pi|x|^2}\mathbf{E}[\sum_{c_j \in \Pi_1} d(c_j, x)\mathbf{1}(O^\alpha_{d(c_j,x)}(x,0) \cap \Pi_1 = \emptyset)\mathbf{1}(d(c_j, x) \leq |x|)]dx$$

$$\tag{A.14}$$

$$= \lambda_1\lambda_2 \int_{\mathbb{R}^2} \int_{\Delta_x} d(y, x)e^{-\lambda_2\pi|x|^2 - \lambda_1|O^\alpha_{d(y,x)}(x,0)|}dydx \tag{A.15}$$

where

$$\Delta_x = \{y \in \mathbb{R}^2 \mid \alpha|x| \leq d(y, x) \leq |x|\} = O^\alpha_{|x|}(x, 0).$$

Here Eq. (A.12) follows by the independence of Π_0 and Π_1, Π_2. Eq. (A.13) follows by the the refined Campbell's Theorem [53], In Eq. (A.14) we have explicitly replaced the void probability for the event associated with Π_2 and used the independence of Π_1 and Π_2. Finally in Eq. (A.15) we have applied Campbell's Theorem once more and explicitly computed void probability of the remaining event. We note however that with a sensor at position 0 and a sink at a location x, the energy cost function associated with using an intervening compressor is bounded by $\alpha|x| \leq |y| + \alpha|y - x| \leq |x|$ hence the range of integration under the integrals.

■

Proof of Theorem 7: We derive an upper bound of the cost incurred at the typical sensor, particularly when this typical sensor is within the cell of a compressor, we start from Eq. (A.15). If we rewrite the righthand side of Eq. (A.15), we have that

$$\lambda_1\lambda_2 \int_{\mathbb{R}^2} \left\{\int_{\Delta_x} d(y, x)e^{-\lambda_1|O^\alpha_{d(y,x)}(x,0)|}dy\right\} e^{-\lambda_2\pi|x|^2}dx \tag{A.16}$$

where

$$\Delta_x = \{y \in \mathbb{R}^2 \mid \alpha|x| \le d(y,x) \le |x|\} = O_{|x|}^\alpha(x,0).$$

Note that this expression is an upper bound of the optimal cost seen by a typical sensor at the origin. The reason is that this expression is obtained by the association rule (2.4), i.e., data of the typical sensor is eventually sent to the *closest sink* rather than the optimal association rule which minimizes the energy equation given in Eq. (2.1). Now let $u = d(y,x) = |y| + \alpha|x - y|$. We see that the inner integration of (A.16) is a function of u only, since x is treated as a constant in the integrand and $|O_{d(y,x)}^\alpha(x,0)| = |O_u^\alpha(x,0)|$ depends only on u. Furthermore, the region of integration is such that

$$\Delta_x = \{y \mid \alpha|x| \le u \le |x|\}.$$

which depends only on u for fixed x. Now let $S(u) = O_u^\alpha(x,0)$. The area of integration region $|\Delta_x|$ is $|S(|x|)|$ and consider the sequence of sets

$$S(u_0), S(u_0 + \Delta u), S(u_0 + 2\Delta u), S(u_0 + 3\Delta u), ..., S(u_1)$$

where $u_0 = \alpha|x|$, $u_1 = |x|$ and Δu is the step size. Note $S(u_l) \subseteq S(u_m)$ if $l \le m$. Let us partition the integration region Δ_x into a sequence of 'bands' by the following.

$$\Delta S_k = S(u_k + \Delta u) - S(u_k), \quad (u_k = u_0 + k\Delta u, \ k = 0, 1, ..., \frac{(1-\alpha)|x|}{\Delta u})$$

Then the sum

$$\sum_k u_k \exp\{-\lambda_1 |S(u_k)|\}|\Delta S_k| \tag{A.17}$$

is a Riemann sum and converges to the inner integral of (A.16), since as $\Delta u \to 0$, the integrand is approximately constant over the 'band' ΔS_k. Also we have, by chain rule,

$$|\Delta S_k| \simeq \left. \frac{\partial |S(u)|}{\partial u} \right|_{u=u_k} \cdot \Delta u = \left. \frac{\partial |O_u^\alpha(x,0)|}{\partial u} \right|_{u=u_k} \cdot \Delta u$$

as Δu vanishes. Thus the inner integral of (A.16) is equal to

$$\int_{\Delta_x} d(y,x) e^{-\lambda_1 |O_{d(y,x)}^\alpha(x,0)|} dy = \lim_{\Delta u \to 0} \sum_k u_k \exp\{-\lambda_1 |S(u_k)|\}|\Delta S_k|$$

$$= \int_{\alpha|x|}^{|x|} u e^{-\lambda_1 |O_u^\alpha(x,0)|} \frac{\partial |O_u^\alpha(x,0)|}{\partial u} du \tag{A.18}$$

116

If we apply (A.18) to (A.16), we have

$$\lambda_1\lambda_2 \int_{\mathbb{R}^2} e^{-\lambda_2\pi|x|^2} \int_{\alpha|x|}^{|x|} u e^{-\lambda_1|O_u^\alpha(x,0)|} \frac{\partial|O_u^\alpha(x,0)|}{\partial u} du dx$$

$$= \lambda_2 \int_{\mathbb{R}^2} e^{-\lambda_2\pi|x|^2} \left\{ -|x|e^{-\lambda_1 f(\alpha)|x|^2} + \alpha|x| + \int_{\alpha|x|}^{|x|} e^{-\lambda_1|O_u^\alpha(x,0)|} du \right\} dx$$

$$\text{(A.19)}$$

The integration in the brackets in (A.19) is still not directly calculable. We will use the following lower bound for the area of a Cartesian oval.

Fact 1. *The area of a Cartesian oval $O_u^\alpha(x,0)$, when $\alpha|x| \le u \le |x|$ and $0 \le \alpha \le 1$, is lower bounded by*

$$|O_u^\alpha(x,0)| \ge \frac{f(\alpha)}{(1-\alpha)^2}(u-\alpha|x|)^2$$

with $f(\alpha)$ as given by (2.3).

Note this bound is tight when $\alpha = 0$. Now we have that

$$\lambda_2 \int_{\mathbb{R}^2} e^{-\lambda_2\pi|x|^2} \left\{ -|x|e^{-\lambda_1 f(\alpha)|x|^2} + \alpha|x| + \int_{\alpha|x|}^{|x|} e^{-\lambda_1|O_u^\alpha(x,0)|} du \right\} dx$$

$$\le \lambda_2 \int_{\mathbb{R}^2} e^{-\lambda_2\pi|x|^2} \left\{ -|x|e^{-\lambda_1 f(\alpha)|x|^2} + \alpha|x| + \int_{\alpha|x|}^{|x|} e^{-\lambda_1 \frac{f(\alpha)}{(1-\alpha)^2}(u-\alpha|x|)^2} du \right\} dx$$

$$= \frac{\alpha}{2\sqrt{\lambda_2}} + \frac{1-\alpha}{2\sqrt{\lambda_2 + \lambda_1 f(\alpha)/\pi}} - \frac{\lambda_2\pi\Gamma(3/2)}{(\lambda_2\pi + \lambda_1 f(\alpha))^{3/2}}$$

Hence the total cost incurred by a typical sensor can be upper bounded as

$$G_{U0} = \underbrace{\frac{\lambda_2\pi\Gamma(3/2)}{(\lambda_2\pi + \lambda_1 f(\alpha))^{3/2}}}_{\text{avg. cost when the typical sensor is in a sink cell}}$$

$$+ \underbrace{\lambda_1\lambda_2 \int_{\mathbb{R}^2} \int_{\Delta_x} d(y,x) e^{-\lambda_2\pi|x|^2 - \lambda_1|O_{d(y,x)}^\alpha(x,0)|} dy dx}_{\text{avg. cost when the typical sensor is in a compressor cell}}$$

$$\le \frac{\lambda_2\pi\Gamma(3/2)}{(\lambda_2\pi + \lambda_1 f(\alpha))^{3/2}} + \frac{\alpha}{2\sqrt{\lambda_2}} + \frac{1-\alpha}{2\sqrt{\lambda_2 + \lambda_1 f(\alpha)/\pi}} - \frac{\lambda_2\pi\Gamma(3/2)}{(\lambda_2\pi + \lambda_1 f(\alpha))^{3/2}}$$

$$= \frac{\alpha}{2\sqrt{\lambda_2}} + \frac{1-\alpha}{2\sqrt{\lambda_2 + \lambda_1 f(\alpha)/\pi}}$$

$$= \hat{G}_{U0}$$

■

Proof of Theorem 8: The proof follows similar steps to Theorem 5, except that in this case, the void probability region of compressors is the circle centered at $\frac{z}{1+\alpha}$ with radius $(\frac{|z|}{1+\alpha})^2$, where the sink is fixed at the origin and $z \in \mathbb{R}^2$ being the position of a sensor. This circle is the trace of compressor locations that provide the same cost for the sensor as that for associating directly to the sink at the origin. This can be seen in Fig. A.1. The minimum cost function in this case is defined as $|z - c(z)|^2 + \alpha|c(z) - s_V(c(z))|^2$ where $c(z)$ is the location of the compressor or the sink that minimizes the above cost and $s_V(c(z))$ is the location of the sink that is closest to the node $c(z)$. Note the cost associated with a sensor at z to directly send to the origin is $|z|^2$. Then the set of compressor positions $(x, y) \in \mathbb{R}^2$ that provides less cost than $|z|^2$ is $\{(x,y)|(x-|z|)^2+y^2+\alpha(x^2+y^2) < |z|^2\}$, which is the inner circle in Fig. A.1, assuming z is on the x axis, i.e. $z = (|z|, 0)$ without loss of generality. So there must be no compressors in the inner circle otherwise the sensor will be associated with a compressor rather than directly sending to the sink. Also there must be no other sinks in the outer circle, which means the sink at the origin must be the closest sink, otherwise the sensor will be associated with the closer sink. Thus we have

$$
\begin{aligned}
\mathbf{E_2^0}[|T_0^{\alpha,2}(\Pi_1, \Pi_2)|] &= \mathbf{E_2^0}[\int_{\mathbb{R}^2} \mathbf{1}(x \in T_0^{\alpha,2}(\Pi_1, \Pi_2))dx] \\
&= \mathbf{E_2^0}[\int_{\mathbb{R}^2} \mathbf{1}(B_{|x|}(x) \cap \Pi_2 = 1)\mathbf{1}(B_{\frac{|x|}{1+\alpha}}(\frac{x}{1+\alpha}) \cap \Pi_1 = \emptyset)dx] \\
&= \int_{\mathbb{R}^2} \mathbf{P_2^0}(B_{|x|}(x) \cap \Pi_2 = \emptyset)\mathbf{P}(B_{\frac{|x|}{1+\alpha}}(\frac{x}{1+\alpha}) \cap \Pi_1 = \emptyset)dx \\
&= \int_{\mathbb{R}^2} e^{-\pi\lambda_2|x|^2}e^{-\pi\lambda_1(\frac{|x|}{1+\alpha})^2}dx \\
&= \frac{1}{\frac{\lambda_1}{(1+\alpha)^2} + \lambda_2}
\end{aligned}
$$

■

Proof of Theorem 9: Using the ideas in the proof of Theorem 8, we can see that the cost incurred by a typical sensor at the origin, when it is in the sink cell, is given

118
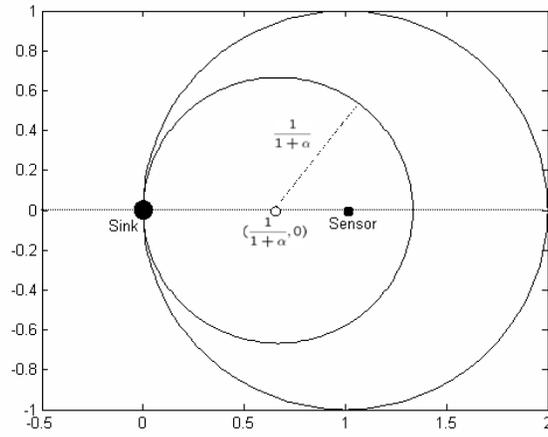
Figure A.1: The figure shows the void probability region for compressors, which is inside of the inner circle when $\gamma = 2$. A sink is at the origin and a sensor is at $(1, 0)$. For a given α the possible positions of compressors that provides same cost for a sensor to directly sending to a sink is the circle centered at $((1+\alpha)^{-1}, 0)$ with radius $(1+\alpha)^{-1}$. In order for this sensor to associate itself directly to the sink, there must be no other compressors in the inner circle and no other sinks in the outer circle.

by

$$\frac{\lambda_2}{\pi(\frac{\lambda_1}{(1+\alpha)^2} + \lambda_2)^2}. \tag{A.20}$$

We derive the expected cost of the case when the typical sensor is within a compressor cell, and under the suboptimal scheme (association rule Eq. (2.7)). Assume that the typical sensor is associated with a compressor at position $c_j \in \Pi_1$ and a sink at $s_k \in \Pi_2$. The cost function, denoted as $d(c_j, s_k)$, is given by

$$d(c_j, s_k) = |c_j|^2 + \alpha|c_j - s_k|^2.$$

There must be no other sinks within $B_{|s_k|}(0)$ otherwise s_k will not be the closest sink. Also there must be no other compressors in a region that can provide lower cost than c_j. One can show that this region is $B_{|c_j - \frac{\alpha}{1+\alpha}s_k|}(\frac{\alpha}{1+\alpha}s_k)$, which is a circle. Finally, $d(c_j, s_k)$ must be smaller than $|s_k|^2$, otherwise the sensor will associate itself directly with s_k rather than via c_j to s_k. Now the expected cost with respect to the Palm probability given the typical sensor at the origin under the suboptimal scheme, can be written as

$$\mathbf{E}_0^0[\sum_{c_j \in \Pi_1} \sum_{s_k \in \Pi_2} d(c_j, s_k)\mathbf{1}(B_{|s_k|}(0) \cap \Pi_2 = \emptyset)\mathbf{1}(B_{|c_j - \frac{\alpha}{1+\alpha}s_k|}(\frac{\alpha}{1+\alpha}s_k) \cap \Pi_1 = \emptyset)$$
$$\times \mathbf{1}(d(c_j, s_k) \leq |s_k|^2]. \tag{A.21}$$

Following the same guideline used in the proof of Theorem 6, (A.21) is equal to

$$\lambda_1\lambda_2 \int_{\mathbb{R}^2} \int_{\Delta_x} d(y, x)e^{-\pi\lambda_1|y - \frac{\alpha}{1+\alpha}x|^2} e^{-\pi\lambda_2|x|^2} dy dx$$

where $\Delta_x = \{y \mid d(y, x) \leq |x|^2\} = B_{\frac{|x|}{1+\alpha}}(\frac{\alpha x}{1+\alpha})$. This integration is rotationally invariant with respect to the position of the sink. Hence we express x with the polar

coordinates (ρ, θ) and express y in the Cartesian coordinates (u, v). Then

$$\lambda_1 \lambda_2 \int_{\mathbb{R}^2} \int_{\Delta_x} d(y, x) e^{-\pi\lambda_1 |y - \frac{\alpha}{1+\alpha}x|^2} e^{-\pi\lambda_2 |x|^2} dy dx \tag{A.22}$$

$$= \lambda_1 \lambda_2 \int_0^\infty \iint_{\Delta_{(\rho,0)}} (u^2 + v^2 + \alpha\{(u-\rho)^2 + v^2\}) e^{-\pi\lambda_1 \{(u - \frac{\alpha}{1+\alpha}\rho)^2 + v^2\}} du dv$$

$$\times e^{-\pi\lambda_2 \rho^2} 2\pi\rho d\rho \tag{A.23}$$

$$= \lambda_1 \lambda_2 \int_0^\infty \int_0^{\frac{\rho}{1+\alpha}} \{(1+\alpha)r^2 + \frac{\alpha}{1+\alpha}\rho^2\} e^{-\pi\lambda_1 r^2} 2\pi r dr \ e^{-\pi\lambda_2 \rho^2} 2\pi\rho d\rho \tag{A.24}$$

$$= \frac{\frac{\lambda_1}{(1+\alpha)^2} \left\{ \frac{\alpha\lambda_1}{(1+\alpha)^2} + (2\alpha+1)\lambda_2 \right\}}{\pi\lambda_2 (1+\alpha) \left\{ \frac{\lambda_1}{(1+\alpha)^2} + \lambda_2 \right\}^2}. \tag{A.25}$$

From Eq. (A.22) to Eq. (A.23), we changed x, y to polar coordinates and used the fact that the integration is invariant to θ. Hence we have fixed the position of the sink at $(\rho, 0)$ which is on the x-axis in Cartesian coordinates. From Eq. (A.23) to Eq. (A.24), we changed from the Cartesian coordinates (u, v) to the polar coordinates (r, ϕ) with the origin at $(\frac{\alpha\rho}{1+\alpha}, 0)$ in Cartesian coordinate. Using the fact the integrand depends only on r, we can scale the integrand with the proper Jacobian, hence the integration reduces to Eq. (A.24). Finally, if we add Eq. (A.20) and Eq. (A.25) to obtain the total cost at the typical sensor, we get Eq. (2.8). ∎

Appendix B

Appendix for Chapter 3

Proof of Lemma 3: We consider only the left half part of the geometry using the symmetric property of the problem. Let V be the grid set of the left half part, i.e. $V := \{(i,j) \mid 0 \leq i \leq l/2, \ |j| \leq \frac{w-1}{2}\}$. We have that

$$\sum_{i,j \in V} e_{i,j}^2 = \left\{ \sum_{k=0}^{(w-1)/2} \sum_{|i|+|j|=k} e_{i,j}^2 \right\} \tag{B.1}$$

$$+ \sum_{|i|+|j|>(w-1)/2} \{e_{i,j}^2\} \tag{B.2}$$

In (B.1), since $\sum_{|i|+|j|=k} e_{i,j} = 1$ regardless of k by the flow conservation, we have that, using Cauchy-Schwarz inequality

$$\sum_{|i|+|j|=k} e_{i,j}^2 \geq \frac{1}{2k+1} \left\{ \sum_{|i|+|j|=k} e_{i,j} \right\}^2 ,$$

and the equality is achieved when $e_{i,j} = \frac{1}{2k+1}$.

For (B.2), consider the case where $|i| + |j| < l/2$. Then considering the w nodes on $|i| + |j| = m$, the similar inequality holds and the minimum is achieved by setting the flow rates of those nodes equally to $1/w$. Consider the flow on the region $\mathcal{A} := \{(i,j) \mid |i| + |j| \geq l/2, \ (i,j) \in V\}$, let us consider the flow rates on the set P_m and Q_m which is defined by

$$P_m := \{(i,j) \mid |i| + |j| = l/2 + m, \ (i,j) \in V\}$$
$$Q_m := \{(l/2, j) \mid -m+1 \leq j \leq m-1\}$$

for $1 \leq m \leq (w-3)/2$ and let $A_m = P_m \cup Q_m$ where $\cup_m A_m = \mathcal{A}$. It is clear that there are $2m-1$ nodes in A_m which lies on MA, i.e., those in Q_m whose rate is fixed

to $1/w$. From flow conservation we have that

$$\sum_{(i,j)\in A_m/Q_m} e_{i,j} \geq 1 - \frac{2m-1}{w}.$$

From this, we again apply Cauchy-Schwarz inequality to find that setting the flow rates of nodes in A_m/Q_m should be set to $1/w$ in order to minimize $\sum_{(i,j)\in A_m/Q_m} e_{i,j}^2$. This holds for all m and we have exhaustively found the flow rates for all nodes in $\mathcal{A}$, which concludes the proof.

$\blacksquare$

Proof of Theorem 13: We have that the energy request arrival process is Poisson with rate λ per unit time per unit space. Since S_i is defined as the energy request in unit time interval, S_i is stochastically equivalent to the shot-noise process in $\mathbb{R}^2$ with intensity λ. From Lemma 2, the n-th order cumulant of S_i is $\chi^{(n)}(1)$, thus we have that

$$C(\theta) = \sum_{n=1}^{\infty} \chi^{(n)} \frac{\theta^n}{n!} = \lambda \sum_{n=1}^{\infty} E\left[\int_{\Phi_0} h(x, \Phi_0)^n dx\right] \frac{\theta^n}{n!}$$

$$= \lambda E\left[\int_{\Phi_0} \sum_{n=1}^{\infty} h(x, \Phi_0)^n \frac{\theta^n}{n!} dx\right]$$

$$= \lambda E\left[\int_{\Phi_0} \{e^{\theta h(x,\Phi_0)} - 1\} dx\right].$$

$\blacksquare$

Appendix C

Appendix for Chapter 4

Proof of Proposition 4:

We have that

$$\mathbf{E}_1^0 \left[\sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in V_0(\Pi_1) \cap \Phi) \right]^2$$

$$= \mathbf{E}_1^0 \sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in V_0(\Pi_1) \cap \Phi)$$

$$+ \sum_{\substack{x_j, x_k \in V_0(\Pi_1) \cap \Pi_0 \\ y_j \neq y_k}} \mathbf{1}(x_i, x_j \in V_0(\Pi_1) \cap \Phi) \qquad (\text{C.1})$$

The first term of the above is already shown to be $\frac{\lambda_0}{\lambda_1} p$. Using Mecke's formula [53] the second term is given by

$$\lambda_0^2 \iint \mathbf{E}_0^{x,y} \mathbf{1}(x, y \in V_0(\Pi_1) \cap \Phi) dx dy$$

$$= \lambda_0^2 \iint \mathbf{E}_0^{x,y} \mathbf{1}(x, y \in V_0(\Pi_1)) \mathbf{P}(x, y \in \Phi) dx dy$$

$$= \lambda_0^2 \iint \mathbf{P}(\{B_{|x|}(x) \cup B_{|y|}(y)\} \cap \Pi_1 = \emptyset) \mathbf{P}(x, y \in \Phi) dx dy. \qquad (\text{C.2})$$

Using the (non-centered) covariance results for stationary Boolean model [53], we have that

$$\mathbf{P}(x, y \in \Phi) = 2p - 1 + (1 - p)^2 \exp(\lambda_e C(r_e, |x - y|))$$

where the function $C(r, d)$ is defined as the area of the intersection of two balls of radius r whose centers are separated by distance d if $d \leq 2r$ and 0 otherwise. If we

divide $P(x, y \in \Phi)$ by p and take the limit as $\lambda_e \to 0$, we have that

$$
\begin{aligned}
\lim_{\lambda_e \to 0} \frac{\mathbf{P}(x, y \in \Phi)}{p} \\
= \lim_{\lambda_e \to 0} \frac{1 - 2e^{-\lambda_e \pi r_e^2} + e^{-2\lambda_e \pi r_e^2 + \lambda_e C(r_e, |x-y|)}}{1 - \exp(-\lambda_e \pi r_e^2)} \\
= \frac{C(r_e, |x - y|)}{\pi r_e^2},
\end{aligned}
\tag{C.3}
$$

using L'Hospital's rule. Thus we have that, with λ_e vanishing and by letting $D(x, y) := |\{B_{|x|}(x) \cup B_{|y|}(y)\}|$,

$$
\begin{aligned}
\lambda_0^2 \iint \mathbf{P}(\{B_{|x|}(x) \cup B_{|y|}(y)\} \cap \Pi_1 = \emptyset)\mathbf{P}(x, y \in \Phi)dxdy \\
\approx p\lambda_0^2 \iint \exp(-\lambda_1 D(x, y))\frac{C(r_e, |x-y|)}{\pi r_e^2}dxdy \\
= p\frac{\lambda_0^2}{\lambda_1^2} \iint \exp(-D(x, y))\frac{C(r_e, |x-y|/\sqrt{\lambda_1})}{\pi r_e^2}dxdy
\end{aligned}
\tag{C.4}
$$

Since (C.4) cannot be reduced to closed form, we use the following technique to obtain an upper bound. Firstly it is easy to verify that

$$
\frac{C(r_e, |x - y|/\sqrt{\lambda_1})}{\pi r_e^2} \leq \exp(-\frac{|x - y|}{2\sqrt{\lambda_1}r_e}).
$$

Let

$$
g(\rho) := \int \int \exp(-D(x, y))\exp(-\rho|x - y|)dxdy.
$$

where $\rho := 1/(2\sqrt{\lambda_1}r_e)$. We make change of variables by letting $|x| = r_1, |y| = r_2, |x - y| = r_3$ then note this can be expressed in triple integral of r_1, r_2, r_3. For simplicity of notation denote $D(x, y)$ by $\bar{D}(\mathbf{r})$ where $\mathbf{r} := (r_1, r_2, r_3)$ and denote the triple integral $\iiint dr_1 dr_2 dr_3$ by $\int d\mathbf{r}$. One can verify that $dxdy$ can be written as $\chi(\mathbf{r})d\mathbf{r}$ where

$$
\chi(\mathbf{r}) := 2\pi r_1 r_2 \frac{\partial}{\partial r_3} \arccos\left(\frac{r_1^2 + r_2^2 - r_3^2}{2r_1 r_2}\right).
$$

Then we can rewrite $g(\rho)$ as

$$
g(\rho) = \int \exp(-\bar{D}(\mathbf{r}))\exp(-\rho r_3)\chi(\mathbf{r})d\mathbf{r}.
$$

We have the following lemma.

Lemma 6. *Let $h(\mathbf{r}) = \exp(-\bar{D}(\mathbf{r}))\chi(\mathbf{r})$. Suppose there exists a positive constant k such that*

$$\frac{\int r_3 h(\mathbf{r})d\mathbf{r}}{\int h(\mathbf{r})d\mathbf{r}} < \frac{\int r_3 h(\mathbf{r})\exp(-kr_3)d\mathbf{r}}{\int h(\mathbf{r})\exp(-2kr_3)d\mathbf{r}} \tag{C.5}$$

holds. Then

$$g(\rho) \leq \frac{1}{1 + c\rho}\int h(\mathbf{r})d\mathbf{r} \tag{C.6}$$

holds for $\rho > 0$ where

$$c := \frac{\int r_3 h(\mathbf{r})d\mathbf{r}}{\int h(\mathbf{r})d\mathbf{r}}. \tag{C.7}$$

Proof. Let $\gamma(\rho) := (1 + c\rho)g(\rho)$, and since $\gamma(0) = \int h(\mathbf{r})d\mathbf{r}$, it suffices to show that $\gamma(\rho)$ is nonincreasing for $\rho > 0$. Let

$$f(r_3, \rho) := \exp(-r_3\rho)(1 + c\rho).$$

Since

$$\int h(\mathbf{r})\frac{\partial f}{\partial \rho}d\mathbf{r}$$

is uniformly convergent in ρ and $h(\rho)$ is finite for all $\rho > 0$, the following relation for derivatives hold[58]:

$$\frac{d\gamma}{d\rho} = \int h(\mathbf{r})\frac{\partial f}{\partial \rho}d\mathbf{r}$$

$$= \int h(\mathbf{r})\exp(-\rho r_3)(c - r_3(1 + c\rho))d\mathbf{r}$$

We will show that $\frac{d\gamma}{d\rho}$ is negative for all $\rho > 0$. Since

$$\frac{d\gamma}{d\rho} = \int h(\mathbf{r})\exp(-\rho r_3)(c - cr_3\rho - r_3))d\mathbf{r} \tag{C.8}$$

$$\leq \int h(\mathbf{r})\exp(-\rho r_3)(c\exp(-\rho r_3) - r_3)d\mathbf{r} \tag{C.9}$$

$$= c\int h(\mathbf{r})\exp(-2\rho r_3)d\mathbf{r} - \int h(\mathbf{r})r_3\exp(-\rho r_3)d\mathbf{r} \tag{C.10}$$

126

where in (C.9) we have used $e^{-x} \geq 1 - x$ for all $x \geq 0$. Let us define the following:

$$f_1(\rho) := \int h(\mathbf{r}) \exp(-2\rho r_3) d\mathbf{r}, \qquad (C.11)$$

$$f_2(\rho) := \int r_3 h(\mathbf{r}) \exp(-\rho r_3) d\mathbf{r}, \qquad (C.12)$$

then we can write (C.10) as $c f_1(\rho) - f_2(\rho)$. Since $f_1(\rho)$ and $f_2(\rho)$ are strictly decreasing function of ρ, the equation $c f_1(\rho) = f_2(\rho)$ has at most one solution. By definition of c we have that $c f_1(0) = f_2(0)$. Thus for $\rho > 0$, $c f_1(\rho) - f_2(\rho)$ is either strictly positive or negative.

However by (C.5) implies there exists a constant[1] $k > 0$ such that $c f_1(k) - f_2(k) < 0$, thus we can conclude that $c f_1(\rho) - f_2(\rho)$ is negative for all $\rho > 0$. Thus $\frac{d\gamma}{d\rho}$ is negative for all $\rho > 0$ which implies $\gamma(\rho)$ is decreasing in ρ, which completes the proof. $\qquad \square$

It can be easily seen that $\int h(\mathbf{r}) d\mathbf{r}$ is the second moment of the area of the typical Voronoi cell induced by a stationary point process of density 1, which gives ≈ 1.28. Also numerical evaluation yields $c \approx 0.66$. Combining (C.6) and (C.1) we have proven Proposition 4. $\qquad \blacksquare$

Proof of Proposition 5:

Note

$$\mathbf{E}\left[\mathbf{1}(O \in \Phi) \sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_0(\Pi_1))\right]$$

$$= \mathbf{E} \sum_{x_i \in \Pi_0} \mathbf{1}(\{x_i, O\} \in \Phi \cap V_0(\Pi_1))$$

$$= 2\pi\lambda_0 \int \mathbf{E}_0^x \mathbf{1}(\{x, O\} \in \Phi \cap V_0(\Pi_1)) dx$$

$$= 2\pi\lambda_0 \int \mathbf{E}_0^x \mathbf{1}(x \in V_0(\Pi_1)) \mathbf{E} \mathbf{1}(\{x, O\} \in \Phi) dx$$

$$= 2\pi\lambda_0 \int \exp(-\pi\lambda_1 |x|^2)$$
$$\times \{2p - 1 + (1 - p)^2 \exp(\lambda_e C(r_e, |x|))\} dx.$$

---

[1] One can verify (C.5) holds for $k = 1$ via numerical evaluation.

By evaluating the first order term with respect to p in the way similar to (C.3), we have that

$$2\pi\lambda_0 p \int \exp(-\pi\lambda_1|x|^2)\frac{C(r_e,|x-y|)}{\pi r_e^2}dx \tag{C.13}$$

$$\leq 2\pi\lambda_0 p \int_0^{2r_e} \rho\exp(-\pi\lambda_1\rho^2)\{1-\frac{\rho}{2r_e}\}d\rho \tag{C.14}$$

$$= \frac{\lambda_0 p}{\lambda_1}\{1-\frac{\mathrm{erf}(2\pi r_e\sqrt{\lambda_1})}{4r_e\sqrt{\lambda_1}}\} \tag{C.15}$$

where in (C.14) we have used the fact that

$$\frac{C(r_e,\rho)}{\pi r_e^2} \leq (1-\frac{\rho}{2r_e})\mathbf{1}(\rho\leq 2r_e)$$

by the convexity of $\frac{C(r_e,\rho)}{\pi r_e^2}$ with respect to ρ. ∎

Bibliography

[1] J.M. Kahn, R.H. Katz, and K.S. Pister, "Next century challenges: Mobile networking for smart dust," in *Proc. MOBICOM*, 1999.

[2] D. Estrin, R. Govindan, J. Heidmann, and S. Kumar, "Next century challenges: Scalable coordination in sensor netwroks," in *Proc. MOBICOM*, 1999.

[3] I.F. Akyldiz, W. Su, Y. Sankarasubermanian, and E. Cayirici, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, August 2002.

[4] A.J. Goldsmith and S.B Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wireless Communications*, vol. 9, no. 4, 2002.

[5] A. Chandrakasan *et. al.*, "Design considerations for distributed microsensor systems," in *Proc. IEEE CICC*, May 1999.

[6] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Hawaii International Conference on System Sciences*, 2000.

[7] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage probability specifications," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, January 2002.

[8] V. Rodoplu and T. H. Meng, "Minimum energy mobile wireless networks," *IEEE JSAC*, vol. 17, no. 8, pp. 1333–1344, 1999.

[9] T. Cover and J. Thomas, *Elements of information theory*, John Wiley & Sons, 1991.

[10] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes: design and construction," in *Proc. of the Data Compression conference*, March 1999.

[11] L. Doherty J. Kusuma and K. Ramachandran, "Distributed compression for sensor networks," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2001.

[12] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes: symmetric rates and applications to sensor networks," in *Proc. of the Data Compression conference*, March 2000.

[13] S. S. Pradhan, J. Kusuma, and K. Ramachandran, "Distributed compression in a dense mircosensor network," *IEEE Signal Processing Magazine*, pp. 51–60, March 2002.

[14] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer, Norwell, MA, 1992.

[15] D. Petrovic, R.C. Shah, , K. Ramchandran, and J. Rabaey, "Data funneling: Routing with aggregation and compression for wireless sensor networks," in *Proc. IEEE Sensor Network Protocols and Applications*, May 2003.

[16] S. Singh, M. Woo, and C. Raghavendra, "Power-aware routing in mobile ad hoc networks," in *Proc. ACM Mobicom*, October 2002.

[17] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic media access for multirate ad hoc networks," in *Proc. ACM Mobicom*, September 2002.

[18] A. Scaglione and S. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," in *Proc. ACM Mobicom*, 2002, pp. 140–147.

[19] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Info. Theory*, vol. 46, no. 2, March 2000.

[20] J. Chou, D. Petrovic, and K. Ramchandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in *Proc. IEEE Infocom*, 2003.

[21] Seema Bandyopadhyay and Edward J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *Proc. IEEE Infocom*, 2003.

[22] J. Edmonds, "Submodular functions, matroids and certain polyhedra," in *Proc. Calgary Int. Conf. Combinatorial Structures and Applications*, 1970, pp. 69–87.

[23] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Information Theory*, vol. 22, pp. 226–228, 1975.

[24] F. Baccelli and S. Zuyev, "Poisson-voronoi spanning trees with applications to the optimization of communication networks," *INRIA Research Report No.3040*, Nov. 1996.

[25] D. Petrovic, R. C. Shah, K. Ramchandran, and J. Rabaey, "Data funneling: Routing with aggregation and compression for wireless sensor networks," in *IEEE ICC Int. Workshop on Sensor Network Protocols and Applications*, 2003.

[26] T. S. Rappaport, *Wireless Communications, Principles and Practices*, Prentice Hall, 2002.

[27] J. Moller, "Random Johnson-Mehl tessellations," *Adv. Appl. Prob.*, vol. 24, pp. 814–844, 1992.

[28] F. Baccelli and P. Brémaud, *Elements of queueing theory*, Springer-Verlag, 1994.

[29] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, no. 5, pp. 51–58, May 1996.

[30] S. Foss and S. Zuyev, "On a certain voronoi aggregative process related to a bivariate poisson process," *Adv. Appl. Probab.*, vol. 28, pp. 965–981, 1996.

[31] V. Rodoplu and T. Meng, "Minimum energy mobile wireless networks," *IEEE Jounal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1333–1344, Aug. 1999.

[32] J. Chang and L. Tassiulas, "Maximum lifetime routing in wireless sensor networks," *IEEE Transactions on Networking*, vol. 12, no. 4, pp. 609–619, Aug. 2004.

[33] C. Toh, H. Cobb, and D. Scott, "Performance evaluation of battery-life-aware routing schemes for wireless ad hoc networks," in *Proceedings of ICC*, June 2001, number 1, pp. 2824–2829.

[34] K. Kar, M. Kodialam, T. Lakshman, and L. Tassiulas, "Routing for network capacity maximization in energy-constrained ad-hoc networks," in *Proceedings of IEEE INFOCOM*, Mar. 2003, vol. 22, pp. 673–681.

[35] L. Lin, N. Shroff, and R. Srikant, "Asymptotically optimal power-aware routing for multihop wireless networks with renewable energy sources," in *Proceedings of IEEE INFOCOM*, 2005, to be published.

[36] S. Servetto and G. Barrenechea, "Constrained random walks on random graphs: Routing algorithms for large scale wireless sensor networks," in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, 2002, pp. 12–21.

[37] M. Keil and C. Gutwin, "Classes of graphs which approximate teh complete euclidean graph," *Discrete Comput. Geom.*, vol. 7, pp. 13–28, 1992.

[38] F. Baccelli, K. Tchoumatchenko, and S. Zuyev, "Markov paths on the Poisson-Delaunay graph with applications to routing in mobile networks," *Adv. Appl. Probab.*, vol. 32, pp. 1–18, 2000.

[39] L. Heinrich and V. Schmidt, "Normal convergence of multidimensional shot noise and rates of this convergence," *Adv. Appl. Prob.*, vol. 17, pp. 709–730, 1985.

[40] D. K. Cheng, *Field and Wave Electrodynamics*, Addison-Wesley, Reading, MA, second edition, 1989.

[41] S. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, New York, 1996.

[42] D. Aldous, *Probability approximations via the Poisson clumping heuristic*, Springer-Verlag, New York, 1989.

[43] R. Adler, *The Geometry of Random Fields*, Wiley, New York, 1981.

[44] J. Cohen, "The M/G/1 fluid model with heavy-tailed message length distributions," Tech. Rep. PNA-R9714, Centrum voor Wiskunde en Informatica, Technical Report, 1997.

[45] W. Whitt and P. Glynn, "Logarithmic asymptotics for steady-state tail probabilities in a single-server queue," *Studies in Applied Probability, Papers in Honour of Lajos Takacs, J. Galambos and J. Gani (eds.), Applied Probability Trust*, pp. 131 – 156, 1994.

[46] G. de Veciana and J. Walrand, "Effective bandwidths: Call admission, traffic policing and filtering for ATM networks," *Queueing Systems*, vol. 20, pp. 37–59, 1995.

[47] X. Su and G. de Veciana, "Predictive routing to enhance QoS for stream-based flows sharing excess bandwidth," *Computer Networks*, , no. 42, pp. 65–82, May 2003.

[48] J. Luo and J. Hubaux, "Joint mobility and routing for lifetime elongation in wireless sensor networks," in *Proceedings of IEEE INFOCOM*, 2005, pp. 1735–1746.

[49] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *Proceedings of IEEE INFOCOM*, 2001, pp. 1360–1369.

[50] R. Shah, S. Roy, S. Jain, and W. Brunette, "Data MULEs: Modeling a three-tier architecture for sparse sensor networks," in *IEEE SNPA Workshop*, May 2003.

[51] W. Wang, V. Srinivasan, and K. Chua, "Using mobile relays to prolong the lifetime of wireless sensor networks," in *Proceedings of ACM MOBICOM*, 2005, pp. 270–283.

[52] D. Slepian and J. Wolf, "Noiseless encoding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, July 1973.

[53] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, J. Wiley & Sons, Chichester, 1995.

[54] S. Baek and G. de Veciana, "Spatial energy balancing through proactive multi-path routing in multi-hop wireless networks," Tech. Rep., Wireless Networking and Communications Group (WNCG), University of Texas at Austin, TX. [Online] Available: http://www.ece.utexas.edu/~gustavo, 2004.

[55] S. Toumpis and L. Tassiulas, "Packetostatics: Deployment of massively dense sensor networks as an electrostatics. problem," in *IEEE INFOCOM*, 2005.

[56] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev, "Stochastic geometry and architecture of communication networks," *J. Telecommunication Systems*, vol. 7, pp. 209–227, 1997.

[57] J. Moller, "Random tessellations in $\mathbb{R}^d$," *Adv. Appl. Prob.*, vol. 21, pp. 37–73, 1989.

[58] I. Gradsteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, Academic, New York, 1965.

Vita

Seung-Jun Baek has received the Bachelor of Science degree in Electrical Engineering from the Seoul National University in 1999, and Master of Science degree in Electrical and Computer Engineering in The University of Texas at Austin in 2003. He has joined DSP Research and Development group of Texas Instruments, Inc. in 2006 where he works as a system architect engineer for broadband communication networks.

Permanent address: 800 W. Renner Rd. Apt. #3218, Richardson, TX,
75080

This dissertation was typeset with LaTeX[†] by the author.

Vita

Seung-Jun Baek has received the Bachelor of Science degree in Electrical Engineering from the Seoul National University in 1999, and Master of Science degree in Electrical and Computer Engineering in The University of Texas at Austin in 2003. He has joined DSP Research and Development group of Texas Instruments, Inc. in 2006 where he works as a system architect engineer for broadband communication networks.

Permanent address: 800 W. Renner Rd. Apt. #3218, Richardson, TX,
75080

This dissertation was typeset with LaTeX[†] by the author.