

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

Seung Jun Baek and Gustavo de Veciana, *Senior Member, IEEE*

Abstract—In this paper we propose a stochastic geometric model to study the energy burdens seen in a large scale hierarchical sensor network. The network makes use of aggregation nodes, for compression, filtering and/or data fusion of locally sensed data. Aggregation nodes (AGNs) then relay the traffic to mobile sinks. While aggregation may substantially reduce the overall traffic on the network it may have the deleterious effect of concentrating loads on paths between AGNs and the sinks – such inhomogeneities in the energy burden may in turn lead to nodes with depleted energy reserves. To remedy this problem we consider how one might achieve a more balanced energy burden across the network by spreading traffic, i.e., using a multiplicity of paths between AGNs and sinks. The proposed model reveals, how various aspects of the task at hand impact the characteristics of energy burdens on the network and in turn the lifetime for the system. We show that the scale of aggregation and degree of spreading can be optimized. Additionally if the sensing activity involves large amounts of data flowing to sinks, then inhomogeneities in the energy burdens seen by nodes around the sinks will be hard to overcome, and indeed the network appears to scale poorly. By contrast if the sensed data is bursty in space and time, then one can reap substantial benefits from aggregation and balancing.

Index Terms—Stochastic geometry, Boolean model, sensor networks, data fusion

I. INTRODUCTION

IN large-scale wireless ad-hoc/sensor networks enabled by multi-hop relaying, it is critical to provide mechanisms to conserve energy due to nodes' limited capabilities to store and replenish energy. Additional challenges may result from inhomogeneous spatial patterns in traffic loads resulting in uneven energy burdens on nodes which shorten network lifetime. We 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 relaying data to an information sink. Unfortunately, such a hierarchy faces an intrinsic problem: the nodes close to sinks will see a disproportionate energy burden as they will see higher loads

Manuscript received XXXX XX, xxxx; revised XXXX xx, xxxx; approved by IEEE TRANSACTIONS ON INFORMATION THEORY Editor X. XXXXX. This work is supported in part by the National Science Foundation under Grant ECS-0225448. An earlier version of this paper has been presented in IEEE SPASWIN 2006. S. Baek is with Communication Systems Laboratory, DSP Solutions R&D Center, Texas Instruments (e-mail: sbaek@ti.com). G. de Veciana is with WNCG (Wireless Networking Communication Group), ECE Dept. of University of Texas at Austin (e-mail: gustavo@ece.utexas.edu).

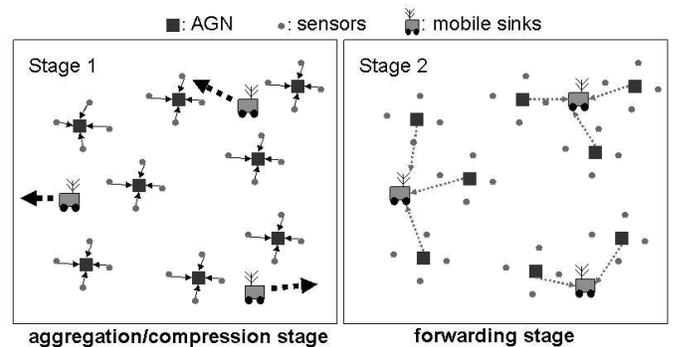


Fig. 1. Stages of operation for sensor network with mobile sinks.

of traffic that flows to sinks. One solution is to increase the density of sinks, however this may end up being too costly. In this paper 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.

A second possibility is to make sinks *mobile*, i.e., have sinks change their locations to balance the energy burdens incurred across the network nodes [1]. As shown in Fig. 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 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 specific 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. In this case the sink mobility increases energy efficiency and throughput capacity [2] substantially.

By contrast let us consider application scenarios where the

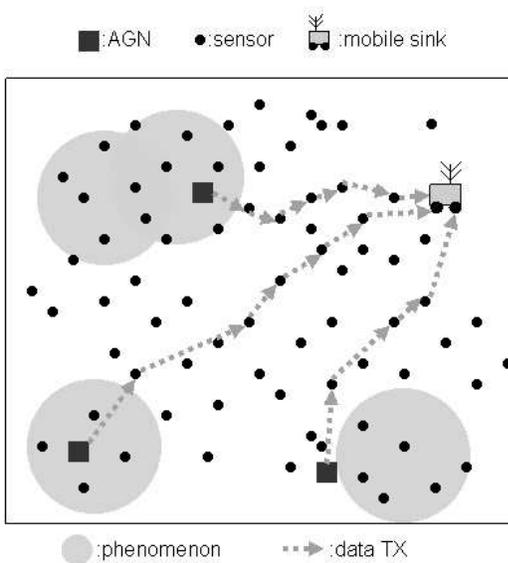


Fig. 2. Bursty phenomena and traffic concentration.

‘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.2 shows an example, where only 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 paper is on modeling 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 paper we devise a spatial model for traffic aggregation and spreading. It is without a doubt a very simple caricature, based on first-order models 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 provide numerical study for the joint optimization of the spatial scales for aggregation and spreading so as to maximize the network lifetime.

This paper is organized as follows: in Section II we discuss related work. In Section III we briefly present our model and assumptions. In Section IV we derive the 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 V. Finally we conclude with Section VI.

II. RELATED WORK

There has been a substantial research effort on energy conserving routing for large-scale ad hoc/sensor networks, for different network scenarios and contexts.

One of the key approaches involves exploiting the correlation structure in sensed data for compression. In [3], [4] compression and routing models are proposed to evaluate theoretical limits on what can be achieved. Also methods to efficiently estimate correlation and perform source coding are proposed in [5], [6], [7]. In this paper we focus more generally on ‘data fusion’ where certain simple filtering and/or processing of locally sensed data is involved, possibly mandated by a given application. We believe this can be efficiently realized if one exploits local aggregation of traffic. Furthermore we consider sensed phenomena such that, either the statistical properties are poorly characterized, or may vary unpredictably. Thus if one wishes to exploit spatial correlations in compressing traffic, it must be performed ‘on-the-fly’ with the data at hand, in which case it makes sense to use local aggregation of sensed data.

The use of mobile relays (or base stations) in large sensor networks was proposed and studied in [8], [9], [1]. The idea was to deploy automated robots to solve scalability problems in sensor networks. Particularly in [1] 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 paper 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 this paper is closely related to our previous work [10] 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. In this paper we

will leverage these results in modelling the spatial patterns of structured, i.e., hierarchical aggregated, traffic. Finally this work is inspired by studies modelling network structure via hierarchies of Voronoi tessellations. Notably modelling of telecommunication networks using stochastic geometry was proposed in [11] to successfully analyze the cost of a network with a hierarchy associated with proximity – we will borrow their framework and notation as appropriate in this paper.

III. MODELS AND ASSUMPTIONS

A. 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 reduced amount of data after the aggregation of n packets. We assume $f(n)$ has the following properties:

- 1) $f(n) \leq n$,
- 2) $\lim_{n \rightarrow \infty} \frac{f(n)}{n} = \alpha$, $0 < \alpha < 1$,
- 3) $\frac{f(n)}{n}$ is a monotonically nonincreasing function of n .

The second assumption captures an asymptotic ratio for the possible reductions in traffic. The parameter α which we refer to as the *compression ratio*, represents the efficiency of the AGN 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 assumptions may apply to generic classes of sensing applications which use compression, filtering and/or data fusion. An example of $f(n)$ can be the joint entropy of n identically distributed discrete random variables whose entropy rate is α .

Among the possible choices of $f(n)$, we define the following for simplicity:

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

It should be noted that $f(n)$ is close to but not linear since $f(1) = 1$, i.e., the size of a single sample cannot be further reduced by itself, otherwise the data fusion of multiple samples is meaningless.

Although we use a specific $f(n)$ for the analysis in this work, 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, in a sensor network in which 5 monitoring sensors see the same event, and report it to their associated AGN, the AGN may choose to filter the data, and send a single warning to a sink, achieving a 5:1 compression ratio.

B. Network Model

1) *Model for Network Operation*: We describe the model for network operation in further detail, recall Fig. 1. There are three levels of hierarchy in the network organization: sensors, AGNs and sinks where a node can act as a sensor or an AGN. The network periodically repeats the following procedure which we refer to as a *round*. A round consists of two stages:

- 1) *Aggregation/compression stage*: If a sensor detects phenomena, it forwards the data to the nearest AGN, or equivalently, if we consider a Voronoi tessellation induced by the locations of AGNs and a sensor is within the cell of an AGN, then the sensor reports to that AGN. The AGNs compress and store the received data. Meanwhile the sinks randomly move around. We denote the time duration of this stage by t_c which is assumed to be constant.
- 2) *Forwarding stage*: All the sinks stop moving and broadcast their location. The AGNs forward compressed data towards the nearest sink. We denote the time duration of this stage by t_f which is assumed to be constant.

There is a high density of possibly redundant nodes in the region of interest. On each round a certain fraction of nodes are put into the ‘sleep’ mode while the rest are assumed to be in the ‘active’ mode. Among the active nodes a certain fraction become AGNs, and the rest of them act as sensors and/or relays. These nodes are active for $t_c + t_f$, i.e., the duration of a round. The randomized active node selection resumes in the next round independent of previous rounds. To provide a consistent coverage for the phenomena of interest as well as relaying capacity the sum of spatial densities of sensors and AGNs is constant in each round.

We assume that t_f , the data collection time at the sinks, is much less than t_c , and that delay requirement for sensing is met by these timescales. In other words, there is a sufficient margin between t_c and the delay tolerance of the system. Furthermore the sink mobility is sufficiently random such that the energy burden concentration problem is effectively mitigated over rounds.

2) *Spatial Model for Network and Sensed Phenomena*: We will assume the locations of active sensors and AGNs constitute independent homogeneous Poisson point processes (PPP) at each round. It is assumed that the locations of sinks during the forwarding stage of each round also constitute a PPP which has fixed intensity and is assumed to be *independent* across rounds¹. The density of sinks is assumed to be much smaller than that of sensors so that a sink is on average associated with hundreds or thousands of sensors simply due to its equipment cost.

We model the spatio-temporal phenomena being sensed by a *Boolean model* [12]. 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 intensity λ_e and a primary grain Φ_0 given by a circular disc with the radius r_e .

¹This is a modelling assumption – in reality the nodes will observe the same set of sinks which merely have changed their location over rounds.

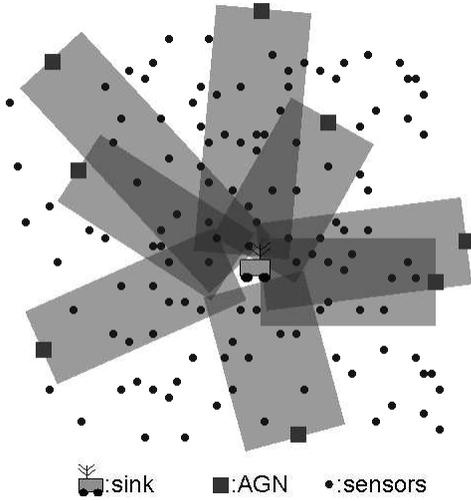
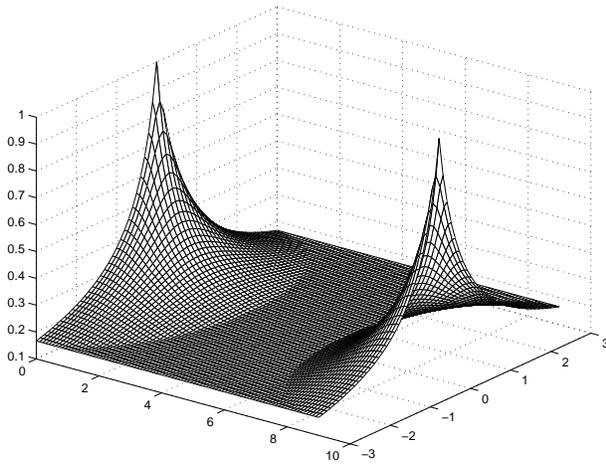


Fig. 3. Strip model.

Fig. 4. Energy burden density $h(\cdot, \cdot, \cdot)$ on a strip.

Specifically Φ is given by

$$\Phi := \bigcup_{x \in \Pi_e} (\Phi_0 + x), \quad (2)$$

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 IV.

C. Spatial Model for Traffic and 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. A routing model for spatially spreading traffic is proposed as follows. We 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. Specifically we assume traffic is spread over rectangular ‘strips’ of certain width which represents the spatial scale of spreading. Fig. 3 depicts the strip model.

In fact energy burdens are not homogenous over a strip. We associate a strip with a function capturing the energy burdens at different locations as shown by Fig. 4. The proposed function is motivated by previous ‘continuum’ models for flows in high-density networks [10], [13]. In those works it is shown that a ‘well-balanced’ spreading of traffic, leads to a harmonic distribution of flow strengths around the 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’ scheme, i.e., a scheme without traffic spreading. The density for energy burdens, i.e., burden per unit area, is given by a function $h_w : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ such that, for $x \in S_w(s, d)$ and $\|s - d\| \geq w$ where $\|\cdot\|$ denotes l_2 norm in \mathbb{R}^2 ,

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

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.

D. Depletion Probability

At each round the energy burden from the perspective of an active sensor depends on the overlapping scenario of strips, i.e., data relaying/forwarding from AGNs to sinks. Assuming that these independent burdens are accumulated over a sufficient number of rounds, we make a second-order approximation, i.e., we apply the central limit theorem for the distribution of cumulative energy burdens at a typical active node².

Under a given strategy S let us denote the energy burden *per round* at a typical active node by a random variable $Y(S)$. Let us denote the mean and the standard deviation of $Y(S)$ by $\mu(S)$ and $\sigma(S)$ respectively. Suppose a node has been active for a total of l rounds, we refer to these as *effective rounds*³. Let us denote the energy burdens on an active node in j th effective round by $Y_j(S)$ and the accumulated energy burdens at round l by $Z_l(S)$, i.e.,

$$Z_l(S) = \sum_{j=1}^l Y_j(S)$$

²It can be show that the energy burden at each round has finite second and higher order moments. The fitness for normality has been validated by Lilliefors-adapted Kolmogorov-Smirnov test [14].

³A node is assumed to be active for l randomly selected rounds out of total number of rounds for the network. For simplicity we assume that the total number of rounds is sufficiently large so that the locations of sensors and AGN are roughly independent across rounds.

where $Y_j(S)$ are independent and identically distributed as $Y(S)$.

Suppose the battery capacity of all nodes is dimensioned proportional to its use, i.e., the number of effective rounds, at the network deployment. We denote the capacity by b which is parameterized by l such that b is some multiple β of l , i.e., $b = \beta l$. Considering the independence of energy burdens across rounds, we see that the mean and variance for energy burdens accumulated for l effective rounds correspond to $l\mu(S)$ and $l\sigma^2(S)$ respectively. For l sufficiently large, the depletion probability can be approximated as follows:

$$P(Z_l(S) > b) \approx \phi\left(\frac{b - l\mu(S)}{\sqrt{l\sigma^2(S)}}\right) = \phi\left(\sqrt{l} \cdot z_\beta(S)\right)$$

where we define the following:

$$z_\beta(S) := \frac{\beta - \mu(S)}{\sigma(S)}, \quad (3)$$

$$\phi(u) := \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-v^2/2} dv. \quad (4)$$

The objective is to minimize the depletion probability of a typical node after l effective rounds, i.e., we would like to maximize $z_\beta(S)$. 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.

IV. DERIVATION OF ENERGY BURDENS UNDER STRIP MODEL

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

A. Formula for Mean Energy Burden

We introduce our notation:

- Π_0, Π_1, Π_2 denote PPPs of sensors, AGNs and mobile sinks respectively. Alternatively $\Pi_k(B)$ represents the intensity measure for the point process Π_k for a Borel set $B \in \mathbb{R}^2$ for $k = 0, 1$ and 2 .
- λ_k is the intensity of point process Π_k .
- $V_z(\Pi)$ is the Voronoi cell with the nucleus at $z \in \Pi$ induced by the point process Π .
- $\mathbf{E}_k^{x_1, x_2, \dots, x_n}$ is the Palm expectation given that n points from Π_k are located at x_1, x_2, \dots, x_n . As a special case \mathbf{E}_k^0 denotes the Palm expectation given that a point in Π_k is at the origin. The similar notation applies to the Palm probability $\mathbf{P}_k^{x_1, x_2, \dots, x_n}$.
- $S_w(y, z)$ 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 model with the primary grain $\Phi_0 := B_{r_e}(O)$ induced by a PPP Π_e with intensity λ_e where O denotes the origin.

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. We will refer to this as *large-cell assumption* which accounts for the relative sparsity of mobile sinks with respect to 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: for this reason 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 although they depend on other system parameters. We start by approximating $\mu(\lambda_1, w)$.

We define the following:

$$\begin{aligned} \mathcal{M}_y &:= \mathbf{1}(y \in \Phi) + \sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_y(\Pi_1)), \\ \mathcal{N}_y &:= f(\mathcal{M}_y). \end{aligned}$$

For an AGN cell with nucleus at y , \mathcal{M}_y accounts for the number of sensors in Π_0 including the AGN, that 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. Recall that $h_w(x, y, z)$ is defined as the energy burden density at location x incurred by a strip originating from the AGN at 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 generated by the AGN 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. For the sake of simplicity we will estimate the mean and variance of energy burdens at a random location in the network.

Since this contribution is stationary, one can write the expectation of the energy burdens from the perspective of the origin:

$$\mu(\lambda_1, w) := \mathbf{E} \left[\sum_{y_j \in V_{Z_0}(\Pi_2) \cap \Pi_1} h_w(O, y_j, Z_0) \mathcal{N}_{y_j} \right] \quad (5)$$

where Z_0 is a random variable in \mathbb{R}^2 which represents the location of the sink closest to the origin: note that 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.5 in the sequel.

Key results used to estimate (5) are the refined Campbell’s theorem [12], and Ryll-Nardzewski–Slivnyak’s formula [15] which is also known as Palm inversion formula. The refined Campbell’s theorem states that for a stationary point process Π_k the expectation of a measurable function on the configuration of Π_k can be converted to an expectation under Palm measure, e.g., the following holds for a non-negative measurable u :

$$\mathbf{E} \left[\sum_{x \in \Pi_k} u(x, \Pi_k) \right] = \mathbf{E}_k^0 \left[\int u(x, \theta_x \Pi_k) \Pi_k(dx) \right].$$

where θ_x is a shifting operation, i.e., $\theta_x B = \{z | z = y - x, y \in B\}$ for a set $B \in \mathbb{R}^2$. In the above formula by setting $u(x, \Pi_k) := \mathbf{1}(x \in V_{X_0}(\Pi_k))v(\Pi_k)$ for a non-negative measurable function v where X_0 is the point in Π_k which is

closest to the origin, we obtain the Palm inversion formula:

$$\mathbf{E}[v(\Pi_k)] = \lambda_k \mathbf{E}_k^0 \left[\int_{V_0(\Pi_k)} v(\theta_x \Pi_k) dx \right].$$

We have the following proposition for our model.

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

$$\lambda_1 \mathbf{E}_1^0[\mathcal{N}_0] \cdot \tilde{m}(w, \lambda_2) \quad (6)$$

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

$$m(w) := \frac{\pi w^2}{(w-1)^2} (w-1 - \log(w)) - \frac{\pi w}{2}, \quad w \geq 1. \quad (7)$$

Proof: From the Palm inversion formula one can rewrite (5) as

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

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

$$\lambda_2 \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} \int_{V_0(\Pi_2)} \mathbf{E}_1^y [h_w(x, y, O) \mathcal{N}_y] \Pi_1(dy) dx \right] \quad (8)$$

$$= \lambda_2 \lambda_1 \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} \int_{V_0(\Pi_2)} h_w(x, y, O) \mathbf{E}_1^0[\mathcal{N}_0] dy dx \right] \quad (9)$$

By changing the order of integration in (9), i.e., integrating with respect to x first, the following holds for the inner integral:

$$\int_{V_0(\Pi_2)} h_w(x, y, O) dx = \int_{V_0(\Pi_2) \cap S_w(y, O)} h_w(x, y, O) dx \quad (10)$$

$$\leq \int_{S_w(y, O)} h_w(x, y, O) dx \quad (11)$$

Thus by defining

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

we have an upper bound for (9) given by

$$\lambda_1 \lambda_2 \mathbf{E}_1^0[\mathcal{N}_0] \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} g_w(y, O) dy \right]. \quad (12)$$

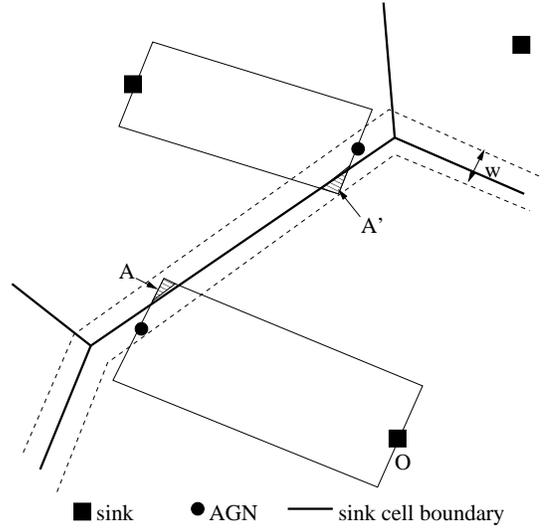


Fig. 5. Edge effects in the typical sink cell $V_0(\Pi_2)$. Note the dimension of the size of strips are exaggerated in the figure for clarity.

We can further write

$$\begin{aligned} & \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} g_w(y, O) dy \right] \\ = & \mathbf{E}_2^0 \left[\int_{\mathbb{R}^2} g_w(y, O) \mathbf{1}(\Pi_2 \cap B_{\|y\|}(y) = \emptyset) dy \right] \\ = & \int_{\mathbb{R}^2} g_w(y, O) \mathbf{E}_2^0[\mathbf{1}(\Pi_2 \cap B_{\|y\|}(y) = \emptyset)] dy \\ = & \int_{\mathbb{R}^2} g_w(y, O) \exp(-\pi \lambda_2 \|y\|^2) dy. \end{aligned} \quad (13)$$

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 (7). We have that

$$\begin{aligned} & \int_{\mathbb{R}^2} g_w(y, O) \exp(-\pi \lambda_2 \|y\|^2) dy \\ = & \int_{\mathbb{R}^2} \|y\| e^{-\pi \lambda_2 \|y\|^2} dy + m(w) \int_{\mathbb{R}^2 \setminus B_w(O)} e^{-\pi \lambda_2 \|y\|^2} dy \\ = & \int_{\mathbb{R}^2} \|y\| e^{-\pi \lambda_2 \|y\|^2} dy + \frac{m(w) \exp(-\pi \lambda_2 w^2)}{\lambda_2}. \end{aligned} \quad (14)$$

We integrate (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 (12) and (14) we obtain the proposed upper bound. \blacksquare

Although (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 paper considering the inequalities involving (11) and (14) as follows.

Indeed (11) is an overestimate of (10) since it counts 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. 5. However we have also ignored the contribution of strips that ‘overlap’ the cell boundary as is the case for A' in Fig. 5, i.e., our expression is already an underestimate of the actual mean cost. Thus considering the homogeneity of Π_0 and Π_1 , such errors will 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. Such region is a subset of a ‘band’ along the boundary as depicted in dotted line in Fig. 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 under large-cell assumption $\mu(\lambda_1, w)$ is strictly increasing in w , i.e., the mean energy burden increases with the degree of spreading since in order to do so some packets take longer ‘detours’. This property of $\mu(\lambda_1, w)$ shows us that, excessive spreading incurs a cost that may eventually counteract the load balancing benefit achieved by the spreading – this will be seen in the sequel.

The following proposition gives an estimate for $\mathbf{E}_1^0[\mathcal{N}_0]$ in 1.

Proposition 2: For our network model we have that

$$\mathbf{E}_1^0[\mathcal{M}_0] = \left(1 + \frac{\lambda_0}{\lambda_1}\right) \{1 - \exp(-\lambda_e \pi r_e^2)\}, \quad (15)$$

$$\begin{aligned} \mathbf{E}_1^0[\mathcal{N}_0] &\leq \alpha \mathbf{E}_1^0[\mathcal{M}_0] \\ &+ (1 - \alpha) \left(1 - \exp\left\{-\lambda_e \left(\frac{1}{\lambda_1} + \frac{4r_e}{\sqrt{\lambda_1}} + \pi r_e^2\right)\right\}\right). \end{aligned} \quad (16)$$

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

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

The first term of the above corresponds to the contact probability of a stationary Boolean model [12] which we denote by p . It is given by

$$p := 1 - \exp(-\lambda_e |\Phi_0|) = 1 - \exp(-\lambda_e \pi r_e^2)$$

where $|\cdot|$ denotes the Lebesgue measure in \mathbb{R}^2 . In other words, p is the probability that a randomly chosen point on \mathbb{R}^2 will intersect with Φ . We have that

$$\begin{aligned} &\mathbf{E}_1^0 \left[\sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_0(\Pi_1)) \right] \\ &= \lambda_0 \mathbf{E}_1^0 \left[\int_{\mathbb{R}^2} \mathbf{1}(x \in \Phi \cap V_0(\Pi_1)) dx \right] \\ &= \lambda_0 \int_{\mathbb{R}^2} \mathbf{E}_1^0[\mathbf{1}(x \in V_0(\Pi_1))] \mathbf{E}[\mathbf{1}(x \in \Phi)] dx \\ &= \lambda_0 \int_{\mathbb{R}^2} \mathbf{P}_1^0(\Pi_1 \cap B_{\|x\|}(x) = \emptyset) \mathbf{P}(x \in \Phi) dx \\ &= \lambda_0 \{1 - \exp(-\lambda_e \pi r_e^2)\} \int_{\mathbb{R}^2} \exp(-\lambda_1 \pi \|x\|^2) dx \end{aligned}$$

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

which proves (15).

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

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

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 has 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 [12]. Thus we have that

$$\begin{aligned} &\mathbf{P}_1^0(V_0(\Pi_1) \cap \Phi \neq \emptyset) \\ &= 1 - \mathbf{E}_1^0 \exp(-\lambda_e |B_{r_e}(O) \oplus V_0(\Pi_1)|) \\ &\leq 1 - \exp(-\lambda_e \mathbf{E}_1^0[|B_{r_e}(O) \oplus V_0(\Pi_1)|]) \quad (17) \\ &= 1 - \exp(-\lambda_e \{\mathbf{E}_1^0[V_0(\Pi_1)] + \mathbf{E}_1^0[\partial V_0(\Pi_1)] r_e + \pi r_e^2\}) \quad (18) \\ &= 1 - \exp\left\{-\lambda_e \left(\frac{1}{\lambda_1} + \frac{4r_e}{\sqrt{\lambda_1}} + \pi r_e^2\right)\right\}. \quad (19) \end{aligned}$$

In (17) we have used Jensen’s inequality. In (18) we have used Steiner’s formula for Minkowski addition since the primary grain Φ_0 is a disc [12], where $\partial V_0(\Pi_1)$ denotes the perimeter of a typical Voronoi cell. In (19) we have used the mean area and perimeter results for a typical cell of a stationary Voronoi tessellation. This proves (16). ■

We introduce the following notation:

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

To summarize the mean energy burden (5) is upper bounded by

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

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

B. 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:

$$\begin{aligned}
& \mathbf{E} \left[\left\{ \sum_{y_j \in V_{z_0}(\Pi_2) \cap \Pi_1} h_w(O, y_j, z_0) \mathcal{N}_{y_j} \right\}^2 \right] \\
&= \lambda_2 \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} \left[\sum_{y_j \in V_0(\Pi_2) \cap \Pi_1} h_w(x, y_j, O) \mathcal{N}_{y_j} \right]^2 dx \right] \\
&= \lambda_2 \mathbf{E}_2^0 \left[\int_{V_0(\Pi_2)} \left[\sum_{y_j \in V_0(\Pi_2) \cap \Pi_1} h_w^2(x, y_j, O) (\mathcal{N}_{y_j})^2 \right. \right. \\
&\quad \left. \left. + \sum_{\substack{y_j, y_k \in V_0(\Pi_2) \cap \Pi_1 \\ y_j \neq y_k}} h_w(x, y_j, O) h_w(x, y_k, O) \mathcal{N}_{y_j} \mathcal{N}_{y_k} \right] dx \right]. \tag{20}
\end{aligned}$$

It turns out that this second moment is difficult to evaluate due to the second term of the above, i.e., (21). Thus we will develop an approximation using the first term (20) of the above as follows:

- 1) Estimate (20) as in Proposition 3.
- 2) In order to do so, we estimate $\mathbf{E}_1^0[\mathcal{N}_0^2]$ up to the first order in p .
- 3) Using the estimate for (20), we find a heuristic approximation to $\sigma^2(\lambda_1, w)$.

It is difficult to directly evaluate $\mathbf{E}_1^0[\mathcal{N}_0^2]$. Thus for simplicity we will only concern ourselves with the terms that are $O(p)$, i.e., up to the first order of 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 the density of nodes/AGNs, 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.

To estimate the variance we first show the following proposition:

Proposition 3: One can upper bound (20) by

$$\lambda_1 \mathbf{E}_1^0[\mathcal{N}_0^2] \cdot \tilde{v}(w, \lambda_2)$$

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

$$v(w) \exp(-\pi \lambda_2 w^2) + \frac{1}{2w\sqrt{\lambda_2}} \tag{22}$$

and $v(w)$ is defined as

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

Proof: The proof is similar to that of Proposition 1. \blacksquare Next we bound $\mathbf{E}_1^0[\mathcal{N}_0^2]$ up to the first order in p . By definition of \mathcal{N}_0 we have that

$$\begin{aligned}
\mathbf{E}_1^0[\mathcal{N}_0^2] &= \alpha^2 \mathbf{E}_1^0[\mathcal{M}_0^2] + 2\alpha(1-\alpha) \mathbf{E}_1^0[\mathcal{M}_0] \\
&\quad + (1-\alpha)^2 \mathbf{P}_1^0(\mathcal{M}_0 > 0).
\end{aligned}$$

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

$$\begin{aligned}
\mathbf{E}_1^0[\mathcal{M}_0^2] &= \mathbf{E}[\mathbf{1}(O \in \Phi)] + \mathbf{E}_1^0 \left\{ \sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_0(\Pi_1)) \right\}^2 \\
&\quad + 2\mathbf{E}_1^0 \left[\mathbf{1}(O \in \Phi) \sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_0(\Pi_1)) \right]. \tag{23}
\end{aligned}$$

The first term is simply the contact probability p . We have the following proposition for the second term of (23).

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

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

where the constant c is by the expression (32) in the appendix.

Proof: See appendix. \blacksquare

Proposition 5: The third term in (23) is, up to the first order in p , bounded above by

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

where $\text{erf}(\cdot)$ is the standard error function, i.e.,

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

Proof: See appendix. \blacksquare

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

$$\begin{aligned}
\alpha^2 p \left[1 + \frac{3\lambda_0}{\lambda_1} - \frac{\lambda_0 \text{erf}(2\pi r_e \sqrt{\lambda_1})}{2r_e \lambda_1 \sqrt{\lambda_1}} + 2.56 \frac{\lambda_0^2 \sqrt{\lambda_1} r_e}{\lambda_1^2 (c + 2\sqrt{\lambda_1} r_e)} \right] \\
+ 2\alpha(1-\alpha) \left(1 + \frac{\lambda_0}{\lambda_1} \right) p + (1-\alpha)^2 p_b + o(p) \tag{24}
\end{aligned}$$

where we will ignore $o(p)$ term. Using this expression for $\mathbf{E}_1^0[\mathcal{N}_0^2]$ we get an estimate for (20) as stated by Proposition 3. Then using the estimate for (20) we resort to simulation of our model and use *heuristic estimation* of the variance as follows.

The simulation results show that, the variance of energy burdens $\sigma^2(\lambda_1, w)$ is well approximated by (20) 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 (21) become less and less correlated since the number of Π_1 cells intersected by a *single* event will become smaller. Due to the ‘scarcity’ of phenomena and the large sink cell size, (21) 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 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 heuristically come up with

the following final estimate for $\sigma^2(\lambda_1, w)$ based on extensive simulations:

$$\sigma^2(\lambda_1, w) \approx \max(1, \gamma(\pi\lambda_1 r_e^2)^\nu) \cdot \lambda_1 \mathbf{E}_1^0[\mathcal{N}_0^2] \tilde{v}(w, \lambda_2) \quad (25)$$

where γ 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 (20). Note $\pi\lambda_1 r_e^2$ 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 (20) this is a crude approximation to $\sigma^2(\lambda_1, w)$ as an effort to capture the degree of ‘clustering effect’ of events with respect to AGN cells.

Simulation results show that $\gamma \approx 0.89$ and $\nu \approx 0.5$ is a fair approximation. Thus we will use (25) as the expression for variance of energy burdens with the above γ and ν in the numerical results which is the next section.

Finally we take the cost associated with aggregation into account. The traffic aggregated at AGN is modelled by strips of width 1 generated at the sensors towards their AGNs where each strip carries traffic of size 1. By the similar techniques as those used for proving Propositions 1, the mean of the total energy burdens including aggregation overhead can be computed by simply adding $\frac{1}{2\sqrt{\lambda_1}}$ to the original expression for the mean, i.e., (6).

As for the variance of energy burdens incurred by aggregation, we make an 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 for future study.

V. 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 III-D, 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 formulated as follows:

$$\begin{aligned} &\text{maximize: } z_\beta(\lambda_1, w) := \left\{ \frac{\beta - \mu(\lambda_1, w)}{\sigma(\lambda_1, w)} \right\}, \\ &\text{subject to: } \lambda_2 < \lambda_1 < 1, w \geq 1, \\ &\text{variables: } w, \lambda_1, \\ &\text{given: } \lambda_2, \alpha, \beta, p, r_e, \lambda_e. \end{aligned}$$

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 . It turns out that with the above constraints $z_\beta(\lambda_1, w)$ admits a unique optimal value.

First let us 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

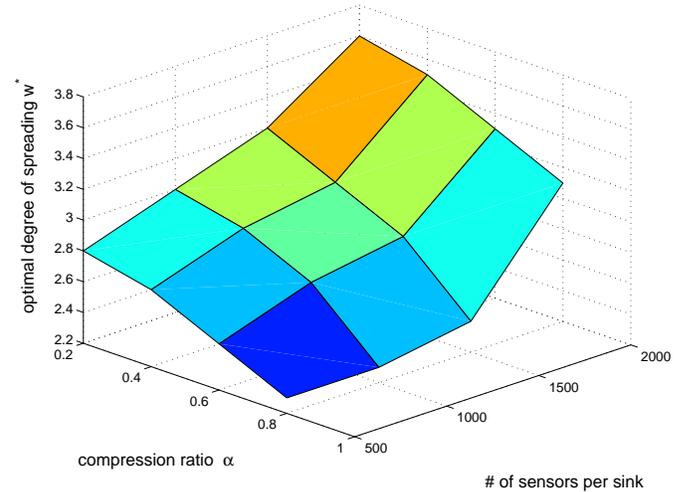


Fig. 6. Optimal degree of traffic spreading as a function of compression ratio and sink density.

the network aims to operate a total of $l = 200$ effective rounds. The maximum energy reserve β 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. 6 and Fig. 7 the radius of event disc r_e is fixed to 4 which implies there are on average about 50 sensors observing a single event disc. The value of r_e is varied in the plots in Fig. 8 and Fig. 9. Note however that the contact probability p is fixed to 0.1 and thus the density of phenomena λ_e is adjusted with r_e .

Fig. 6 (Fig. 7) shows 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 corresponds to $\lambda_2 = 10^{-3}$. The number of sensors per sink ranges from 500 to 2000, and the compression ratio 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. 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. Below we discuss our main observations.

Longer routes require more spreading of traffic. By inspecting Fig. 6 we can make the following observation. When we fix the compression ratio the optimal spreading width tends to increase as the number of sensors per sink increases. We see that in Fig. 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 6 indicates that one should spread traffic *more*.

This is intuitive since for larger sink cells, the number of *overlapping* strips will be greater, on average, as compared to that for smaller sink cells although the density of AGN, i.e., the total number of strips per unit area, remains fixed. This in turn implies that the imbalance in the energy burden profile is more of a concern, thus one should consider additional traffic spread to mitigate this effect.

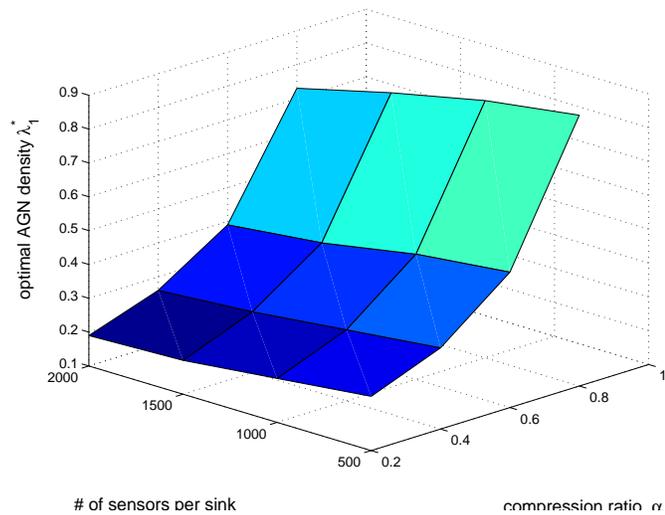


Fig. 7. Optimal aggregation density as a function of compression ratio and sink density.

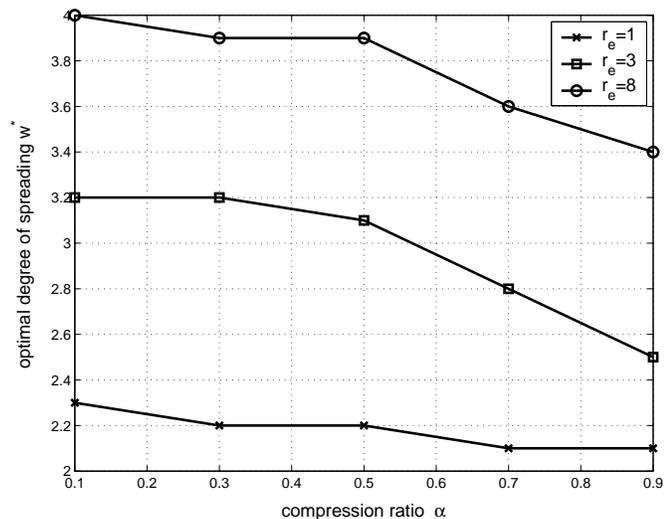


Fig. 8. Optimal degree of traffic spreading as a function of compression ratio and spatial scale of events.

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 which 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. 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 a larger role in a typical sensor's lifetime with worsening compression ratios when the size of sink cells is fixed. However if we consider Fig. 7 at the same time, the overall traffic per strip *increases* when 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 $\alpha = 0.8$

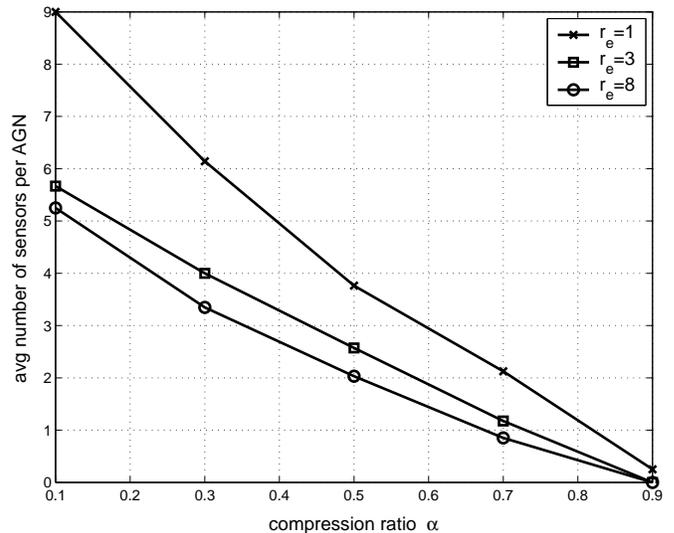


Fig. 9. Optimal average number of sensors per AGN as a function of compression ratio and spatial scale of events.

and (3.6, 0.2), i.e., roughly 4 sensors per AGN, when $\alpha = 0.2$. Thus each strip carries an average of 0.6 and 2 units of load 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, so strips with larger loads should spread *more* given that the average length of strips is fixed. Finally in Fig. 7 we see that the degree of aggregation increases, i.e., the densities for AGN decrease with improving compression performance.

The above observations are intuitive: if one should traverse certain distance and has more traffic, then one should spread the traffic more. Also if some traffic has to traverse longer distances, then one should spread the traffic more. This is beneficial to 'future' energy balance of the network in that the upcoming, possibly overlapping routes over the network will see more balanced energy burden profiles along their paths.

When the sensed phenomena are more 'bursty' one should further spread the traffic: Fig. 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.

We have imposed an important constraint on this numerical result: the contact probability p which represents the fraction of total area covered by Φ in \mathbb{R}^2 space, is *fixed*, i.e., the larger dimension of phenomena r_e corresponds to the lower phenomena density λ_e . For example a larger r_e would mean the phenomena tend to more spatially 'clustered' or the phenomena has more 'bursty' nature. Thus a larger r_e parameterizes a sensed phenomena for various burstiness of the spatial scale of phenomena.

As one can see in the figure, it is clear that for larger r_e one should spread more. The intuition is that, as the spatial scale of an individual event grows the spatial concentration of the traffic will be more severe since the traffic strips from

AGNs will overlap with each other with higher probability, as compared to that for smaller r_e . Thus the spreading of traffic becomes increasingly critical if the phenomena of interest has a high degree of burstiness.

When physical phenomena are less bursty the system admits further aggregation of traffic: Fig. 9 represents the the optimal degree of aggregation versus compression ratio where each curve is associated with different values of r_e . Specifically the y axis represents the average optimal number of sensors per AGN. Clearly if events are less ‘clustered’ one can aggregate more traffic and exploit the benefits of compression further. If the events are highly bursty then more aggregation exacerbates 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 a negative one that, perhaps surprisingly even for a very good compression performance, the degree of aggregation should not be too aggressive. For example see Fig. 9 for the case where $r_e = 1$ and $\alpha = 0.1$, i.e., even though one can reduce 90% of traffic compression, the optimal degree of AGN organization is such that only about 1 out of 10 sensors becomes an AGN on average. This corresponds to a benign situation where the spatial events are highly ‘scattered’, i.e., $r_e = 1$ implies there are only about 3 sensors per event disc. When the events are more bursty, i.e., with larger r_e , our results show that a only a few sensors per AGN is the optimal scale of aggregation.

The reason can be explained as follows: if one aggregates more traffic, one then needs to address concentration of energy burdens between AGNs and sinks. Although the problem can be mitigated by spreading traffic at a proper scale, one 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 reducing overall traffic by more aggregation versus avoiding excessive degree of aggregation to prevent concentration of energy burdens.

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 in large scale sensor networks.

VI. CONCLUSION AND FUTURE WORK

In this paper we address 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 argue that network operation can still be optimized to enhance the the operational lifetime

of a typical sensor. The key idea is to spread out aggregated traffic when it is forwarded from AGNs to sinks in order to smooth inhomogeneities in the energy burdens the network will incur.

In particular we introduce a model for bursty spatial phenomena, which serves as a caricature of some practical sensing applications. Such burstiness has 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. 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 paper 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 considered the joint optimization of these scales via numerical methods, and found 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 our 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 burden in a 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 the 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 one is able to devise practical ways of incorporating more elegant way of providing compression/fusion gains, e.g., distributed compression or network coding one may be able to overcome such burstiness with or without the traffic aggregation.

APPENDIX

Proof of Proposition 4: We have that

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

$$\begin{aligned}
&= \mathbf{E}_1^0 \left[\sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in V_0(\Pi_1) \cap \Phi) \right. \\
&\quad \left. + \sum_{\substack{x_j, x_k \in V_0(\Pi_1) \cap \Pi_0 \\ y_j \neq y_k}} \mathbf{1}(x_i, x_j \in V_0(\Pi_1) \cap \Phi) \right] \quad (26)
\end{aligned}$$

The first term of the above is already shown to be $\frac{\lambda_0}{\lambda_1}p$. Using the expression for the second-order moment measure of PPP [12] the second term is given by

$$\begin{aligned}
&\lambda_0^2 \iint \mathbf{E}_0^{x,y} [\mathbf{1}(x, y \in V_0(\Pi_1) \cap \Phi)] dx dy \\
&= \lambda_0^2 \iint \mathbf{E}_0^{x,y} [\mathbf{1}(x, y \in V_0(\Pi_1))] \mathbf{P}(x, y \in \Phi) dx dy \\
&= \lambda_0^2 \iint \mathbf{P}(\{B_{\|x\|}(x) \cup B_{\|y\|}(y)\} \cap \Pi_1 = \emptyset) \mathbf{P}(x, y \in \Phi) dx dy. \quad (27)
\end{aligned}$$

Using the (non-centered) covariance results for stationary Boolean model [12], 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 $\mathbf{P}(x, y \in \Phi)$ by p and take the limit as $\lambda_e \rightarrow 0$, we have that

$$\begin{aligned}
&\lim_{\lambda_e \rightarrow 0} \frac{\mathbf{P}(x, y \in \Phi)}{p} \\
&= \lim_{\lambda_e \rightarrow 0} \frac{1 - 2e^{-\lambda_e \pi r_e^2} + e^{-2\lambda_e \pi r_e^2 + \lambda_e C(r_e, \|x - y\|)}}{1 - \exp(-\lambda_e \pi r_e^2)} \\
&= \frac{C(r_e, \|x - y\|)}{\pi r_e^2}, \quad (28)
\end{aligned}$$

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

$$\begin{aligned}
&\lambda_0^2 \iint \mathbf{P}(\{B_{\|x\|}(x) \cup B_{\|y\|}(y)\} \cap \Pi_1 = \emptyset) \mathbf{P}(x, y \in \Phi) dx dy \\
&\approx p \lambda_0^2 \iint \exp(-\lambda_1 D(x, y)) \frac{C(r_e, \|x - y\|)}{\pi r_e^2} dx dy \quad (29) \\
&= p \frac{\lambda_0^2}{\lambda_1^2} \iint \exp(-D(x, y)) \frac{C(r_e, \|x - y\|/\sqrt{\lambda_1})}{\pi r_e^2} dx dy
\end{aligned}$$

Since (29) cannot be reduced to closed form, we use the following technique to obtain an upper bound. Firstly it is easy to verify that

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

Let

$$g(\rho) := \iint \exp(-D(x, y)) \exp(-\rho \|x - y\|) dx dy.$$

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 $dx dy$ 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 1: Let $h(\mathbf{r}) = \exp(-\bar{D}(\mathbf{r})) \chi(\mathbf{r})$. Suppose there exists a positive constant k such that

$$\frac{\int r_3 h(\mathbf{r}) d\mathbf{r}}{\int h(\mathbf{r}) d\mathbf{r}} < \frac{\int r_3 h(\mathbf{r}) \exp(-kr_3) d\mathbf{r}}{\int h(\mathbf{r}) \exp(-2kr_3) d\mathbf{r}} \quad (30)$$

holds. Then

$$g(\rho) \leq \frac{1}{1 + c\rho} \int h(\mathbf{r}) d\mathbf{r} \quad (31)$$

holds for $\rho > 0$ where

$$c := \frac{\int r_3 h(\mathbf{r}) d\mathbf{r}}{\int h(\mathbf{r}) d\mathbf{r}}. \quad (32)$$

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 [16]:

$$\begin{aligned}
\frac{d\gamma}{d\rho} &= \int h(\mathbf{r}) \frac{\partial f}{\partial \rho} d\mathbf{r} \\
&= \int h(\mathbf{r}) \exp(-\rho r_3) (c - r_3(1 + c\rho)) d\mathbf{r}
\end{aligned}$$

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

$$\frac{d\gamma}{d\rho} = \int h(\mathbf{r}) \exp(-\rho r_3) (c - cr_3\rho - r_3) d\mathbf{r} \quad (33)$$

$$\leq \int h(\mathbf{r}) \exp(-\rho r_3) (c \exp(-\rho r_3) - r_3) d\mathbf{r} \quad (34)$$

$$= c \int h(\mathbf{r}) \exp(-2\rho r_3) d\mathbf{r} - \int h(\mathbf{r}) r_3 \exp(-\rho r_3) d\mathbf{r} \quad (35)$$

where in (34) we have used $e^{-x} \geq 1 - x$ for all $x \geq 0$. Let us define the following:

$$f_1(\rho) := \int h(\mathbf{r}) \exp(-2\rho r_3) d\mathbf{r}, \quad (36)$$

$$f_2(\rho) := \int r_3 h(\mathbf{r}) \exp(-\rho r_3) d\mathbf{r}, \quad (37)$$

then we can write (35) as $cf_1(\rho) - f_2(\rho)$. One can show that using the monotonic properties of $f_1(\rho)$ and $f_2(\rho)$, the equation $cf_1(\rho) = f_2(\rho)$ has at most one solution. By

definition of c we have that $cf_1(0) = f_2(0)$. Thus for $\rho > 0$, $cf_1(\rho) - f_2(\rho)$ is either strictly positive or negative.

However by (30) implies there exists a constant⁴ $k > 0$ such that $cf_1(k) - f_2(k) < 0$, thus we can conclude that $cf_1(\rho) - f_2(\rho)$ 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. ■

It can be easily seen that $\int h(\mathbf{x})d\mathbf{x}$ 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 (31) and (26) we have proven Proposition 4.

Proof of Proposition 5: Note

$$\begin{aligned} & \mathbf{E} \left[\mathbf{1}(O \in \Phi) \sum_{x_i \in \Pi_0} \mathbf{1}(x_i \in \Phi \cap V_0(\Pi_1)) \right] \\ &= \mathbf{E} \left[\sum_{x_i \in \Pi_0} \mathbf{1}(\{x_i, O\} \in \Phi \cap V_0(\Pi_1)) \right] \\ &= 2\pi\lambda_0 \int \mathbf{E}_0^x [\mathbf{1}(\{x, O\} \in \Phi \cap V_0(\Pi_1))] dx \\ &= 2\pi\lambda_0 \int \mathbf{E}_0^x [\mathbf{1}(x \in V_0(\Pi_1))] \mathbf{E} [\mathbf{1}(\{x, O\} \in \Phi)] dx \\ &= 2\pi\lambda_0 \int \exp(-\pi\lambda_1 \|x\|^2) \\ & \quad \times \{2p - 1 + (1 - p)^2 \exp(\lambda_e C(r_e, \|x\|))\} dx. \end{aligned}$$

By evaluating the first order term with respect to p in the way similar to (28), we have that

$$2\pi\lambda_0 p \int \exp(-\pi\lambda_1 \|x\|^2) \frac{C(r_e, \|x - y\|)}{\pi r_e^2} dx \quad (38)$$

$$\leq 2\pi\lambda_0 p \int_0^{2r_e} \rho \exp(-\pi\lambda_1 \rho^2) \left\{1 - \frac{\rho}{2r_e}\right\} d\rho \quad (39)$$

$$= \frac{\lambda_0 p}{\lambda_1} \left\{1 - \frac{\text{erf}(2\pi r_e \sqrt{\lambda_1})}{4r_e \sqrt{\lambda_1}}\right\} \quad (40)$$

where in (39) we have used the fact that

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

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

REFERENCES

- [1] 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.
- [2] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *Proceedings of IEEE INFOCOM*, 2001, pp. 1360–1369.
- [3] 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.
- [4] D. Marco, E. Duarte-Melo, M. Liu, and D. Neuhoff, "On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data," in *International Workshop on Information Processing in Sensor Networks*, 2003.
- [5] 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.

⁴One can verify (30) holds for $k = 1$ via numerical evaluation.

- [6] E. Duarte-Melo, M. Liu, and A. Misra, "A computational approach to the joint design of distributed data compression and data dissemination in a data-gathering wireless sensor network," in *34th Allerton Conference on Communications and Control*, 2003.
- [7] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, pp. 51–60, March 2002.
- [8] 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.
- [9] 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.
- [10] S. BaeK and G. de Veciana, "Spatial energy balancing through proactive multipath 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.
- [11] F. Baccelli and S. Zuyev, "Poisson-voronoi spanning trees with applications to the optimization of communication networks," *INRIA Research Report No.3040*, Nov. 1996.
- [12] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, J. Wiley & Sons, Chichester, 1987.
- [13] S. Toumpis and L. Tassiulas, "Packetostatics: Deployment of massively dense sensor networks as an electrostatics problem," in *IEEE INFOCOM*, 2005.
- [14] H. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American Statistical Association*, vol. 62, pp. 399–402, 1967.
- [15] F. Baccelli and P. Brémaud, *Elements of queueing theory*, Springer-Verlag, 1994.
- [16] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, Academic, New York, 1965.



Seung Jun Baek (S'04) received his B.S. degree from Seoul National University in 1998, and M.S. and Ph.D. degree from University of Texas at Austin in 2003 and 2006 respectively, all in Electrical and Computer Engineering. He is currently with Communications Systems Laboratory, DSP Solutions R&D Center, Texas Instruments. His research includes modeling and analysis of wireless ad-hoc/sensor networks, and design architecture for broadband wireless networks such as WiMAX and 3GPP LTE.



Gustavo de Veciana (S'88-M'94-SM'01) received his B.S., M.S., and Ph.D. in electrical engineering from the University of California at Berkeley in 1987, 1990, and 1993 respectively. He is currently a Professor at the Department of Electrical and Computer Engineering at the University of Texas at Austin and the Director of the Wireless Networking and Communications Group (WNCG). His research focuses on the design, analysis and control of telecommunication networks. Current interests include: measurement, modeling and performance evaluation; wireless and sensor networks; architectures and algorithms to design reliable computing and network systems. Dr. de Veciana has been an editor for the IEEE/ACM Transactions on Networking. He is the recipient of General Motors Foundation Centennial Fellowship in Electrical Engineering and a 1996 National Science Foundation CAREER Award, co-recipient of the IEEE William McCalla Best ICCAD Paper Award for 2000, and co-recipient of the Best Paper in ACM Transactions on Design Automation of Electronic Systems, Jan 2002–2004.