Traffic Matrix Estimation in a Dense Multihop Wireless Network

Esa Hyytiä
Centre for Quantifiable Quality of Service in Communication Systems, Centre of Excellence,†
Norwegian University of Science and Technology, Trondheim, Norway

***Abstract*— The traffic matrix estimation based on the measurements at the certain points of a fixed network poses an interesting problem, which has also been studied extensively in the literature. In this paper, we consider a similar problem in the setting of dense multihop wireless network. In particular, we assume a large number of nodes with multihop routes using the shortest path routing, so that the routes can be modelled as straight line segments. Furthermore, we assume that we are able to measure the number of transmissions occurring in the different parts of the network during the measurement periods. In this setting we study the problem of inferring the end-to-end traffic demands (traffic matrix) based on the available information. As this information is not sufficient we make some additional Poissonian assumptions on the nature of the traffic in order to have a well-defined problem with a unique solution. Analysing the problem in the framework of stochastic geometry, we are able to give an exact solution for the formulated traffic matrix estimation problem. The methodology is further illustrated by numerical examples.**

***Index Terms*— traffic matrix, dense wireless multihop network, sensor networks, Poisson point process**

I. Introduction

The knowledge on the traffic matrix is valuable information for the IP network operators as it is the prime input parameter for making decisions both on the network design and also on how the current network should be operated. The topic has been studied extensively since the publication [1] by Vardi where the term "network tomography" is used to describe the problem of inferring the information on the end-to-end traffic demands based on the partial information. For details on different approaches and problem formulations we refer to [2], [3], [4].

The same fundamental question can be presented also for multihop wireless networks (MWN), i.e., based on the "link activity" measurements what can we infer on the end-to-end traffic demands between the different locations or parts of the network. The possible applications for this problem include both civilian and military applications. For example, in military radio reconnaissance one important task is to obtain the information on how the enemy units communicate. The necessary data for the analysis can be collected by a sensor network deployed, e.g., by an airplane flying over the enemy territory. Similarly, in a multihop mesh network or in a sensor network, the knowledge of the end-to-end traffic demands may be useful, e.g., for the load balancing or network design purposes.

In this paper we will present a framework based on the stochastic geometry for inferring the end-to-end traffic demands in a large scale MWN from the information on the transmission activities in the different parts of the network. The taken continuum approach is a valid assumption for a dense MWN. For sparse MWNs, with a smaller number of nodes, the presented approach still provides a viable approximation for studying end-to-end traffic demands.

The rest of the paper is organized as follows. In Section II we describe the assumptions made on the wireless network and the means how the measurements are conducted. In Section III the mathematical framework is presented together with the necessary results. Section IV contains the formal problem formulation and the exact solution to it. The framework is further illustrated by numerical examples in Section V, and Section VI concludes the paper.

II. Model

We consider a dense multihop wireless network (MWN), where, the nodes can be assumed to exist "everywhere" and the transmission range is several orders of magnitude smaller than a typical distance between two given nodes. At this limit the route taken by a packet can be modelled as a continuous path from the source to the destination. Moreover, if we assume the shortest paths routing, then the routes correspond to the line segments between the origin-destination pairs [5], [6], [7], [8].

† "Centre for Quantifiable Quality of Service in Communication Systems, Centre of Excellence" is appointed by The Research Council of Norway, and funded by the Research Council, NTNU and UNINETT. (http://www.ntnu.no/Q2S/)

A. Measurements

In a fixed network one typically can measure the link loads (or link packet counts), i.e., we know the point-to-point traffic volumes between the neighbouring nodes (considering the nodes as neighbours if there is a link between them). However, the situation with traffic measurements in a wireless multihop network is quite different from the traditional fixed networks. In a wireless environment one uses broadcast transmissions and it is not possible to identify the receiver (or even the transmitter in our case). Consequently, the direction of the traffic must be inferred by indirect means.

We assume that the measurement data is based on the observations of the transmission events in different parts of the network. The actual content of the packets and, in particular, their source and destination are not known to us. For example, the packets may have been encrypted. And, as mentioned already, we are not able to identify the local direction of the packet, i.e., at each (measurement) point we only know the number of packet transmissions that occurred during the measurement time interval in the given neighbourhood. This information allows as to estimate the local traffic load in the measurement points. By traffic load we mean, roughly speaking, the rate of transmissions occurring in the neighbourhood of the given node. We will give a more formal definition for the traffic load later in Section III. Finally, we assume that the retransmissions are so rare that they can be neglected, or that we can identify them somehow and count each packet only once (in the corresponding neighbourhood).

The actual measurement data is as follows. Let n denote the number of measurement points (or devices) deployed in the given area. Using these n devices we have conducted K independent measurements, where at each measurement point we have counted the number of transmissions that occured in the neighbourhood of the given measurement point (for simplicity we assume that each packet is counted only once). Hence, our sample set consists of $K \cdot n$ values each representing the transmission activity in the neighbourhood of a given node during a certain time interval Δt. Let vectors $\mathbf{m}^{(k)}$ denote the number of packets observed in n different measurement points during the kth time interval, $k = 1, \ldots K$. Then, the mean number of packets observed in different points is given by

$$\hat{\mathbf{m}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{m}^{(k)}, \tag{1}$$

and the corresponding covariance matrix is given by,

$$\hat{\mathbf{\Sigma}} = \frac{1}{K} \sum_{k=1}^{K} (\mathbf{m}^{(k)} - \hat{\mathbf{m}})(\mathbf{m}^{(k)} - \hat{\mathbf{m}})^{\mathrm{T}}. \tag{2}$$

III. Mathematical Framework

Next we will give some mathematical definitions and results which allow us the model and analyse the traffic flows in MWN. In the limit of dense (wireless) network the traffic matrix and traffic load can be expressed using continuous functions. Thus, it is natural the define the traffic matrix as follows:

Definition 1 (traffic matrix) *The traffic matrix, denoted by $\lambda(\mathbf{r}_1, \mathbf{r}_2)$ defines the rate at which packets are generated from a small area of dA around $\mathbf{r}_1$ to a small area of dA around $\mathbf{r}_2$.*

The total packet generation rate in the network is simply,

$$\Lambda = \int_{\mathcal{A}} \int_{\mathcal{A}} \lambda(\mathbf{r}_1, \mathbf{r}_2)\, d^2\mathbf{r}_2 \; d^2\mathbf{r}_1, \tag{3}$$

where $\mathcal{A}$ corresponds to the area (in plane) where the network is located. As a part of our earlier work, in [7], [8], we have formulated the spatial traffic load in this setting as follows:

Definition 2 (traffic load in dense multihop network) *Traffic load, denoted by $\Phi(\mathbf{r})$, is the scalar packet flux.*

A computational formula for the (scalar) packet flux in a dense MWN is given in [7] and [8], and here we briefly restate the definitions and the relevant results for our purposes.

Let $a(\mathbf{r}, \phi)$ denote the distance from $\mathbf{r}$ to the boundary of $\mathcal{A}$ in direction ϕ. The *angular packet flux*, denoted by $\varphi(\mathbf{r}, \phi)$, corresponds to the the rate at which packets arrive from a differential angle interval $(\phi, \phi{+}d\phi)$ across a perpendicular differential line segment at $\mathbf{r}$ divided by the differential angle $d\phi$ and the length of the differential line segment ds. This is illustrated in Fig. 1. Thus, the angular flux corresponds to the packet flux per unit angle, per unit distance, and per unit time. As explained in [7] the **angular flux** is given by

$$\varphi(\mathbf{r}, \phi) = \int_0^{a_1} \int_0^{a_2} \lambda(\mathbf{r}_1, \mathbf{r}_2) \cdot (t_1 + t_2)\, dt_2\, dt_1 \tag{4}$$

where $\mathbf{r}_1 = \mathbf{r} + t_1 \cdot (\cos\phi, \sin\phi)$, $\mathbf{r}_2 = \mathbf{r} + t_2 \cdot (\cos(\phi+\pi), \sin(\phi+\pi))$, and $a_1 = a(\mathbf{r}, \phi)$ and $a_2 = a(\mathbf{r}, \phi+\pi)$ correspond to the distance to the boundary from $\mathbf{r}$ in the directions $\phi + \pi$ and ϕ, respectively (see Fig. 1). Consequently, the **scalar packet flux** is obtained by taking an integral of (4) over all the possible angles,

$$\Phi(\mathbf{r}) = \int_0^{2\pi} \int_0^{a_1} \int_0^{a_2} \lambda(\mathbf{r}_1, \mathbf{r}_2) \cdot (t_1 + t_2)\, dt_2\, dt_1\, d\phi. \tag{5}$$

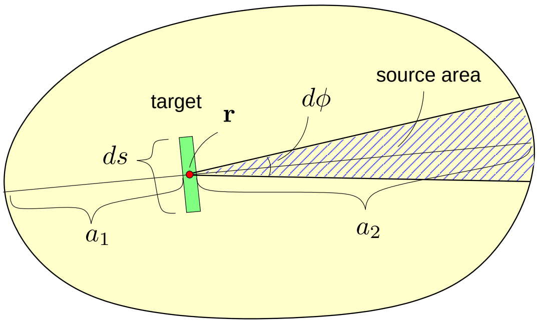

Fig. 1: The angular flux is the rate at which packets originating from the shaded area cross the differential perpendicular line segment at r divided by the differential angle $d\phi$ and the length of the differential line segment ds.

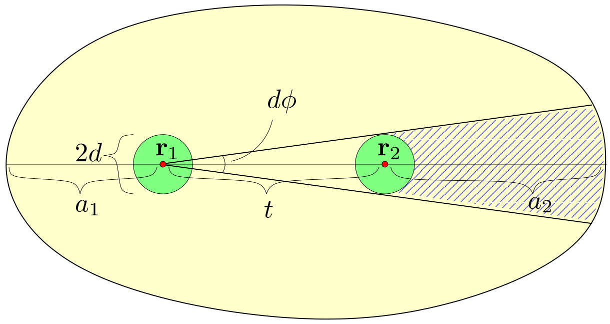

Fig. 2: The rate of packets originating behind the disk at $\mathbf{r}_2$ and travelling through the disk at $\mathbf{r}_1$ can be related to the concept of angular flux.

For details we refer to [7], [8]. We note that the concept of packet flux is similar to fluxes encountered in physics (see, e.g., [9]). For example, we have

$$\Phi(\mathbf{r}) = n(\mathbf{r}) \cdot v(\mathbf{r}), \qquad (6)$$

where $n(\mathbf{r})$ is the packet density (per unit area) and $v(\mathbf{r})$ a constant (local) velocity at point $\mathbf{r}$. Moreover, for small values of d we can estimate the arrival rate of packets into a $\mathbf{r}$-centric disk with radius d by

$$q(\mathbf{r}, d) \approx 2d \cdot \Phi(\mathbf{r}), \qquad (7)$$

which justifies the use of the packet flux as the measure of the traffic load in a dense MWN.

A. Common traffic

Next we will derive an expression for determining the rate of packets, which are seen by two different nodes with reception ranges equal to d. To this end, consider two small non-overlapping disks with radius d at points $\mathbf{r}_1$ and $\mathbf{r}_2$ corresponding to the observation areas. The situation is illustrated in Fig. 2.

The common traffic, corresponding to the packets travelling through both disks, can be obtained by considering the angular packet flux. Let ϕ denote the direction of $\mathbf{r}_2$ from $\mathbf{r}_1$, $a_1 = a(\mathbf{r}_1, \phi + \pi)$ and $a_2 = a(\mathbf{r}_2, \phi)$. Note that here the role of a_2 has changed a bit, i.e., a_2 corresponds to the distance to the boundary from $\mathbf{r}_2$ (not from $\mathbf{r}_1$). Consider next the packets which travel first through the disk at $\mathbf{r}_2$, and then later through the disk at $\mathbf{r}_1$. The height of the perpendicular “target area” at $\mathbf{r}_1$ is $2d$ (cf. ds in Fig. 1). Similarly, the differential angle $d\phi$ from which the packets behind the disk at $\mathbf{r}_2$ may arrive is equal to $2d/t$. Moreover, from (4) we can identify that the fraction of the angular flux originating from a distance further than t,

$$t = |\mathbf{r}_1 - \mathbf{r}_2|,$$

is given by

$$\int_0^{a_1} \int_t^{t+a_2} \lambda(\mathbf{r}_1,\, \mathbf{r}_2) \cdot (t_1 + t_2)\; dt_2\, dt_1.$$

Multiplying the above by $2d \cdot 2d/t$ gives us the rate of packets observed first by the node at $\mathbf{r}_2$ and then later by the node at $\mathbf{r}_1$. By considering the traffic flows from the both directions, ϕ and $\phi + \pi$, (note that (4) is symmetric) one obtains that the rate of packets going through both observation areas is given by

$$\frac{4d^2}{t} \int_0^{a_1} \int_0^{a_2} s(\mathbf{r}_1 - t_1\mathbf{u},\, \mathbf{r}_2 + t_2\mathbf{u}) \cdot (t + t_1 + t_2)\; dt_2\, dt_1,$$

where $s(\mathbf{r}_1, \mathbf{r}_2)$ is the **bidirectional packet rate** density,

$$s(\mathbf{r}_1, \mathbf{r}_2) = s(\mathbf{r}_2, \mathbf{r}_1) = \lambda(\mathbf{r}_1, \mathbf{r}_2) + \lambda(\mathbf{r}_2, \mathbf{r}_1), \qquad (8)$$

and $\mathbf{u}$ is a unit vector from $\mathbf{r}_1$ towards to $\mathbf{r}_2$,

$$\mathbf{u} = \frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|}.$$

With a slight abuse of notation, let $\Phi(\mathbf{r}_1, \mathbf{r}_2)$ denote the common packet flux (per unit distance to power of two) travelling through both $\mathbf{r}_1$ and $\mathbf{r}_2$, for which we have

$$\Phi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{t} \int_0^{a_1} \int_0^{a_2} s(\mathbf{r}_1 - t_1\mathbf{u}, \mathbf{r}_2 + t_2\mathbf{u}) \cdot (t + t_1 + t_2)\; dt_2\, dt_1, \qquad (9)$$

This relation turns out to be especially useful in the context of the traffic matrix estimation, as, together with some additional assumptions, it allows us to differentiate between the different traffic flows. For a constant packet rate density, $s(\mathbf{r}_1, \mathbf{r}_2) = s$, (9) reduces into

$$\Phi(\mathbf{r}_1, \mathbf{r}_2) = s \cdot a_1\, a_2 \left(1 + \frac{a_1 + a_2}{2t}\right).$$

Additionally, note that with $s(\mathbf{r}_1, \mathbf{r}_2)$ the expression for the packet flux, (5), can be written as

$$\Phi(\mathbf{r}) = \int_0^{\pi} \int_0^{a_1} \int_0^{a_2} s(\mathbf{r}_1,\, \mathbf{r}_2) \cdot (t_1 + t_2)\, dt_2\, dt_1\, d\phi. \quad (10)$$

IV. Problem Formulation

In [7] we have studied the resulting traffic load in a dense MWN when the end-to-end traffic demands are known and the shortest paths are used. In [8] this work is taken further by considering the load balancing problem by using curvilinear paths instead of the shortest paths. In this paper we address the inverse problem, i.e., we try to determine the end-to-end traffic demands based on the knowledge on the traffic load in different parts of the network. In order to keep the problem tractable we assume that the shortest paths, corresponding to line segments, are used, and that the reception range of the nodes is constant (or known which allows the scaling the measured packet rates to equivalent rates). We are interested in finding $\lambda(\mathbf{r}_1,\, \mathbf{r}_2)$ in (5) for a given packet flux $\Phi(\mathbf{r})$, for which we have the obvious estimate,

$$\hat{\Phi}(\mathbf{r}) = m(\mathbf{r})/(2d \cdot \Delta t).$$

As we have no means to differentiate between the packets travelling from $\mathbf{r}_1$ to $\mathbf{r}_2$ and packets travelling the opposite direction, from $\mathbf{r}_2$ to $\mathbf{r}_1$, we have to settle us with estimating the rate of bidirectional traffic, $s(\mathbf{r}_1, \mathbf{r}_2)$, given by (8) instead of the actual traffic matrix $\lambda(\mathbf{r}_1, \mathbf{r}_2)$.

It is easy to convince oneself that, without making any additional a priori assumptions, the problem of determining the (bidirectional) traffic matrix based on the knowledge of the spatial packet flux is strongly underdetermined and the solution is not unique. Firstly, the measurements are essentially a function from $\mathbb{R}^2$ to $\mathbb{R}$, while the quantity we are interested in is a function from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$. Secondly, any path carrying a certain amount of information can be split into several parts which together have the same contribution to the traffic load distribution. Thus, we have to make additional assumptions in order to have a well-defined problem with a unique solution.

A. Poisson point process

In particular, let us assume that during the observation period the number of packets (or bits) transmitted between two locations obeys Poisson distribution with some parameter and that the different traffic flows are independent. More precisely, we assume that the packet arrival process corresponds to a non-uniform Poisson point process in $\mathbb{R}^2 \times \mathbb{R}^2$ with unknown intensities $\lambda(\mathbf{r}_1, \mathbf{r}_2)$. In other words, let $X(\mathbf{r}_1, \mathbf{r}_2; t_1, t_2)$ denote the number of packets sent from the differential area dA around $\mathbf{r}_1$ to the differential area dA around $\mathbf{r}_2$ during a time interval of (t_1, t_2). Then, we assume that $X_1 = X(\mathbf{r}_1, \mathbf{r}_2; t_1, t_2)$ and $X_2 = X(\mathbf{r}_1', \mathbf{r}_2'; t_1', t_2')$ obey Poisson distributions with parameters $\lambda(\mathbf{r}_1, \mathbf{r}_2) \cdot dA^2 \cdot (t_2 - t_1)$ and $\lambda(\mathbf{r}_1', \mathbf{r}_2') \cdot dA^2 \cdot (t_2' - t_1')$, respectively. Moreover, if $\mathbf{r}_1 \neq \mathbf{r}_1'$ or $\mathbf{r}_2 \neq \mathbf{r}_2'$ or the time intervals (t_1, t_2) and (t_1', t_2') do not overlap, then X_1 and X_2 are independent.

The key property we will utilize is the independence. Let $\{X_i\}$ be a set of independent random variables, and $A = \sum_{i \in I_A} X_i$, and $B = \sum_{i \in I_B} X_i$. Then the covariance $\mathrm{Cov}\,[A,\, B]$ is simply the sum of the variances of the common terms,

$$\mathrm{Cov}\,[A,\, B] = \sum_{i \in I_A \cap I_B} \mathrm{V}\,[X_i]\,.$$

For Poisson random variable $X_i \sim \mathrm{Poisson}(\lambda_i)$ we have $\mathrm{E}\,[X_i] = \mathrm{V}\,[X_i] = \lambda_i$, and consequently

$$\mathrm{Cov}\,[A,\, B] = \sum_{i \in I_A \cap I_B} \lambda_i.$$

Note that the assumptions on the nature of the traffic are essentially the same as in [1] for the fixed networks.

B. Traffic matrix estimation problem

For our model of MWN, (5) gives the mean packet rate seen at $\mathbf{r}$, and (9) the mean rate of the common traffic seen at two different locations $\mathbf{r}_1$ and $\mathbf{r}_2$. Consequently,

$$\begin{aligned} \mathrm{E}\,[m(\mathbf{r})] &= 2d \cdot \Phi(\mathbf{r}) \cdot \Delta t, \\ \mathrm{E}\,[m(\mathbf{r}_1, \mathbf{r}_2)] &= 4d^2 \cdot \Phi(\mathbf{r}_1, \mathbf{r}_2) \cdot \Delta t. \end{aligned}$$

Note that the above relations depend only on the (mean) end-to-end traffic demands $\lambda(\mathbf{r}_1, \mathbf{r}_2)$. Moreover, as we are not able identify the packets, we cannot measure the common traffic $m(\mathbf{r}_1, \mathbf{r}_2)$ directly. On the other hand, with the Poissonian assumption for the traffic we have

$$\begin{aligned} m(\mathbf{r}) &\sim \mathrm{Poisson}(2d \cdot \Phi(\mathbf{r}) \cdot \Delta t), \\ m(\mathbf{r}_1, \mathbf{r}_2) &\sim \mathrm{Poisson}(4d^2 \cdot \Phi(\mathbf{r}_1, \mathbf{r}_2) \cdot \Delta t). \end{aligned}$$

Moreover, the covariance between the packet counts measured at two different locations, $\mathbf{r}_1$ and $\mathbf{r}_2$, is equal to the expected mean number of common packets, i.e.,

$$\mathrm{Cov}\,[m(\mathbf{r}_1),\, m(\mathbf{r}_2)] = 4d^2 \cdot \Phi(\mathbf{r}_1, \mathbf{r}_2) \cdot \Delta t. \quad (11)$$

By performing enough measurements we can, in theory, estimate both $\Phi(\mathbf{r})$ and $\Phi(\mathbf{r}_1, \mathbf{r}_2)$. Thus, we can formulate the traffic matrix estimation problem as follows:

Definition 3 (traffic matrix estimation problem) *Find such end-to-end traffic demands* $s(\mathbf{r}_1, \mathbf{r}_2)$ *for a given scalar packet flux,* $\Phi(\mathbf{r})$ *and a common packet flux,* $\Phi(\mathbf{r}_1, \mathbf{r}_2)$*, that satisfy both* (10) *and* (9).

C. Solution

Consider an arbitrary line segment $\ell = \ell(\mathbf{r}_1, \mathbf{r}_2)$ cutting the area $\mathcal{A}$ from boundary to boundary on which points $\mathbf{r}_1 \in \mathcal{A}$ and $\mathbf{r}_2 \in \mathcal{A}$ reside. Let L denote the length of the line segment ℓ, and r_1 and r_2 denote the distance from the arbitrarily chosen end point of ℓ to the points $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. Without loss of generality we can assume that $0 < r_1 < r_2 < L$, and (9) becomes

$$z(r_1, r_2) = \int_0^{r_1} \int_{r_2}^{L} s(t_1, t_2) \cdot (t_2 - t_1)\, dt_2\, dt_1, \tag{12}$$

where

$$z(r_1,\, r_2) = (r_2 - r_1) \cdot \Phi(r_1, r_2).$$

Taking the partial derivate of (12) in respect to both r_1 and r_2 yields

$$s(r_1, r_2) = -\frac{1}{r_2 - r_1} \cdot \frac{\partial^2}{\partial r_1 \partial r_2} z(r_1, r_2), \tag{13}$$

which is an explicit formula for computing the end-to-end traffic demand densities, $s(r_1, r_2)$, for a given $\Phi(r_1, r_2)$ (the measurement data).

D. Finite number of measurements

Assume next that instead of having unrealistically distributed the measurement devices everywhere, we have only deployed a finite set of n devices, as illustrated in Fig. 3. The assumption of a dense MWN implies that the distance between any two measurement devices is large when compared to the length of a typical hop.

Let the $\mathbf{r}_i$ denote the locations of the measurement devices in $\mathcal{A}$, and random variables m_i the corresponding measured packet counts during a time interval of Δt. A straightforward approach is to use the linear interpolation to estimate the packet counts $m(\mathbf{r})$ outside the measurement points $\{\mathbf{r}_i\}$. Then, for each $\mathbf{r} \in \mathcal{A}$ we have

$$m(\mathbf{r}) = \sum_i a_i(\mathbf{r}) \cdot m_i,$$

where the sum of the interpolation constants $a_i(\mathbf{r})$ is equal to 1. The covariance is linear and we obtain

$$\mathrm{Cov}\left[m(\mathbf{r}_1),\, m(\mathbf{r}_2)\right] = \sum_{i,j} a_i(\mathbf{r}_1)\, a_j(\mathbf{r}_2)\, \mathrm{Cov}\left[m_i,\, m_j\right],$$

and consequently,

$$\hat{\Phi}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{4d^2\, \Delta t} \sum_{i,j} a_i(\mathbf{r}_1)\, a_j(\mathbf{r}_2)\, \mathrm{Cov}\left[m_i,\, m_j\right]. \tag{14}$$

Let $\mathbf{\Sigma}$ denote the covariance matrix with $\mathrm{Cov}\left[m_i,\, m_j\right]$ as its elements, and $\mathbf{a}(\mathbf{r})$ the row vector consisting of the linear interpolation weights $a_i(\mathbf{r})$. Then (14) can be written in matrix form,

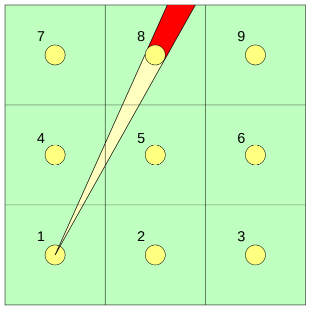

Fig. 3: Example: 3×3-measurement grid and d=0.1.

$$\hat{\Phi}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{4d^2\, \Delta t}\, \mathbf{a}(\mathbf{r}_1)\, \mathbf{\Sigma}\, \mathbf{a}(\mathbf{r}_2)^{\mathrm{T}} \tag{15}$$

Combining (13) with (14) (or (15)) allows us to determine the end-to-end traffic demand densities, $s(\mathbf{r}_1, \mathbf{r}_2)$, between any two given locations $\mathbf{r}_1$ and $\mathbf{r}_2$. Note that even though (13) is an exact result, there are two sources in the proposed approximation contributing to the overall error: 1) the finite number of n measurement points causing inaccuracy in the areas between the measure points, and 2) the finite number of K measurements causing inaccuracy in the estimate for the covariance (2).

V. Numerical Examples

The numerical examples assume a regular $m{\times}m$-grid depicted in Fig. 3 with a measurement node located in the middle of each cell. Thus, there are in total $n = m^2$ measurement points and, consequently, m^4 cross-covariances between the measurement points. For simplicity, and due to the lack of space, we will only briefly describe two simple examples with a 3×3 grid.

A. Uniform traffic demands

The first example is with the uniform end-to-end traffic demands. The nodes are enumerated as illustrated in Fig. 3 and the area is also 3×3. The reception range is assumed to be $d = 0.1$. The traffic rate density is a constant, $\lambda(\mathbf{r}_1, \mathbf{r}_2) = 1$, the measurement time interval $\Delta t = 1$ and the number of samples is $K = 100000$. With these, the (normalized) theoretical cross-covariances corresponding to the amount of common traffic between the number of observed packets in node 1 and node i, $i =, 2, \ldots, 9$ are

$$\left(1.50,\quad 0.31,\quad 1.50,\quad 3.00,\quad 0.39,\quad 0.31,\quad 0.39,\quad 0.63\right),$$

while the numerical simulation gives,

$$\left(1.57,\quad 0.38,\quad 1.44,\quad 2.47,\quad 0.54,\quad 0.44,\quad 0.59,\quad 0.60\right).$$

The resemblance is obvious even though the scenario is rather far from the assumptions of the model.

B. Mesh network with gateway nodes

Mesh networks are a special case of the general problem. The mesh network consists of one (or few) gateway nodes which are connected to the backbone network (e.g., Internet). All the traffic in the network can be assumed to be between the nodes and the gateway(s), i.e., at each point of the network (except at the gateways) the packets are traversing either to or from the nearest gateway. This is in strike contrast to the previous example, where no node has a special role.

In this case, it should be rather easy to identify the locations of the gateways as they correspond to the local maximums of the traffic load (i.e., the packet flux). Moreover, when moving radially away from the gateway and considering the rate at which traffic load decreases one can determine the rate of packets between the current position and the gateway.

In the example we have located the gateway node in the middle with the measurement point 5. The theoretical (normalized) covariances between node 5 and node i are

$$\big(1.25,\, 0.625,\, 1.25,\, 0.625,\, -,\, 0.625,\, 1.25,\, 0.625,\, 1.25\big)\,,$$

while the similar simulation setup gives

$$\big(1.033,\, 0.858,\, 1.027,\, 0.852,\, -,\, 0.835,\, 1.001,\, 0.848,\, 1.046\big)\,.$$

Other cross-covariances were close to zero, which is also the correct behaviour as in this case there is no common traffic between the other node pairs.

VI. Conclusions and Future Work

We have presented an analytical framework for estimating the traffic matrix in a wireless environment based on the transmission rates measured in the different parts of the network. In particular, we have assumed a dense wireless multihop network where a typical distance between two nodes is several magnitudes longer than the transmission range. In this limit, the paths the packets traverse can be assumed to be straight line segments. The analysis is carried out in the framework of stochastic geometry. The packet arrival process is modelled using a Poisson point processes in 4-dimensional space corresponding to the origin-destination pairs in a two-dimensional area $\mathcal{A}$ the network is located in.

In this setting, the spatial traffic load can be related to a concept of (scalar) packet flux. Together with the Poissonian assumptions, we were able to give an exact expression for the covariance between the measured traffic volumes in two different locations of the network as a function of the end-to-end traffic demands. The traffic matrix estimation problem is the inverse problem, i.e., the problem of inferring the end-to-end traffic demands from the knowledge of the traffic volume statistics in the different locations of the network. For this we were able to give an explicit result.

The exact result relies on the capability to measure the scalar packet flux in all points $\mathbf{r}\in\mathcal{A}$, which is a rather unrealistic assumption. Instead, in a more realistic scenario a finite number of measurement devices have been deployed in the area of interest. By using linear interpolation between the measurement points we were able to give an explicit expression for the end-to-end traffic demands also in this case. The finite number of measurement points and samples implies inaccuracy to the estimate, which is a topic of future research.

The approach taken in this paper shares the similar Poissonian assumptions on the nature of traffic as the methodology developed by Vardi in [1] for the traffic matrix estimation in fixed networks. Since then several other approaches have been proposed in the context of fixed networks. The future research includes studying the applicability of such approaches for the traffic matrix estimation problem in wireless multihop networks.

References

[1] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *Journal of the Americal Statistical Association*, vol. 91, no. 433, pp. 365–377, Mar. 1996.

[2] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic matrix estimation: existing techniques and new directions," in *SIGCOMM'02: Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, Pittsburgh, Pennsylvania, USA, 2002, pp. 161–174.

[3] Anders Gunnar, Mikael Johansson, and Thomas Telkamp, "Traffic matrix estimation on a large IP backbone: a comparison on real data," in *IMC'04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, Taormina, Sicily, Italy, 2004, pp. 149–160.

[4] Augustin Soule, Anukool Lakhina, Nina Taft, Konstantina Papagiannaki, Kave Salamatian, Antonio Nucci, Mark Crovella, and Christophe Diot, "Traffic matrices: balancing measurements, inference and modeling," in *Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, Banff, Alberta, Canada, 2005, pp. 362–373.

[5] Peter P. Pham and Sylvie Perreau, "Performance analysis of reactive shortest path and multi-path routing mechanism with load balance," in *Proceedings of Infocom '03*, San Francisco, USA, March-April 2003, IEEE, vol. 1, pp. 251–259.

[6] Yashar Ganjali and Abtin Keshavarzian, "Load balancing in ad hoc networks: Single-path routing vs. multi-path routing," in *Proceedings of Infocom '04*, Hong Kong, March 2004, IEEE.

[7] Esa Hyytiä, Pasi Lassila, Aleksi Penttinen, and János Roszik, "Traffic load in dense wireless multihop network," in *The 2nd ACM International Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks (PE-WASUN 2005)*, Montreal, Quebec, Canada, Oct. 2005, pp. 9–17.

[8] Esa Hyytiä and Jorma Virtamo, "On load balancing in a dense wireless multihop network," submitted for publication, Dec. 2005.

[9] G.I. Bell and S. Glasstone, *Nuclear Reactor Theory*, Reinhold, 1970.