Estimation of Message Source and Destination From Network Intercepts

Derek Justice and Alfred Hero, *Fellow, IEEE*

***Abstract*—We consider the problem of estimating the endpoints (source and destination) of a transmission in a network based on partial measurement of the transmission path. Possibly asynchronous sensors placed at various points within the network provide the basis for endpoint estimation by indicating that a specific transmission has been intercepted at their assigned locations. During a training phase, test transmissions are made between various pairs of endpoints in the network and the sensors they activate are noted. Sensor activations corresponding to transmissions with unknown endpoints are also observed in a monitoring phase. A semidefinite programming relaxation is used in conjunction with the measurements and linear prior information to produce likely sample topologies given the data. These samples are used to generate Monte Carlo approximations of the posterior distributions of source/destination pairs for measurements obtained in the monitoring phase. The posteriors allow for maximum a posteriori (MAP) estimation of the endpoints along with computation of some resolution measures. We illustrate the method using simulations of random topologies.**

***Index Terms*—Channel and network models, data acquisition and sensor models, detection and identification of anomalous events, network tomography and surveillance.**

I. Introduction

We present a method to estimate the endpoints (source and destination) of a data transmission in a network whose logical topology is unknown. We assume there are a number of asynchronous sensors placed on some subset of elements (links or nodes) in a network. A sensor is activated, and its activation recorded, whenever the path of a data transmission is intercepted on the element where the sensor is situated. The measurement apparatus is illustrated on a sample network in Fig. 1. Measurements are taken at discrete time instances, and the subscript k is used throughout the paper to index time. If multiple sensors are activated by a single transmission, they may not be capable of providing the precise order in which they were activated. In general, a probability distribution on the possible orders of activation $P_k(\rho)$ is observed for each measurement; here the argument $\rho \in \{1, 2, \ldots, \}$ is simply a natural number used to indicate a specific ordering of the sensors activated at time k. For example, a transmission with endpoints $u_1 = (\sigma_1, \delta_1)$ in Fig. 1 might activate $y_1 = \{\gamma_2, \gamma_3\}$-suppose this is the first measurement so $k = 1$. The ordering (γ_2, γ_3), corresponding to $\rho = 1$, might have probability $P_1(1) = 3/4$, while the ordering (γ_3, γ_2), where $\rho = 2$, has probability $P_1(2) = 1/4$. Since the orderings are defined over distinct sensor sets, we implicitly assume the transmission does not cycle in its path-that is, a particular sensor is activated at most once by a single transmission. During a preliminary training phase, the network is probed by transmitting data packets between various pairs of probing sites $\{u_k = (\sigma_k, \delta_k)\}_{k=1}^{K_o-1}$, and the sensors $\{y_k\}_{k=1}^{K_o-1}$ activated by each transmission are recorded along with the distributions on orderings $\{P_k(\rho)\}_{k=1}^{K_o-1}$. A monitoring phase begins at time instant K_o and continues until some final time K, whereby we observe sensor activation sets $\{y_k\}_{k=K_o}^{K}$ and associated ordering distributions $\{P_k(\rho)\}_{k=K_o}^{K}$ for which the endpoints are unknown.

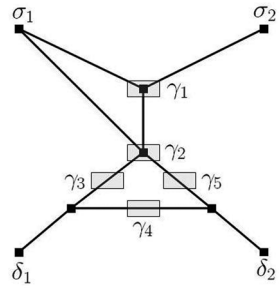


Fig. 1. Diagram of the measurement apparatus on a sample network. Probing sites are sources $\Sigma = \{\sigma_1, \sigma_2\}$ and destinations $\Delta = \{\delta_1, \delta_2\}$. A box on a link or node represents a sensor that indicates when a transmission path intercepts that link/node. We see γ_1 and γ_2 monitor nodes while γ_3, γ_4, and γ_5 monitor links.

The probing data $\{x_k\}_{k=1}^{K_o-1} \equiv \{u_k, y_k, P_k(\rho)\}_{k=1}^{K_o-1}$, the monitored data $\{x_k\}_{k=K_o}^{K} \equiv \{y_k, P_k(\rho)\}_{k=K_o}^{K}$, and some prior information about the network topology are processed to produce Monte Carlo estimates of the posterior distributions of possible endpoints of those transmissions observed in the monitoring phase. We allow prior information of the form $Q(\bar{A}) = b$ on the logical ($\{0, 1\}$) adjacency matrix A describing connections among sensors and probing sites. $\bar{A}$ is some subset of the elements of A, Q is a fixed linear operator, and b is a vector. Thus the prior information is essentially a set of linear equalities that the adjacency matrix A ought to satisfy. The linear operator Q can be expressed as an equivalent matrix if the elements of $\bar{A}$ are organized in a vector a. The linear prior information is then of the form $Qa = b$. In general, we make no assumptions about the structure of Q, so that given arbitrary Q and b the computation of feasible solutions to the linear equation is known to be an NP-Complete problem [1]. We consider the associated min-

Manuscript received April 25, 2005; revised March 30, 2006. This work was supported by the Department of Electrical Engineering and Computer Science, University of Michigan Graduate Fellowship of the first author and in part by the National Science Foundation under ITR Contract CCR-0325571. The Associate Editor coordinating the review of this manuscript and approving it for publication was Prof. Mohan S. Kankanhalli.

The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122 USA (e-mail: justiced@umich.edu; hero@umich.edu).

Digital Object Identifier 10.1109/TIFS.2006.879291

imum norm problem $\min \|Qa-b\|_\Lambda^2$ where $a \in \{0,1\}^n$ and $\|\cdot\|_\Lambda$ is a quadratic norm with respect to the positive definite matrix Λ. It is known that combinatorial optimization problems of this type may be successfully approximated by "lifting" them into a higher dimensional matrix space where $X_{ij} = a_i a_j$ and $X \in \{0,1\}^{n\times n}$ [2].

With the advent of polynomial time interior point methods for linear programming that can be extended to semidefinite programming [3], [4], it is convenient to consider a semidefinite programming (SDP) relaxation of the higher dimensional problem. Indeed, SDP relaxations have proven to be powerful tools for approximating hard combinatorial problems [5]–[8]. The SDP, however, is solved over a continuous domain so it is necessary to retrieve a 0–1 solution from the possibly fractional SDP solution. One possibility is a branch and bound scheme whereby certain variables are fixed and the SDP is repeated until a discrete solution is found [1], [8]. The branch and bound algorithm can take an exponential amount of time, depending on how tight the desired bound is. A randomized rounding scheme was developed in [6] for SDP relaxations of the maximum cut (MAXCUT) and maximum 2-satisfiability (MAX2SAT) problems. This scheme is shown to produce solutions of expected value at least 0.878 times the optimal value in [6]. We develop an SDP relaxation of the 0–1 minimum norm problem and apply the randomized rounding method in conjunction with samples from the ordering distributions $\{\rho_{1:K}^m\}_{m=1}^M$ to produce a number of network topology adjacency matrices $\{A^m\}_{m=1}^M$ that approximately satisfy the linear prior information $Q(\bar{A}) = b$. We derive an expression for the expected value of the squared error $\mathbf{E}\left[\|Qa-b\|_\Lambda^2\right]$ of samples produced in this way. This expression depends on the solution of the SDP relaxation, but an upper bound on the error independent of the SDP solution is also given.

We wish to produce posterior distributions given the data and prior information of the endpoints of transmissions observed in the monitoring phase $P(u_k|x_{1:K}, Q(\bar{A}) = b)$ for $k \geq K_o$. The network topology and sensor ordering samples are used in conjunction with prior distributions on the endpoints of measurements made during the monitoring phase $P_k(u)$ for $k = K_o, K_o+1, \ldots, K$ to compute Monte Carlo approximations of the desired posterior distributions via Bayes rule. Bayes formula for this problem essentially reduces to the expected value of a functional of the topology A and sensor ordering ρ; our approximation of the endpoint posterior thus becomes an average of the values of this functional at each sample topology A^m and ordering set $\rho_{1:K}^m$. It is readily apparent that this functional requires the conditionals $P(y|u, \rho, A)$-these path likelihood functions are the conditional probabilities of a sensor activation set y given the endpoints u and activation order ρ in a topology A. We propose a path likelihood model inspired by shortest path routing, whereby the length of a path determines its probability. Since the model is probabilistic, it is also well suited to dynamic algorithms, such as distance vector routing [9], which may not always choose the same path for a single endpoint pair. With the endpoint posterior distribution in hand, we can immediately give the MAP estimate of u_k (with $k \geq K_o$) or an a posteriori confidence region of probable source/destination pairs.

The related area of network tomography has recently been a subject of substantial research. It refers to the use of traffic measurements over parts of a network to infer characteristics of the complete network. Some characteristics of interest include the following: source/destination traffic rates [10], [11], link-level packet delay distributions [12], [13], link loss [14], and link topology [15], [16]. For an overview of relevant tomography problems for the Internet see [17]. In many applications, the tomography problem is ill posed since data are insufficient to determine a unique topology or delay distribution.

Our work is related to the internally sensed network tomography application described in [18], [19]. These works propose a methodology for estimating the topology of a telephone network using the measurement apparatus illustrated in Fig. 1. The data transmissions are of course telephone calls and the asynchronous sensors are located on trunk lines. A simple argument in [19] demonstrates that the number of topologies consistent with the data measured during the probing phase $\{x_k\}_{k=1}^{K_o-1}$ is exponential in the number of sensors. Indeed the problem is ill-posed as the data required to provide a reasonable estimate of the topology will never be available in practice. We sidestep the difficulties of developing a single topology estimate by averaging over many probable topologies in computing the endpoint posterior distribution.

The solution approach we develop is very general, and we suspect it might have application in all sorts of networks: including telephone networks as described in [18], the Internet, social networks (such as command and control structures), or biological networks (such as protein-protein interaction networks) [20], [21]. Since we allow for sensor placement on arbitrary network elements, the method is equally applicable to networks where it may be more convenient to monitor nodes (as in the Internet) or monitor links (as in the telephone network of [18]). Also, the ordering distributions allow for situations involving sensors ranging from asynchronous to perfectly synchronized. At one extreme, the sensors are exactly synchronized-in which case the distribution $P_k(\rho)$ reduces to a delta function with all mass concentrated on the known ordering of sensors. A natural source of such information would be the noisy time stamp assigned by each sensor to when it saw the message. Indeed, this is an issue faced in many active probing scenarios. Methods involving GPS and calibration of PC clocks have been described in [22] and [23], respectively, for addressing asynchronous sensors in active probing of technological networks. One might derive the ordering distributions from some noise model for the time stamps. In the present work, we assume the ordering distributions themselves are provided since the chronological order in which the sensors intercept a message is the crucial information.

Although the monitored network topology is unknown, the linear prior information permits inclusion of reasonably available information relevant to the topology. This is a generalization of the frequently used vertex degree prior. Vertex degree priors are used quite often due to the fact that many real world networks are characterized by specific degree distributions [24]. For example, studies have suggested a power-law distribution describes vertex degrees in the Internet [25]. Such priors have recently been applied to research involving models of social and

biological networks [20], [21], [26]. Since the degree of a vertex is equal to the sum over the row of the adjacency matrix describing connections to that vertex, one can easily construct a linear operator Q so that $Q(\bar{A}) = b$ expresses the degree prior for a given vector of vertex degrees b.

The approach described here might also find utility in systems conveniently modelled by graphs, such as finite state automata. The problem of machine identification is a classic problem in the theory of automata testing [27], [28]. Here, we are given a black box with an automaton inside whose transition function is unknown. Based on the response of the system to certain input sequences, we wish to reconstruct the transition function. The link to the network topology recovery aspect of our problem is clear, since a graph provides a convenient representation for the transition function of interest. The probing sites chosen in the probing phase of our problem is analogous to the input sequences to the black box automaton. Similarly, link sensors correspond to events in the automaton's observable event set. An exhaustive algorithm for solving this problem is given in [27] and shown to have exponential run time. Our methods might be adapted to provide a polynomial time approximation algorithm. This would involve partitioning measurements with cycles (whereby an observable event occurs more than once in the same string) to satisfy the direct path assumption and selecting a different conditional path likelihood $P(y|u, \rho, A)$ since the shortest path routing model we suggest might not be appropriate.

The outline of this paper is as follows. We review the problem, describe in detail each component of the endpoint estimation system, and analyze its complexity in Section II. In Section III, we provide some simulations of random graphs. In Section IV we conclude with some extensions of the method presented here and give directions for future work utilizing feedback for adaptive probing.

II. Model and Theory for Source-Destination Estimation

Let $G(V, E, f)$ be a simple graph defined by the vertex set V, edge set E, and incidence relation $f : E \to V \times V$ giving the vertices connected by each edge. We allow G to be either directed or undirected; however, it should be known a priori which is the case. In our application, E defines the set of links in the network topology, V defines the routers or switches connected by these links, and f determines the pair of routers/switches connected by each link. The graph G is unknown to us.

Let Γ denote a set of sensors we place in the network. Sensors are placed on some subset of graph elements; that is sensors may be placed on vertices, edges, or both. A sensor will indicate whenever a transmission through the network passes the element it is monitoring. Probing sites are selected from the vertex set V. The source vertex set $\Sigma \subseteq V$ is the set of vertices from which transmissions may originate, and the destination vertex set $\Delta \subseteq V$ are those vertices at which transmissions may terminate. A path observed at time $k < K_o$ between probing sites $s_k \in \Sigma$ and $d_k \in \Delta$ is given by $y_k \subseteq \Gamma$, where y_k contains the sensors activated by the transmission from s_k to d_k. We assume the first $K_o - 1$ measurements correspond to probes of the network (i.e., active measurements) so that the sources and destinations of these measurements are known (since we choose them). Because the sensors are in general asynchronous, the paths are unordered sets. However, along with each y_k, a discrete probability distribution $P_k(\rho)$ is observed on possible orderings indexed by ρ of the set y_k; the observation data is then $x_k \equiv (u_k, y_k, P_k(\rho))$ for $k < K_o$. We assume a transmission does not cycle in its path from source to destination, so that only orderings of distinct elements of y_k are considered. It follows that if y_k has $|y_k|$ distinct elements, then $P_k(\rho)$ is defined over $|y_k|!$ different orderings. Note that the case of perfectly synchronized sensors is easily handled in this framework: simply take $P_k(\rho) = \delta(\rho - \rho_k)$ where ρ_k is the known order in which the sensors y_k were activated.

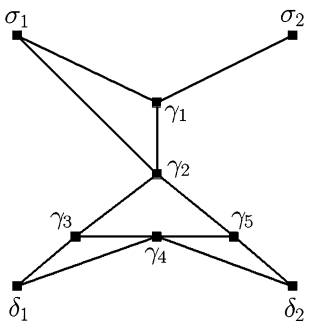


Fig. 2. Example logical topology $G_A(V_A, E_A)$ for the monitored network G in Fig. 1. The vertex set of G_A consists of sensors $\Gamma = \{\gamma_i\}_{i=1}^5$ and probing sites $\Sigma = \{\sigma_1, \sigma_2\}$, $\Delta = \{\delta_1, \delta_2\}$, so that $V_A = \Gamma \cup \Sigma \cup \Delta$. The edges of G_A summarize logical adjacencies among sensors and probing sites with any intervening unmonitored elements short-circuited.

At time K_o, we proceed with monitoring of the network, that is observing activated sensor sets with unknown source and destination. The observation data in the monitoring phase is $x_k \equiv (y_k, P_k(\rho))$ for $k \geq K_o$. The purpose of our system is to estimate the source and destination $u_k = (s_k, d_k)$ of an activated sensor set y_k. In order to estimate the endpoints of such a measurement, it is necessary to have some idea of the logical topology of the network. Instead of considering the logical adjacencies implied by the actual network $G(V, E, f)$, we are concerned with adjacency relationships among only those elements (vertices and edges) that are either monitored with a sensor or used as a probing site. For example, we cannot hope to pinpoint the position of a link e in the original network that is not monitored by a sensor. We assume unmonitored elements are essentially 'short-circuited' in the original network G. The idea here is to assure two elements are logically adjacent even if they are physically separated by an element (or subgraph of elements) that is not monitored. The particular topology we consider is then $G_A(V_A, E_A)$ where $V_A = \Gamma \cup \Sigma \cup \Delta$ is the set of sensors and probing sites, and $E_A \subseteq V_A \times V_A$ describes the logical adjacencies among these elements. G_A may be undirected or directed depending upon the nature of the network G. For computational purposes, we represent G_A by its adjacency matrix A where $A_{ij} = 1$ if and only if $(i, j) \in E_A$ for $i, j \in V_A$ and $A_{ij} = 0$ otherwise. An example logical topology G_A is given in Fig. 2 for the monitored network G in Fig. 1.

We assume independence of measurements at different times and utilize a Bayesian framework to produce suitable approximations of the endpoint posterior distribution

$$P(u_k|x_{1:K}, Q(\bar{A}) = b) = \mathbf{E}_{A,\rho_{1:K}}\left[\frac{P(y_k|u_k,\rho_k,A)P_k(u_k)}{\sum_u P(y_k|u,\rho_k,A)P_k(u)}|x_{1:K}, Q(\bar{A}) = b\right] \quad (1)$$

where the expression is obtained quite simply by using the law of total probability to expand the distribution $P(u_k|x_{1:K}, Q(\bar{A}) = b)$ over the random variables A, $\{\rho_k\}_{k=1}^K \equiv \rho_{1:K}$ and applying Bayes rule with appropriate independence assumptions to $P(u_k|x_{1:K}, \rho_{1:K}, A)$. Of course $k \geq K_o$ in (1) so that we are considering the endpoint posterior of a passive measurement. We have available linear prior information on some of the logical adjacency elements $\bar{A}$ (a submatrix of A) of the form $Q(\bar{A}) = b$ where Q is a fixed linear operator and a prior distribution on endpoints $P_k(u)$. It is assumed the endpoint pair of a passive measurement is independent of the particular topology A, in other words, the parties communicating do not know the network topology either. However, if there is no connection between a given endpoint pair u in a topology A, one would expect such a pair to have probability zero; we shall use a model for the term multiplying $P_k(u)$ to ensure the product is zero in this case. Here $x_{1:K} \equiv \{x_k\}_{k=1}^K$ represents all measured data ($x_k \equiv (u_k, y_k, P_k(\rho))$ for $k < K_o$ and $x_k \equiv (y_k, P_k(\rho))$ for $k \geq K_o$), and ρ_k is the ordering of the sensors activated in measurement y_k. The conditional expectation is therefore taken over all logical adjacency matrices A and sensor orderings for all measurements $\rho_{1:K}$. We introduce a shortest path routing model for the conditional path probabilities $P(y|u, \rho, A)$. The conditional expectation in (1) is approximated in a Monte Carlo fashion by summing over the argument evaluated at a number of adjacency matrix and sensor ordering samples. The sensor orderings $\rho_{1:K}$ are drawn independently from observed distributions $P_k(\rho)$ for $k = 1, 2, \ldots, K$. These are used in conjunction with the solution to a semidefinite programming relaxation that incorporates the prior information $Q(\bar{A}) = b$ to produce adjacency matrix samples A that are likely given both the data and the prior information. With the approximate endpoint posterior distribution in hand, we can provide MAP estimates of the endpoints of the passive measurement and compute appropriate error measures.

In the following, we first elaborate on probing of the network and the characterization of measurements obtained. Then we describe the distribution $P(A, \rho_{1:K}|x_{1:K}, Q(\bar{A}) = b)$ and how it may be efficiently sampled using the given ordering distributions and a semidefinite programming relaxation. Next we discuss how the samples are used to approximate the endpoint posterior and produce MAP estimates. Finally, we analyze the complexity of our algorithm.

A. Probing the Network and Taking Measurements

The set of all available measurements $\{x_k\}_{k=1}^K$ is partitioned into two disjoint sets. The measurements for $k = 1, 2, \ldots K_o - 1$ correspond to a training phase for the probing sites Σ, Δ. For each $k < K_o$, we select a probing pair $u_k \in \Sigma \times \Delta$ and pass a transmission between this pair to observe the sensors y_k activated and a distribution $P_k(\rho)$ on the $|y_k|!$ possible orderings of the activated sensors. The measurement data therefore consists of both the endpoints and the activated sensor set/ordering distribution $x_k = (u_k, y_k, P_k(\rho))$ for $k < K_o$. Such a measurement is referred to as an *active* measurement. The remaining measurements are due to monitored transmissions so that the endpoints are not available: $x_k = (y_k, P_k(\rho))$ for $K_o \leq k \leq K$. These are referred to as *passive* measurements since they were not due to active probing of the network on our behalf. It is assumed that the endpoints of these measurements are realizations of a random probing site pair described by the known distribution $P_k(u)$ defined on $\Sigma \times \Delta$. We desire to estimate the particular probing site pair between which a transmission was passed resulting in a given passive measurement.

B. Generating Topology and Sensor Ordering Samples

In order to produce a Monte Carlo estimate of the conditional expectation in (1), we need to specify and sample from the distribution $P(A, \rho_{1:K}|x_{1:K}, Q(\bar{A}) = b)$. We first expand this distribution as

$$\begin{aligned} &P(A, \rho_{1:K}|x_{1:K}, Q(\bar{A}) = b) \\ &\quad = P(A|x_{1:K}, \rho_{1:K}, Q(\bar{A}) = b) \\ &\qquad \times \prod_{k=1}^{k} P(\rho_k|x_{1:K}, Q(\bar{A}) = b) \end{aligned} \quad (2)$$

where independence over the measurement time index k is used to write the second term in product form. We now note that each measurement x_k contains a distribution over orderings $P_k(\rho)$ for the activated sensor set. Since these distributions are observations, it is reasonable to suspect that all topological considerations are folded into them. We therefore assume that given the ordering distributions, the particular orderings ρ_k are independent of the linear prior on topology. Equation (2) therefore becomes

$$\begin{aligned} &P\left(A, \rho_{1:K}|x_{1:K}, Q(\bar{A}) = b\right) \\ &\qquad = P\left(A|x_{1:K}, \rho_{1:K}, Q(\bar{A}) = b\right) \prod_{k=1}^{K} P_k(\rho_k). \end{aligned} \quad (3)$$

The factored form of the distribution in (3) suggests the first thing we should do in generating our samples is to select orderings ρ_k independently from the distributions P_k for each $k = 1, 2, \ldots, K$. This is a simple matter since each P_k is a discrete distribution defined over a finite number of orderings.

Consider now what a measurement x_k equipped with an ordering ρ_k implies about the adjacency matrix A. Let $x_{k\rho_k}$ denote the ordered sensor activation set where, if x_k is an active measurement, the source probing site is taken as the first element followed by the ordering ρ_k of the activated sensors and the destination probing site is taken as the last element. If x_k is a passive measurement, $x_{k\rho_k}$ is simply the ordering ρ_k of the activated sensors. The fact that the transmission passes from the lth element of $x_{k\rho_k}$, given by $x_{k\rho_k}^l$, to $x_{k\rho_k}^{l+1}$ implies there must be

a logical connection between $x^l_{k\rho_k}$ and $x^{l+1}_{k\rho_k}$. Thus if we select an ordering ρ_k for each measurement (i.e., for $k = 1, 2, \ldots K$), then every adjacency element in the set $A^{x\rho}$ must be 1, where $A^{x\rho}$ is defined by

$$A^{x\rho} = \left\{ A_{ij} \mid \exists k, l : \left(x^l_{k\rho_k}, x^{l+1}_{k\rho_k} \right) = (i, j) \right\}. \quad (4)$$

Once we draw orderings $\rho_{1:K}$ as previously described, the adjacency matrix elements in $A^{x\rho}$ are immediately fixed at unity by these. It remains, however, to select the remaining adjacency elements. In drawing these, we must account for the prior information $Q(\bar{A}) = b$. Since Q is a linear operator, we may re-express this information as $Qa = b$ where Q is now understood to be a matrix and $a \in \{0, 1\}^n$ is a vectorized version of the adjacency elements $\bar{A}$. For arbitrary Q, finding a 0–1 vector a that satisfies the equation $Qa = b$ is an NP-Complete problem [29]. We will shortly discuss how randomized rounding of a semidefinite programming relaxation may be used to find approximate solutions. The randomized rounding will induce a distribution on $P(A|x_{1:K}, \rho_{1:K}, Q(\bar{A}) = b)$, the remaining factor in (3). The induced distribution will have the desirable property that it assigns high probability to samples that approximately satisfy the linear constraint $Q(\bar{A}) = b$.

Consider the matrix equation $Qa = b$ equivalent to the linear prior information $Q(\bar{A}) = b$. Producing vectors a that satisfy this equation amounts to finding several solutions to the problem

$$\begin{aligned} \text{find} \quad & a \in \{0, 1\}^n \\ \text{such that} \quad & Qa = b. \end{aligned} \quad (5)$$

Unfortunately, the problem in (5) is NP-complete for an arbitrary, unstructured matrix Q [29]. We consider an equivalent restatement of (5)

$$\begin{aligned} \text{minimize} \quad & (Qa - b)^T \Lambda (Qa - b) \\ \text{such that} \quad & a \in \{0, 1\}^n \end{aligned} \quad (6)$$

where Λ is a (symmetric) positive definite matrix that may be chosen to emphasize the relative importance of the different constraints. Obviously any optimal solution of the problem in (6) with zero value solves the feasibility problem in (5). The problem in (6) is no easier than the original statement, however, it has been shown that problems of this type (0–1 quadratic programs) can be approximated quite well using a semidefinite relaxation [7].

We now proceed to derive the SDP relaxation of (6). Our relaxation is similar to the one derived in [6] for MAX2SAT. First note that the optimization in (6) is equivalent to

$$\begin{aligned} \text{minimize} \quad & a^T D a - 2 d^T a \\ \text{such that} \quad & a \in \{0, 1\}^n \end{aligned} \quad (7)$$

where $D = Q^T \Lambda Q$ and $d = Q^T \Lambda b$. This is easily seen by expanding the objective in (6) and dropping the constant term. Now note that $a_i^2 = a_i$ since $a_i \in \{0, 1\}$; this fact this allows (7) to be re-expressed as

$$\begin{aligned} \text{minimize} \quad & \sum_{i,j} D_{ij} a_i a_j - 2 \sum_j d_j a_j^2 \\ \text{such that} \quad & a \in \{0, 1\}^n. \end{aligned} \quad (8)$$

We now introduce variables $w_i \in \{-1, 1\}$ for each $a_i \in \{0, 1\}$ for $i = 1 \ldots n$ along with an additional $w_{n+1} \in \{-1, 1\}$ so that the change of variables is given by

$$a_i = \frac{1}{2}(1 + w_{n+1} w_i). \quad (9)$$

The identities in (10) follow from this change of variables:

$$\begin{aligned} a_i a_j &= \frac{1}{4} \left[(1 + w_i w_j) + (1 + w_{n+1} w_i) \right. \\ &\quad \left. + (1 + w_{n+1} w_j) - 2 \right] \\ -a_i a_j &= \frac{1}{4} \left[(1 - w_i w_j) + (1 - w_{n+1} w_i) \right. \\ &\quad \left. + (1 - w_{n+1} w_j) - 4 \right]. \end{aligned} \quad (10)$$

If we introduce a negative sign in the objective, then the optimization in (8) becomes

$$\begin{aligned} \max \quad & \frac{1}{4} \sum_{i,j} [B_{ij}(1 + w_i w_j) + C_{ij}(1 - w_i w_j) - 4 D_{ij}] \\ \text{such that} \quad & w \in \{-1, 1\}^{n+1} \end{aligned} \quad (11)$$

where e is a vector of ones and matrices B, C are given by

$$\begin{aligned} B &= \begin{pmatrix} 0 & 2d \\ 2d^T & 0 \end{pmatrix} \\ C &= \begin{pmatrix} D & De \\ (De)^T & 0 \end{pmatrix}. \end{aligned} \quad (12)$$

In order to obtain a semidefinite program, define the matrix $W = ww^T$. It is simple to show that $W = ww^T$ for some vector w if and only if $W \succeq 0$ (i.e., W is positive semidefinite) and $rank(W) = 1$. We drop the nonconvex rank-1 constraint to obtain the SDP relaxation

$$\begin{aligned} \text{maximize} \quad & Tr[(B - C)W] \\ & \text{diag}(W) = e \\ \text{such that} \quad & W \succeq 0 \end{aligned} \quad (13)$$

where $Tr[\cdot]$ indicates the trace operation and the constraint $\text{diag}(W) = e$ is added to enforce $w_i^2 = 1$. The equivalence of the objective functions in (13) and (11) can be seen easily by replacing $w_i w_j$ with W_{ij} and dropping constant terms. The SDP in (13) may be solved in polynomial time using a primal-dual path following algorithm [4]. The result of this optimization W^* will in general be a non-integer symmetric positive semidefinite matrix. In [6], a randomized rounding methodology is proposed to recover a -1,1 vector w from the SDP solution W^*. The strategy is to first perform the Cholesky factorization $W^* = V^T V$. A random hyperplane through the origin with normal vector r is then chosen by selecting r from the uniform distribution on the surface of the unit hypersphere $S_n = \{r \in \mathbf{R}^{n+1} | r^T r = 1\}$. The value of w_i is then determined by whether the corresponding column v_i of V lies above or below the hyperplane, i.e., $w_i = 1$ if $v_i^T r \geq 0$ and $w_i = -1$ if $v_i^T r < 0$. The ith element of the vectorized adjacency sample $\hat{a}$ is then given by

$$\hat{a}_i = \begin{cases} 1, & \text{if } sign\left(v_i^T r\right) = sign\left(v_{n+1}^T r\right) \\ 0, & \text{if } sign\left(v_i^T r\right) \neq sign\left(v_{n+1}^T r\right). \end{cases} \quad (14)$$

This result can be seen by applying the rounding method and then using the change of variable formula given in (9).

We now proceed to derive the mean squared error $\mathbf{E}\left[||Q\hat{a}-b||_\Lambda^2\right]$ of the sample adjacency in (14) and thereby quantify how close the samples produced in this way come to satisfying the linear prior information on average. First note that the rounding scheme used implies the following identities:

$$\begin{aligned}\mathbf{E}[1+w_iw_j] &= 2P\left(sign\left(v_i^Tr\right)=sign\left(v_j^Tr\right)\right)\\ \mathbf{E}[1-w_iw_j] &= 2P\left(sign\left(v_i^Tr\right)\neq sign\left(v_j^Tr\right)\right)\end{aligned} \quad (15)$$

where r is a random vector from the uniform distribution on S_n as previously defined. We may evaluate the probabilities in (15) quite easily via the observation in [6]. Note that symmetry of the distribution implies $P\left(sign\left(v_i^Tr\right)\neq sign\left(v_j^Tr\right)\right) = 2P\left(v_i^Tr\geq 0, v_j^Tr<0\right)$. And if $\theta=\arccos\left(v_i^Tv_j\right)$ is the angle between the vectors v_i and v_j then it follows $P\left(v_i^Tr\geq 0, v_j^Tr<0\right)=\theta/2\pi$ since the distribution of r is uniform on S_n. A similar argument applies to the case of matching sign. The results are summarized below

$$\begin{aligned}P\left(sign\left(v_i^Tr\right)=sign\left(v_j^Tr\right)\right) &= 1-\frac{1}{\pi}\arccos\left(v_i^Tv_j\right)\\ P\left(sign\left(v_i^Tr\right)\neq sign\left(v_j^Tr\right)\right) &= \frac{1}{\pi}\arccos\left(v_i^Tv_j\right).\end{aligned} \quad (16)$$

If we define the matrix Z such that $Z_{ij}=\arccos(W^*_{ij})$ where W^* is the solution of the SDP relaxation in (13) and note that the objective function in (11) is exactly equal to $b^T\Lambda b-||Q\hat{a}-b||_\Lambda^2$, then we may take the expectation of the objective in (11) and apply the identities in (15) and (16) to obtain the mean squared error as

$$\mathbf{E}\left[||Q\hat{a}-b||_\Lambda^2\right] = ||Qe-b||_\Lambda^2 - \frac{1}{2\pi}Tr[(C-B)Z] \quad (17)$$

where e is a vector of ones.

We may obtain a bound on the expected value of the squared error in (17) independent of the solution to the SDP. As in [6], define the constant α

$$\alpha = \min_{z\in[0,\pi]}\frac{2}{\pi}\frac{z}{1-\cos z}. \quad (18)$$

From this definition of α, the following identities follow immediately:

$$\begin{aligned}\frac{1}{2}\alpha(1+\cos z) &\leq 1-\frac{1}{\pi}z\\ \frac{1}{2}\alpha(1-\cos z) &\leq \frac{1}{\pi}z.\end{aligned} \quad (19)$$

We take the expected value of the objective function in (11) and apply the identities in (19) with $Z_{ij}=\arccos(W^*_{ij})$ to give

$$\begin{aligned}&b^T\Lambda b-\mathbf{E}\left[||Q\hat{a}-b||_\Lambda^2\right]\\ &\geq \alpha\frac{1}{4}\left(\sum_{i,j}[B_{ij}+C_{ij}]+Tr[(B-C)W^*]\right)-e^TDe.\end{aligned} \quad (20)$$

Now suppose the equation $Qa=b$ has at least one feasible solution a^0. Let w^0 be the corresponding -1,1 vector and $W^0=w^0(w^0)^T$. We then have

$$\begin{aligned}0=||Qa^0-b||_\Lambda^2 &= e^TDe+b^T\Lambda b\\ &-\frac{1}{4}\left(\sum_{i,j}[B_{ij}+C_{ij}]+Tr[(B-C)W^0]\right).\end{aligned} \quad (21)$$

But since W^* solves the SDP in (13), it follows:

$$\begin{aligned}Tr[(B-C)W^*] &\geq Tr[(B-C)W^0]\\ &= 4e^TDe+4b^T\Lambda b-\sum_{i,j}[B_{ij}+C_{ij}].\end{aligned} \quad (22)$$

We may now combine the inequalities in (20) and (22) and rearrange to obtain a bound on the expected value of the squared error that is independent of the SDP solution

$$\mathbf{E}\left[||Q\hat{a}-b||_\Lambda^2\right] \leq (1-\alpha)\left(||Qe||_\Lambda^2+||b||_\Lambda^2\right). \quad (23)$$

In practice, the bound in (23) tends to exceed the true expected value in (17) by a large amount. However, it is of theoretical interest since it gives a general idea of how close samples produced in this way will come to satisfying the linear prior information, given the matrix Q and vector b specifying this information. One must be careful to apply this bound only when all elements of Q and b are nonnegative (such as when a vertex degree prior is used). A similar bound can be derived when some elements of Q or b are negative, but we will omit it here.

A naive procedure for generating the necessary samples using these procedures would be to first draw the ordering variables $\rho_{1:K}$ then fix the adjacency elements in $A^{x\rho}$ corresponding to the draw. One could then reduce the system $Q(\bar{A})=b$ by eliminating elements in $\bar{A}\cap A^{x\rho}$ and proceed to formulate and solve the SDP for use in randomized rounding. This approach is computationally prohibitive, however, because it requires solving a new SDP for every single sample. Instead, we prefer to solve a single SDP and use its solution to generate all samples. The single SDP is derived from the system $Q(\bar{A})=b$ where the eliminated variables A_{ij} are those whose probability of being in the set $A^{x\rho}$ exceeds a threshold. The probability $P\left(A_{ij}\in A^{x\rho}\right)$ is computed from the ordering distributions $P_k(\rho)$ as

$$P(A_{ij}\in A^{x\rho}) = \max_k \sum_{\rho|\exists l: x^l_{k\rho}=i, x^{l+1}_{k\rho}=j} P_k(\rho). \quad (24)$$

Note that by fixing the variables that are likely to be in $A^{x\rho}$ and eliminating them from the prior constraints $Q(\bar{A})=b$, we are throwing away some prior information. Provided the threshold is fairly high, the eliminated variables will most often be set to unity anyway due to the ordering samples. In the interest of keeping down computational costs, this is a reasonable approach.

There may be adjacency matrix elements that are not in $\bar{A}$ and have zero probability of being in $A^{x\rho}$. Define $A^o\equiv\left\{A_{ij}\mid A_{ij}\notin\bar{A}, P(A_{ij}\in A^{x\rho})=0\right\}$; A^o then denotes the adjacency matrix elements that we have no information about. We adopt the principle of parsimony and assume all elements in A^o are zero. A summary of our procedure for generating M sample adjacency matrices and orderings follows.

- Compute $P\left(A_{ij}\in A^{x\rho}\right)$ for all $A_{ij}\in\bar{A}$ as in (24).
- Eliminate $\left\{A_{ij}\mid P(A_{ij}\in A^{x\rho})\geq\delta\right\}$ from $\bar{A}$ and adjust the system $Q(\bar{A})=b$ with these variables fixed at 1.
- Solve the SDP corresponding to $Q(\bar{A})=b$ for the optimum W^*.
- Compute and store the Cholesky factor V of the SDP solution W^*.
- For $m=1,2,\ldots M$.
  — Draw ρ_k from $P_k(\rho)$ for $k=1,2,\ldots,K$.

— Determine $A^{x\rho}$ as in (4) and set $A_{ij} = 1$ for all $A_{ij} \in A^{x\rho}$.
— Draw r from the uniform distribution on S_n.
— Take inner products of the Cholesky factors with r to determine $A_{ij} \notin A^{x\rho}$ that are organized in the vector a as shown in (14).
— Set all remaining adjacency elements to 0.

We may now write down the conditional distribution $P(A|x_{1:K}, \rho_{1:K}, Q(\bar{A}) = b)$ from which the SDP rounding method is sampling. First define the set $H(A_{ij})$ as

$$H(A_{ij}) = \begin{cases} \left\{r \in S_n \mid sign\left(v_{ij}^T r\right) = sign\left(v_{n+1}^T r\right)\right\}, & \text{if } A_{ij} = 1 \\ \left\{r \in S_n \mid sign\left(v_{ij}^T r\right) \neq sign\left(v_{n+1}^T r\right)\right\}, & \text{if } A_{ij} = 0 \end{cases} \tag{25}$$

where S_n is the surface of the unit hypersphere and v_{ij} is the appropriate column of the Cholesky factor V corresponding to the variable A_{ij} as defined earlier. Since the only random elements of A given $x_{1:K}$, $\rho_{1:K}$ and $Q(\bar{A}) = b$ are those in $\bar{A} - A^{x\rho}$, the desired conditional distribution is given by

$$P\left(A|x_{1:K}, \rho_{1:K}, Q(\bar{A}) = b\right) = \frac{Vol\left(\bigcap_{A_{ij} \in \bar{A} - A^{x\rho}} H(A_{ij})\right)}{Vol(S_n)}. \tag{26}$$

The expression in (26) is a rather complicated distribution shaped by the prior data Q and b through the solution of the SDP in (13) formulated from this data. Luckily, we do not need to evaluate it. The crucial point is that samples from this distribution will approximately satisfy the prior information $Q(\bar{A}) = b$. In order to investigate the quality of individual samples, define the sublevel set S_ϵ of adjacency matrices as follows:

$$S_\epsilon = \left\{A \in \{0,1\}^{n \times n} \mid \frac{||Q(\bar{A}) - b||_\Lambda^2}{||Q(ee^T)||_\Lambda^2 + ||b||_\Lambda^2} \leq \epsilon\right\} \tag{27}$$

where we have resumed the operator notation for Q and ee^T is a matrix of ones used to coincide with the vectorized notation in (23). It follows from the definition that if $\epsilon_1 \leq \epsilon_2$ then $S_{\epsilon_1} \subseteq S_{\epsilon_2}$. Ideally we would prefer samples from S_0 so that the prior linear equalities are exactly satisfied, however we settle for adjacency samples from the larger set S_ϵ for some tolerance $\epsilon > 0$. For $\epsilon \geq 1 - \alpha \approx 0.88$, we can apply the Markov inequality along with the bound in (23) to give the following general result for sample adjacencies $\hat{A}$ produced using this method:

$$P(\hat{A} \in S_\epsilon) \geq 1 - \frac{1 - \alpha}{\epsilon}. \tag{28}$$

Although the method is capable (in principle) of producing any adjacency matrix in S_∞, a lower bound for the proportion of samples falling in the tolerance set S_ϵ is given by (28). One might investigate further the shape of the sampling distribution in (26) in order to determine the relative likelihood of different adjacencies, however we will conclude our analysis here.

C. Approximating the Endpoint Posterior

We use the topology and sensor ordering samples obtained in the previous section to derive an approximate endpoint posterior distribution of a passive measurement indexed by k as given in (1). If $\{A^m\}_{m=1}^M$ are the topology samples and $\{\rho_{1:K}^m\}_{m=1}^M$ are the sensor ordering samples (for each measurement), then the strong law of large numbers suggests a Monte Carlo estimate of the conditional expectation given by

$$\hat{P}(u_k|x_{1:K}, Q(\bar{A}) = b) = \frac{1}{\kappa} \sum_{m=1}^{M} \frac{P(y_k|u_k, \rho_k^m, A^m) P_k(u_k)}{\sum_u P(y_k|u, \rho_k^m, A^m) P_k(u)} \tag{29}$$

where κ is a normalization constant inserted to ensure the total mass of the approximate posterior is unity. Since we are given a distribution on the endpoints of the passive measurement $P_k(u)$, we need only specify a model for the conditional path probability $P(y|u, \rho, A)$ in order to approximate the posterior as in (29). Routing mechanisms and traffic data might figure prominently into such a model. We propose a simple model whereby the length of a path determines its probability (as in shortest path routing). If $|y_\rho|$ denotes the length of the ordered path y_ρ, and $y_\rho^{u,A}$ denotes the shortest ordered path between endpoints u in topology A, then the conditional distribution is given by

$$P(y|u, \rho, A) = \begin{cases} \theta, & \text{if } |y_\rho| = |y_\rho^{u,A}| < \infty \\ 1 - \theta, & \text{if } |y_\rho^{u,A}| < |y_\rho| < \infty \\ 0, & \text{if } |y_\rho| = \infty. \end{cases} \tag{30}$$

The model basically says that the shortest path between endpoints u in topology A is chosen with probability θ, and all other valid paths (that is, paths of finite length) have probability $1 - \theta$. If a path does not connect the endpoints u in the given topology A, then naturally it has zero probability. Note that for arbitrary θ, we need to run Dijkstra's algorithm (or some other shortest path routing algorithm) for each topology sample A^m in order to compute the conditional path probability in (30) [1]. This is not necessary, however, in the case that $\theta = 1/2$.

We may give maximum a posteriori (MAP) estimates of the endpoints u_k of a passive measurement y_k after computing the posterior distribution estimate in (29). Indeed, the MAP estimate is simply given by

$$\hat{u}_k = \arg\max_{u \in \Sigma \times \Delta} \hat{P}(u|x_{1:K}, Q(\bar{A}) = b). \tag{31}$$

Recall that $u \equiv (s, d)$, thus MAP estimates of s_k or d_k individually may be obtained by maximizing the appropriate marginal $\hat{P}(s|x_{1:K}, Q(\bar{A}) = b)$ or $\hat{P}(d|x_{1:K}, Q(\bar{A}) = b)$, respectively.

We use as an error measure the ratio $\Lambda_u(k)$ below for the estimated endpoints $\hat{u}_k$

$$\Lambda_u(k) = \frac{\hat{P}(\hat{u}_k|x_{1:K}, Q(\bar{A}) = b)}{\hat{P}(\hat{u}_k|x_{1:K}, Q(\bar{A}) = b) + \max_{u \in \Sigma \times \Delta - \hat{u}_k} \hat{P}(u|x_{1:K}, Q(\bar{A}) = b)}. \tag{32}$$

It is also useful to compute the corresponding ratios associated with the marginalized distributions $\Lambda_s(k)$ and $\Lambda_d(k)$, as it may be the case that either the source or destination of a passive measurement is more accurately determined individually than are both collectively. These are defined exactly as in (32), except

u is replaced with s or d throughout (so that the appropriate marginal distribution is considered). It is clear that the ratio in (32) must lie in the interval $[1/2, 1]$. Larger values of this ratio in a sense indicates more 'confidence' in the MAP estimate since a value of 1 is achieved only when all of the mass of the estimated posterior distribution is concentrated at the MAP estimate.

Algorithm Complexity

We now analyze the complexity of the source/destination estimation scheme developed here. The two fundamental quantities that determine the size of the problem are denoted by N and h; N is the total number of sensors plus probing sites, so that $N = |\Gamma| + |\Sigma \cup \Delta|$, while h is the maximum number of activated sensors in any measurement, so that $|y_k| \leq h$ for all $k = 1, 2, \ldots, K$. The maximum number of hops h may be a function of N, depending upon the type of network considered. For networks that obey the small world effect, as many real world networks do, h will remain approximately constant with increasing N [20], [30]. The number of measurements K and the number of Monte Carlo samples M also affect the complexity; however, we shall see the complexity dependence on these is always linear.

First note that we must store the ordering distributions $P_k(\rho)$ for all measurements. Since each distribution is defined over $O(h!)$ orderings, this requires $O(Kh!)$ space. The adjacency matrix A considers all logical connections among sensors and probing sites, so that A has $O(N^2)$ elements. In the worst case, the linear prior information $Q(\bar{A}) = b$ will constrain all elements of this matrix so that $\bar{A} = A$. It will therefore take $O(KN^2h!)$ time to compute $P\left(A_{ij} \in A^{x\rho}\right)$ for all $A_{ij} \in \bar{A}$. Now in the worst case, thresholding these probabilities will produce a negligible reduction in the size of the system $Q(\bar{A}) = b$, so that we still have to contend with $O(N^2)$ variables in solving the SDP relaxation. Typically interior point methods are used to solve SDP's to within ϵ of the optimal solution. These are based on Newton's method; therefore at each iteration it is necessary to solve a linear system of equations for the Newton directions ($O(n^3)$ for a system of size n). An algorithm given in [31] is shown to take $O(|\log\epsilon|\sqrt{n})$ iterations for a problem of size n-this performance is typical for all interior point algorithms. Our problem has dimension $O(N^2)$, thus solving the SDP takes $O((N^2)^{3.5})$ or $O(N^7)$ time. A Cholesky factorization is then performed on the SDP solution, which takes $O((N^2)^3)$ or $O(N^6)$ time.

After solving the SDP, the M topology and ordering samples may be produced relatively quickly. For each sample, we need to draw an ordering for each of the K measurements, thus requiring $O(MK)$ time to produce the ordering samples. Given the K orderings for a single sample, $A^{x\rho}$ may be generated in $O(Kh)$ time. Finally, we may draw the vector r and take inner products to determine the remaining elements of the topology sample. Since the time required for each inner product is linear, it takes a total of $O(MN^2 + MKh)$ time to produce the M topology samples.

The final step is to compute the Monte Carlo approximation of the endpoint posterior distribution of a passive measurement. A quick inspection of (29) reveals that we need to determine the conditional path probabilities $P(y|u, \rho, A)$ for every endpoint pair u-there are $O(N^2)$ such pairs. Also, computing each path probability for a given ordered path y_ρ requires tracing this path through the topology A, which takes $O(h)$ time. Now, if $\theta \neq 1/2$ we must take $O(MN^3)$ time to run a shortest path algorithm on each sample [1]. Therefore, it takes $O(MN^2h + MN^3)$ to produce the approximate endpoint posterior for $\theta \neq 1/2$; this reduces to $O(MN^2h)$ for $\theta = 1/2$.

The factors that give some cause for concern in this algorithm are the $h!$ in considering all possible orderings and the N^7 in the SDP solution complexity. If we are dealing with small world networks, then h might be around four or five so that $h!$ is still manageable. And if this is not the case, one would hope that the ordering distributions $P_k(\rho)$ are nonzero only over a reasonable number of orderings since we need only consider ρ with $P_k(\rho) > 0$. In practice, the actual SDP complexity is likely to be significantly less than the worst case bound of $O(N^7)$ after reducing the system $Q(\bar{A}) = b$, especially if the original prior only constrains some small subset of the adjacency elements. Our algorithm would still benefit from speedy SDP algorithms as solving the relaxation takes the most time in the worst case. A parallel implementation of an interior point algorithm for SDP's might reduce the time requirements if multiple processors are available [32].

III. Simulations

We performed some numerical simulations to demonstrate the utility of the method described in this paper. We generated undirected random graphs with 25 nodes to serve as test networks. The number of edges in each graph was fixed at 40 by randomly selecting 40 of the possible 300 vertex pairs and connecting the selected pairs by an edge. We randomly chose 12 of the 25 nodes to serve as probing sites-this set was then partitioned in half so that both the source set Σ and destination set Δ each had 6 distinct elements. Sensors were placed on links in the network for two cases: 100% sensor coverage (in which all 40 links were monitored by a sensor) and 75% sensor coverage (in which 30 of the 40 links were selected at random for hosting a sensor). In the 75% coverage case, networks were generated in a rejection sampling manner so that every measurement (whether passive or active) activated at least one sensor.

In order to probe a network, we randomly selected 18 of the 36 distinct pairs in $\Sigma \times \Delta$ to serve as endpoints for active measurements. This set of 18 endpoint pairs is denoted $L \subset \Sigma \times \Delta$; the remaining pairs are denoted by $L^c \equiv \Sigma \times \Delta - L$. Sensor activations in response to transmissions between all pairs in $\Sigma \times \Delta$ were observed in the monitoring phase. All transmissions were routed through the network using shortest path routing, and activated sensor sets y_k were observed. Thus for each network we had $K = 54$ data points: $K_o - 1 = 18$ active measurements $x_{1:18} \equiv (u_{1:18}, y_{1:18}, P_{1:18}(\rho))$ and 36 passive measurements $x_{19:54} \equiv (y_{19:54}, P_{19:54}(\rho))$. For each data point ($k = 1, 2, \ldots K$), a distribution on the order in which sensors were activated $P_k(\rho)$ was generated as follows: first the true ordering of sensors ρ_k was noted, then noise $n(\rho)$ was drawn independently from the $Uniform[0, 0.2]$ distribution for

$\rho = 1, 2, \ldots |y_k|!$, finally the distribution $P_k(\rho)$ was generated by normalizing the corrupted delta function distribution as

$$P_k(\rho) = \frac{\delta(\rho - \rho_k) + n(\rho)}{\sum_{\rho=1}^{|y_k|!} \delta(\rho - \rho_k) + n(\rho)}. \quad (33)$$

The linear prior information was generated from degree information on the logical topology A. Indeed vertex degree information is a commonly used special case of the more general linear prior specified by $Q(\bar{A}) = b$ [21], [26]. The sensor degree, that is the number of sensors b_i to which the ith vertex in the logical topology is adjacent, was known for all $v_i \in V_A$. In addition to knowing the sensor degrees of vertices in the logical topology G_A, a random subset consisting of no more than 60% of the sensors not adjacent to a given vertex were also known. For the ith vertex, the ith row of the operator $Q_i(\bar{A})$ therefore sums over the elements of A for which adjacency to vertex i is uncertain, and the ith element of b, b_i, is simply the known sensor degree of vertex i. As an example, consider vertex γ_5 of the logical topology in Fig. 3. Vertex γ_5 is adjacent to sensors $\{\gamma_2, \gamma_4\}$, therefore its sensor degree is two. Since there are two sensors not adjacent to γ_5, $\lfloor 2 * 60\% \rfloor = 1$ sensor, say γ_x, is selected at random from the set $\{\gamma_1, \gamma_3\}$. Let Γ_i denote the set of sensors known to be nonadjacent to vertex i, so that $\Gamma_{\gamma_5} = \{\gamma_x\}$ in this example. The row of the prior $Q(\bar{A}) = b$ corresponding to γ_5 is then given by $\sum_{j \in \Gamma - \gamma_5 - \Gamma_{\gamma_5}} A_{\gamma_5 j} = 2$. The remaining rows are constructed in a similar fashion.

Given the sensor degree prior information and the ordering distributions, we eliminated those adjacency elements whose probability of being in the set $A^{x\rho}$ exceeded 1/2 from $\bar{A}$, where $P(A_{ij} \in A^{x\rho})$ was computed as in (24). The reduced system $Q(\bar{A}) = b$ was then used to formulate the SDP relaxation in (13) for the minimum norm solution with the weight matrix Λ taken as the identity. The relaxation was solved with a predictor-corrector path following algorithm given in [4]. A publicly available C implementation of this algorithm was used [33]. The SDP solution was used along with the ordering distributions $P_k(\rho)$ to produce $M = 500$ samples of measurement orderings $\rho_{1:K}$ and adjacency matrices A for computing the Monte Carlo estimates of the endpoint posteriors.

We assumed the endpoint priors $P_k(u)$ were uniform over $\Sigma \times \Delta$ for all 36 passive measurements $k = 19, 20, \ldots, 54$. Also, the parameter θ in the conditional path probabilities of (30) was taken as 1/2 so that it was not necessary to run a shortest path routing algorithm on every sample topology. The 500 ordering and topology samples were then used to compute the approximate endpoint posteriors for all passive measurements as given in (29). These were used to produce joint MAP estimates of the transmission endpoints and to compute the resolution measures $\Lambda_u(k)$. An example endpoint posterior is given in Fig. 3, for which the correct endpoint pair is source no. 6 and destination no. 3. It is clear that the MAP estimate will result in the correct pair in this case. Also indicated in the figure is the second most likely pair $u = (6, 5)$; this is used in computing the resolution measure Λ_u as in (32)-$\Lambda_u(k) = 0.60$ for this case. Marginal distributions are obtained by summing the approximate posterior over either source or destination. These were used in individual MAP estimation of source and destination. It is clear that the individual estimates will match the joint estimate for this case; the resolution measures were a bit lower though with $\Lambda_s(k) = 0.58$ and $\Lambda_d(k) = 0.59$. This completes the simulation process for a single graph.

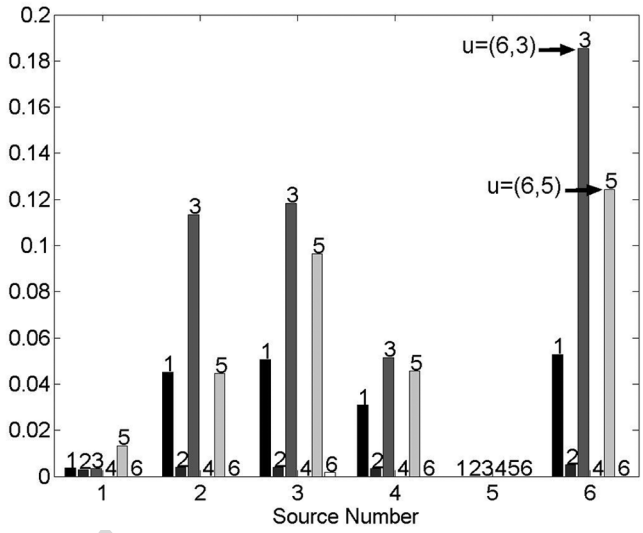


Fig. 3. Example endpoint posterior distribution $\hat{P}(u_k | x_{1:K}, Q(\bar{A}) = b)$ for a passive measurement at time $k \geq K_o$ with endpoints $u = (s, d) = (6, 3)$. The probabilities are grouped by source, with each of 6 bars in a group corresponding to a different destination (noted above the individual bar). The largest and second largest values of the posterior are indicated—it is these values that are used in computing the resolution ratio Λ_u of (32), calculated as $\Lambda_u(k) = 0.60$. It is clear in this example that the endpoints of this transmission will be correctly estimated by the joint MAP estimate.

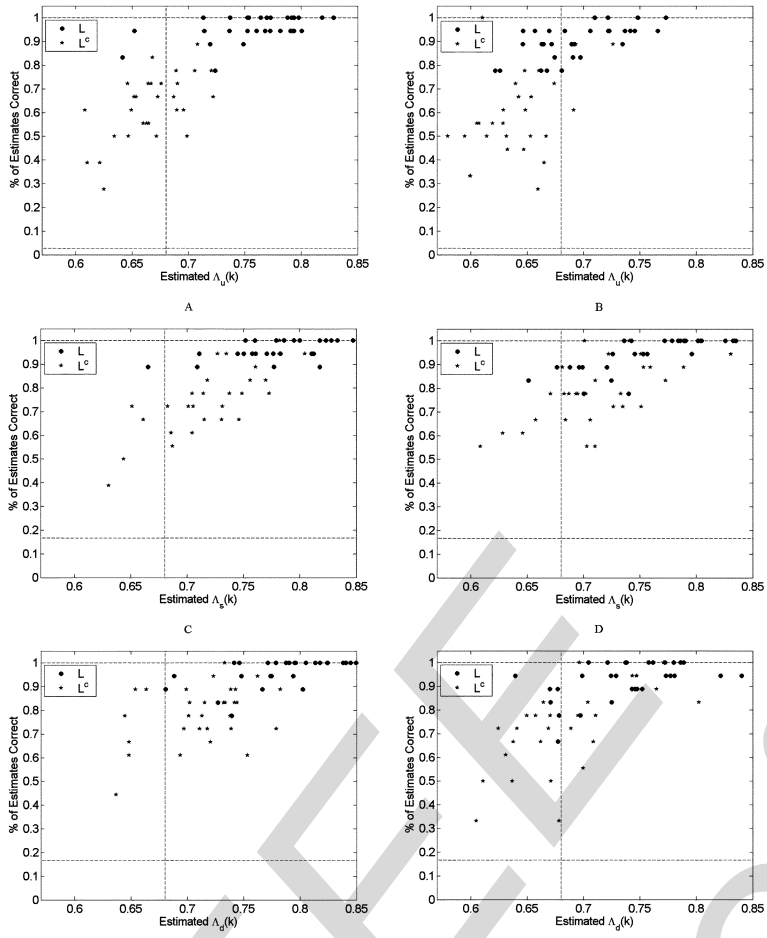


Fig. 4. Plots of proportion of endpoint estimates correct for a given set (L or L^c) versus the resolution ratios of (32) averaged over the corresponding set for the two simulation cases: 100% sensor coverage in the first column and 75% sensor coverage in the second. Circles indicate averages over paths from set L and pentagrams indicate averages over paths from set L^c. The first row (Λ_u) is for joint MAP estimation of $u_k = (s_k, d_k)$ from joint distribution $\hat{P}(u_k|x_{1:K}, Q(\bar{A}) = b)$. The second row ($\Lambda_s$) is for individual estimation of s_k from marginal distribution $\hat{P}(s_k|x_{1:K}, Q(\bar{A}) = b)$. The third row ($\Lambda_d$) is for individual estimation of d_k from marginal distribution $\hat{P}(d_k|x_{1:K}, Q(\bar{A}) = b)$. Some reference lines are also plotted: a horizontal line indicating the chance line for randomly selecting endpoints (1/36 for joint estimation and 1/6 for individual estimation), and a vertical line at 0.68. Note that above $\Lambda(k) = 0.68$, an approximately linear behavior is observed. This behavior is somewhat washed out for the marginalized estimates, however marginalizing tends to increase the percent of correct estimates. It is not surprising that there appears to be some degradation in the quality of the estimates when only 75% of the links are equipped with sensors. A: Λ_u, 100% coverage. B: Λ_u, 75% coverage. C: Λ_s, 100% coverage. D: Λ_s, 75% coverage. E: Λ_d, 100% coverage. F: Λ_d, 75% coverage.

We repeated the simulation procedure for 30 networks with 100% sensor coverage and 30 networks with 75% sensor coverage. The SDP randomized rounding algorithm was quite effective for producing topology samples that approximately agree with the sensor degree prior information. We computed the normalized squared topology sample error $(1/||Qe||^2 + ||b||^2)(1/M)\sum_m ||Q\hat{a}^m - b||^2$ averaged over the $M = 500$ samples along with the normalized expected squared error $(1/||Qe||^2 + ||b||^2)\mathbf{E}\left[||Q\hat{a} - b||^2\right]$ as in (17) for each graph. When averaged over all graphs with 100% coverage, these agreed at 0.022 with two significant figures. For 75% coverage, the expected and observed errors agreed at 0.013 with two significant figures. The bound derived in (23) assures the expected squared error can never exceed $1 - \alpha \approx 0.12$. We see that graphs with 75% sensor coverage tend to have lower error values.

Plots of proportion of passive measurement endpoint estimates correct for a given set (L or L^c) versus the resolution ratio from (32) averaged over the corresponding set are given in Fig. 4. Plots are shown for joint estimates of u_k via the joint distribution as well as for individual estimates of s_k and d_k from the marginals. We observe an approximately linear relation between the proportion of correct estimates and the appropriate Λ ratio when the Λ ratio exceeds 0.68. In this regime, the Λ ratio might be used as a measure of confidence in the endpoint estimates. Also note that transmissions in set L tend to have higher Λ ratios (and are correct more often) than those in set L^c because it is the transmissions in set L that are used in training the probing sites. We see that marginalized MAP estimates are often better than joint MAP estimates. Marginalization certainly blurs the linear relation in the higher confidence regime. We also observe some degradation in the quality of the estimates when only 75% of the links are equipped with sensors; this is to be expected though. Recall that these results are obtained with completely random placement of sensors and random choices for the (s, d) pairs to use in the probing phase. These two factors will clearly affect the estimates of passive measurement endpoints, and therefore provide an interesting direction for future work.

IV. Summary and Extensions

In this paper, we have developed a methodology for estimating the endpoints of a transmission in a network using link-level transmission interceptions. The estimation is done using Monte Carlo simulation in a Bayesian framework. A semidefinite programming relaxation is used to generate logical network topology samples that approximately agree with linear prior information. It is possible to envision applications of the method in all sorts of networks, or systems with key features modeled by networks. We have displayed simulations of its utility on some random networks. We now discuss some extensions of the theory presented here and possibilities for future work on this problem.

It is possible to extend our algorithm for source/destination estimation to the cases of noisy sensors and sensor excitation due to multiple transmissions without much trouble. Consider first when the sensors are noisy: then the observed set of activated sensors y may not match the true set of sensors $\tilde{y}$ passed by a particular transmission. Suppose that each sensor $\gamma \in \Gamma$ has an associated miss probability $\alpha_m(\gamma) = P(\gamma \notin y | \gamma \in \tilde{y})$ and false alarm probability $\alpha_f(\gamma) = P(\gamma \in y | \gamma \notin \tilde{y})$. The probing mechanism then repeats the data transmission from σ_k to $\delta_k N$ times for each k. These N measurements are used to construct a maximum likelihood estimate $\hat{y}_k$ of each path $\tilde{y}_k$ according to the following model. Along the lines of a generalized likelihood approach, the measurement mechanism passes along the maximum likelihood path estimates for each $\tilde{y}_k$ for use in approximating the endpoint posterior. Note that we will likely have to settle for $N = 1$ for passive measurements.

Define the path indicator vector ν whose elements are given by $\nu(j) = I_y(\gamma_j)$ for all $j = 1, 2, \ldots |\Gamma|$ where $I_A : A \to \{0, 1\}$ is the usual indicator function. If we assume sensor errors are independent across paths and measurements, then the joint probability mass function of the N observed path vectors for a given source/destination pair ν_i is

$$\begin{aligned} &P(\nu_1, \nu_2, \ldots \nu_N | \tilde{\nu}) \\ &\quad = \prod_{i=1}^{N} \prod_{j=1}^{|\Gamma|} \alpha_m(\gamma_j)^{(1-\nu_i(j))\tilde{\nu}(j)} \beta_m(\gamma_j)^{\nu_i(j)\tilde{\nu}(j)} \\ &\quad \times \alpha_f(\gamma_j)^{\nu_i(j)(1-\tilde{\nu}(j))} \beta_f(\gamma_j)^{(1-\nu_i(j))(1-\tilde{\nu}(j))} \end{aligned} \tag{34}$$

where $\beta_m(\gamma) \equiv 1 - \alpha_m(\gamma)$ and $\beta_f(\gamma) \equiv 1 - \alpha_f(\gamma)$. If we define the likelihood function $L(\tilde{\nu})$ as the logarithm of the expression in (34), then it may be written explicitly as

$$\begin{aligned} L(\tilde{\nu}) = &\sum_{j=1}^{|\Gamma|} \left(N \log \beta_f(\gamma_j) + \sum_{i=1}^{N} \nu_i(\gamma_j) \log \frac{\alpha_f(\gamma_j)}{\beta_f(\gamma_j)} \right) \\ &+ \sum_{j=1}^{|\Gamma|} \left(N \log \frac{\alpha_m(\gamma_j)}{\beta_f(\gamma_j)} + \sum_{i=1}^{N} \nu_i(\gamma_j) \log \frac{\beta_e(\gamma_j)}{\alpha_e(\gamma_j)} \right) \\ &\times \tilde{\nu}(\gamma_j) \end{aligned} \tag{35}$$

where $\alpha_e(\gamma) \equiv \alpha_m(\gamma)\alpha_f(\gamma)$ and $\beta_e(\gamma) \equiv \beta_m(\gamma)\beta_f(\gamma)$. Since only the second term in (35) depends on $\tilde{\nu}$ and $\tilde{\nu} \in \{0, 1\}^{|\Gamma|}$, the maximum likelihood path estimate may be written quite compactly as

$$\hat{y} = \left\{ \gamma_j \in \Gamma \mid N \log \frac{\alpha_m(\gamma_j)}{\beta_f(\gamma_j)} + \sum_{i=1}^{N} \nu_i(j) \log \frac{\beta_e(\gamma_j)}{\alpha_e(\gamma_j)} \geq 0 \right\}. \tag{36}$$

As another extension, suppose that for passive measurements the activated sensor set y_k is due to transmissions passed between n source/destination pairs u_{ki} for $i = 1, 2, \ldots, n$ where n is known. The strategy here is to introduce a random variable η_k for each passive measurement that represents a partition of the activated sensor set y_k into sets y_{ki} for $i = 1, 2, \ldots, n$, where the sensors in each y_{ki} are activated in response to a single transmission. We may then split the single measurement y_k into n different passive measurements y_{ki} according to the value of the partition variable η_k and proceed with the previous theoretical development. In this case, the endpoint posterior of (1) becomes

$$\begin{aligned} &P(u_k | x_{1:K}, Q(\bar{A}) = b) \\ &\quad = \mathbf{E}_{A, \rho_{1:K}, \eta_{K_o:K}} \\ &\qquad \times \left[\frac{P(y_k | u_k, \rho_k, A) P_k(u_k)}{\sum_u P(y_k | u, \rho_k, A) P_k(u)} \mid x_{1:K}, Q(\bar{A}) = b \right] \end{aligned} \tag{37}$$

where we must now also take the expectation over partition variables $\eta_{K_o:K}$ of all passive measurements. The first step of the Monte Carlo sampling would then be to draw a partition variable for each passive measurement from some (presumably available) distribution $P_k(\eta)$. Given the partition variable, appropriate orderings may be drawn and so on as before.

One can similarly account for the case of random linear prior information $Q(\bar{A}) = b$. Suppose that instead of being given a fixed operator Q and vector b, we are given a distribution on these $P(Q, b)$. This might occur, for example, when we know that the vertex degrees follow a power-law distribution [25]- in which case a distribution on b is induced. We must now also take the expectation over Q and b, so that the endpoint posterior becomes

$$\begin{aligned} &P(u_k | x_{1:K}) \\ &\quad = \mathbf{E}_{A, \rho_{1:K}, Q, b} \left[\frac{P(y_k | u_k, \rho_k, A) P_k(u_k)}{\sum_u P(y_k | u, \rho_k, A) P_k(u)} \mid x_{1:K} \right]. \end{aligned} \tag{38}$$

A Monte Carlo approximation of (38) would therefore require drawing Q and b then proceeding as before. Unfortunately, a new SDP must be solved for every Q and b in order to produce topology samples A. If the SDP relaxation is not too large, this might be reasonable. If the size is prohibitive, one might approximate the expectation by selecting only a few of the most likely realizations of (Q, b) and solving the SDP for these. The distribution $P(Q, b)$ is then restricted to be nonzero only at elements of this preselected dictionary so that the Monte Carlo simulation selects those only those values for which we have already solved the SDP.

An interesting direction for future work would be to develop an adaptive probing scheme. It is obvious that the quality of endpoint estimates for suspect transmissions will depend on which

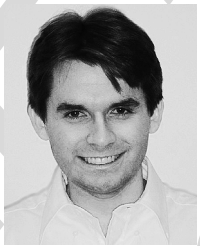
endpoints were used in the probing phase. The idea here is to use the approximate endpoint posterior distributions to suggest additional active measurements that should be made in order to improve the estimates. One can hypothesize criteria for determining the new probing pairs. For example, nodes that tend to have similar posterior probabilities over several suspect paths might be selected for probing so as to distinguish them more explicitly in the constraints. The question of efficient online implementation naturally arises in this context. A forgetting factor could be used in conjunction with existing topology and ordering samples so that an entirely new batch would not be required at each probing cycle.

ACKNOWLEDGMENT

The first author is also grateful to M. Rabbat and S. Sridharan for valuable discussions related to this work.

REFERENCES

[1] C. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
[2] L. Lovasz and A. Schrijver, "Cones of matrices and set-functions and 0–1 optimization," *SIAM J. Optim.*, vol. 1, no. 2, pp. 166–190, 1991.
[3] R. Saigal, *Linear Programming: A Modern Integrated Analysis*. Norwell, MA: Kluwer, 1995.
[4] C. Helmberg, F. Rendl, R. Vanderbei, and H. Wolkowicz, "An interior-point method for semidefinite programming," *SIAM J. Optim.*, vol. 6, no. 2, pp. 342–361, May 1996.
[5] F. Alizadeh, "Interior point methods in semidefinite programming with applications to combinatorial optimization," *SIAM J. Optim.*, vol. 5, no. 1, pp. 13–51, 1995.
[6] M. Goemans and D. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, Nov. 1995.
[7] M. Goemans and F. Rendl, "Semidefinite programming in combinatorial optimzation," in *Handbook of Semidefinite Programming*, H. Wolkowicz, R. Saigal, and L. Venberghe, Eds. Norwell, MA: Kluwer, 2000, pp. 343–360.
[8] C. Helmberg, "Fixing variables in semidefinite relaxations," *SIAM J. Matrix Anal. Apps.*, vol. 21, no. 3, pp. 952–969, 2000.
[9] A. Tanenbaum, *Computer Networks*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.
[10] Y. Vardi, "Network tomography: Estimating the source-destination traffic intensities from link data," *J. Amer. Statist. Assoc.*, vol. 91, pp. 365–377, 1996.
[11] G. Liang and B. Yu, "Maximum pseudo likelihood estimation in network tomography," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2043–2053, Aug. 2003.
[12] M.-F. Shih and A. Hero, "Unicast-based inference of network link delay distributions with finite mixture models," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2219–2228, Aug. 2003.
[13] Y. Tsang, M. Coates, and R. Nowak, "Network delay tomography," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2125–2136, Aug. 2003.
[14] R. Caceres, N. Duffield, J. Horowitz, and D. Towsley, "Multicast-based inference of network-internal loss characteristics," *IEEE Trans. Inform. Theory*, vol. 45, no. 7, pp. 2462–2480, Nov. 1999.
[15] M. Coates, R. Castro, and R. Nowak, "Maximum likelihood network topology identification from edge-based unicast measurements," *ACM Sigmetric*, Jun. 2002.
[16] N. Duffield, J. Horowitz, F. L. Presti, and D. Towsley, "Multicast topology inference from measured end-to-end loss," *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 26–45, Jan. 2002.
[17] M. Coates, A. Hero, R. Nowak, and B. Yu, "Internet tomography," *IEEE Signal Process. Mag.*, vol. 19, no. 3, pp. 47–65, May 2002.
[18] J. Treichler, M. Larimore, S. Wood, and M. Rabbat, "Determining the topology of a telephone system using internally sensed network tomography," in *Proc. 11th Digital Signal Processing Workshop*, Aug. 2004.
[19] M. Rabbat and R. Nowak, Telephone Network Topology Inference Univ. Wisconsin, Madison, 2004, Tech. Rep..
[20] M. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, pp. 167–256, 2003.
[21] Q. Morris, B. Frey, and C. Paige, "Denoising and untangling graphs using degree priors," *Proc. Neural Info. Process. Sys.*, no. 16, 2003.
[22] A. Pasztor and D. Veitch, "A precision infrastructure for active probing," in *Proc. Workshop on Passive and Active Networking*, Apr. 2001.
[23] ——, "PC based precision timing without GPS," in *Proc. ACM SIGMETRICS*, Jun. 2002.
[24] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
[25] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in *Proc. CM SIGCOMM*, Aug. 1999, pp. 251–262.
[26] S. Gomez and A. Rzhetsky, "Towards the prediction of complete protein-protein interaction networks," in *Proc. Pacific Symp. Biocomputing*, 2002, pp. 413–424.
[27] E. Moore, "Gedanken-experiments on sequential machines," in *Automata Studies, Annals of Mathematics Studies*. Princeton, NJ: Princeton Univ. Press, 1956, pp. 129–153.
[28] D. Lee and M. Yannakakis, "Principles and methods of testing finite state matchines—A survey," *Proc. IEEE*, vol. 84, no. 8, pp. 1090–1123, Aug. 1996.
[29] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA: Freeman, 1979.
[30] S. Milgram, "The small world problem," *Psychol. Today*, vol. 2, pp. 60–67, 1967.
[31] C.-J. Lin and R. Saigal, A Predictor Corrector Method for Semidefinite Linear Programming Dept. Industrial and Operations Engineering, Univ. Michigan, Ann Arbor, MI, 1995, Tech. Rep. TR95-20.
[32] M. Nayakkankuppam and Y. Tymofyeyev, "A parallel implementation of the spectral bundle method for large-scale semidefinite programs," in *Proc. 8th SIAM Conf. App. Lin. Alg.*, 2003.
[33] B. Borchers, "CSDP: A C library for semidefinite programming," *Optim. Meth. Softw.*, vol. 11, no. 1, pp. 613–623, 1999.



Derek Justice received the B.S. degree in electrical and computer engineering and the B.S. degree in physics from North Carolina State University, Raleigh, in 2003 and the M.S. degree in electrical engineering systems from the University of Michigan, Ann Arbor, in 2004. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering and Computer Science at the University of Michigan.

He was with the Space and System Engineering Group, Los Alamos National Laboratory, Los Alamos, NM, in 2003. From 1998 to 2000, he was with the North Carolina School of Science and Mathematics, Durham, NC. his research interests include inference and scheduling in sensor networks and other systems with combinatorial aspects.

Mr. Justice was a Goldwater Scholar and was inducted into the Phi Beta Kappa and Phi Kappa Phi honor societies.

Alfred Hero (S'79–M'80–SM'96–F'98) received the B.S. degree in electrical engineering (Hons.) from Boston University, Boston, MA, in 1980, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1984.

Since 1984, he has been with the University of Michigan, Ann Arbor, where he is a Professor in the Department of Electrical Engineering and Computer Science and, by courtesy, in the Department of Biomedical Engineering and the Department of Statistics. He has held visiting positions at I3S University of Nice, Sophia-Antipolis, France in 2001; Ecole Normale Supérieure de Lyon, Lyon, Frqnce in 1999; Ecole Nationale Supérieure des Télécommunications, Paris, in 1999; Scientific Research Labs of the Ford Motor Company, Dearborn, MI, in 1993; Ecole Nationale Superieure des Techniques Avancees (ENSTA), Ecole Superieure d'Electricite, Paris, France, in 1990; and the M.I.T. Lincoln Laboratory from 1987–1989. His research interests include inference for sensor networks, bioinformatics, and statistical signal and image processing.

Dr. Hero is a member of Tau Beta Pi, the American Statistical Association (ASA), the Society for Industrial and Applied Mathematics (SIAM), and the U.S. National Commission (Commission C) of the International Union of Radio

Science (URSI). He has received a IEEE Signal Processing Society Meritorious Service Award (1998), IEEE Signal Processing Society Best Paper Award (1998), the IEEE Third Millennium Medal and a 2002 IEEE Signal Processing Society Distinguished Lecturership. He is currently President of the IEEE Signal Processing Society.