Statistical Inference for Internal Link Parameters in a Network

R. Cáceres* N.G. Duffield† J. Horowitz‡ D. Towsley§

Abstract

Performance measurements on individual links in a large network such as the Internet are expected to suffer from poor scaling properties. Furthermore, the results of measurements are not guaranteed to be available to end users. To overcome these problem we develop statistical techniques for estimating packet loss rates and other characteristics of internal links based on observations at the endpoints of a tree connecting a source and multiple receivers. We consider a specific model for packet transmission through the tree, show that it is identifiable, determine the MLEs for the link loss rates, and show them to be asymptotically normal and consistent. The methods are illustrated with examples of simulated network traffic.

1 Introduction

Two approaches to evaluating performance in large communications networks, such as the Internet, have been to use (i) network management packages to collect data at internal nodes and report on link-level performance, and (ii) endpoint measurements of point-to-point traffic behavior to characterize network performance.

The first approach suffers from the drawback that it may be difficult to gain access to a wide range of routers in a large network, and introducing new measurement tools into the routers would require companies to make changes in their products. Moreover, the problem of integrating link-level information to obtain a picture of overall network performance is not completely understood.

The second approach is an area of continuing investigation. `pathchar` [3] is under evaluation for inferring link-level statistics from end-to-end point-to-point measurements. Several measurement infrastructure projects in progress are based on the exchange of unicast probe packets on a mesh of network paths between measurement servers; see [1, 2, 4, 10]). `mtrace` reports per hop statistics from a multicast source to a receiver. For differing reasons, these approaches potentially suffer from poor scaling properties in large networks, due to the volume of measurement-related traffic required to obtain a comprehensive picture of network performance.

In this paper, we consider a new approach to link-level loss behavior in a network based on *multicast* probe traffic, that is, traffic between a source and several receivers. When a probe is multicast, only one copy of it traverses each link on the tree that describes the path through the network from the source to the receivers. (With unicast measurement, there would be one probe for each receiver). Multicast introduces dependence in the losses measured at the receivers, which can, in turn, be used to infer the loss rates at links in the routing tree spanning the sender and receivers. Apart from the analytic possibilities of multicast-based measurements, the use of multicast probes introduces less traffic that unicast probes: it has better scaling properties for large networks. We envisage implementation of our method on infrastructural projects such as those cited above.

The results presented here are as follows. We derive maximum likelihood estimators (MLEs) of the link loss rates based on a sequence of n independent probes, under the assumption that losses are independent on different links. These estimators are shown to be strongly consistent and asymptotically normal as the number of probes tends to infinity. We present two simple examples based on simulated network traffic. In the first, the assumption of independent losses is built into the model, and convergence of the MLEs to their target parameters is rapid. The second, more realistic, example, illustrates losses due to queue overflows, in which case there are dependencies between the

*AT&T Labs–Research, Rm. A173, 180 Park Avenue, Florham Park, NJ 07932, USA; E-mail: ramon@research.att.com

†AT&T Labs–Research, Rm. A175, 180 Park Avenue, Florham Park, NJ 07932, USA; E-mail: duffield@research.att.com

‡Department of Mathematics & Statistics Lederle Graduate Research Tower, Box 34515 University of Massachusetts Amherst, MA 01003-4515 USA; E-mail: joeh@math.umass.edu

§Dept. of Computer Science University of Massachusetts Amherst, MA 01003-4610, USA; E-mail: towsley@cs.umass.edu

links which cause a slight asymptotic bias in the MLEs. The extent of this bias and some approaches to overcoming it are discussed at the end of the paper.

Before continuing, we draw the reader's attention to two articles on different problem in network inference; that of inferring source-destination traffic intensities from a set of link-aggregated rates in a network; see [12, 13].

2 Description of the Basic Model

Let $\mathcal{T} = (V, L)$ denote the *logical* (as opposed to physical) multicast tree, consisting of the set of nodes V, including the source and receivers, and the set of links L, which are ordered pairs (j, k) of nodes, indicating a (directed) link from j to k. The set of *children* of node j is denoted by $d(j)$; these are the nodes with a link coming from j. For each node j, other than the root 0, there is a unique node $f(j)$, the *parent* of j, such that $j \in d(f(j))$. Each link can therefore be identified by its "child" endpoint. We define "ancestors" (grandparents and the like) in an obvious way, and likewise "descendants". The difference between a logical and a physical tree is that, whereas it is possible for a node to have only one child in the physical tree, in the logical tree each node must have at least two children. A physical tree can be converted into a logical tree by deleting all nodes, other than the root, which have an only child and adjusting the links accordingly.

The root $0 \in V$ represents the source of the probes and the set of *leaf* nodes $R \subset V$ (i.e., those with no children) represents the receivers.

A probe packet is sent down the tree starting at the root. If it reaches a node j a copy of the packet is produced and sent down the link toward each child of j. As a packet traverses a link k (recall k denotes the endpoint), it is lost with probability $\bar{\alpha_k} = 1 - \alpha_k$ and arrives at k with probability α_k. We shall use the notation $\bar{\alpha} = 1 - \alpha$ for any quantity α (with or without subscripts) between 0 and 1. The losses on different links are assumed to be independent and to occur with the probabilities $\bar{\alpha_k}$ as described. Later we discuss how realistic this model is and how it can be corrected if there are dependencies between the losses.

We describe the passage of probes down the tree by a stochastic process $X = (X_k)_{k \in V}$ where each X_k equals 0 or 1: $X_k = 1$ signifies that a probe packet reaches node k, and 0 that it does not. The packets are generated at the source, so $X_0 = 1$. For all other $k \in V$, the value of X_k is determined as follows. If $X_k = 0$ then $X_j = 0$ for the children j of k (and hence for all descendants of k). If $X_k = 1$, then for j a child of k, $X_j = 1$ with probability α_j, and $X_j = 0$ with probability $\overline{\alpha}_j$, independently for all the children of k. We write $\alpha_0 = 1$ to simplify expressions concerning the α_k.

3 Results on Maximum Likelihood Estimators

If a probe is sent down the tree from the source, the outcome is a record of whether or not a copy of the probe was received at each receiver. Expressed in terms of the process X, the outcome is a configuration $X_{(R)} = (X_k)_{k \in R}$ of zeroes and ones at the receivers (1 = received, 0 = lost). Notice that only the values of X at the receivers are observable; the values at the internal nodes are invisible. The state space of the observations $X_{(R)}$ is thus the set of all such configurations, $\Omega = \{0, 1\}^R$. For a given set of link probabilities $\alpha = (\alpha_k)_{k \in V}$, the distribution of $X_{(R)}$ on Ω will be denoted by $\mathbf{P}_\alpha$. The probability mass function for a single outcome $x \in \Omega$ is $p(x; \alpha) = \mathbf{P}_\alpha(X_{(R)} = x)$.

Let us dispatch n probes, and, for each $x \in \Omega$, let $n(x)$ denote the number of probes for which the outcome x is obtained. The probability of n independent observations $x^1, \ldots, x^n$ (with each $x^m = (x_k^m)_{k \in R}$) is then

$$p(x^1, \ldots, x^n; \alpha) = \prod_{m=1}^{n} p(x^m; \alpha) = \prod_{x \in \Omega} p(x; \alpha)^{n(x)} \tag{1}$$

We estimate α using maximum likelihood based on the data $(n(x))_{x \in \Omega}$, and we find that the usual regularity conditions that imply good large-sample behavior of the MLE are satisfied in the present situation. This will be useful for the applications we have in mind because (a) we will want to assess the accuracy of our estimates via confidence intervals, and (b) it will be important to determine the smallest number n of probes needed to achieve the desired accuracy. We want to minimize n because, although sending out probes is inexpensive in itself, networks are subject to various fluctuations (e.g., [7]) which can perturb the model, and the measurement process itself ties up network resources.

We begin with a summary of our main results on the existence and uniqueness of the MLE. Another question, which we shall not treat here, but which is important for applications, is the feasibility and organization of the computations. We work with the log-likelihood function

$$\mathcal{L}(\alpha) = \log p(x^1, \ldots, x^n; \alpha) = \sum_{x \in \Omega} n(x) \log p(x; \alpha). \tag{2}$$

In the notation we suppress the dependence of $\mathcal{L}$ on n and $x^1, \ldots, x^n$. For each node k, let $\Omega(k)$ be the set of outcomes $x \in \Omega$ such that $x_j = 1$ for at least one receiver $j \in R$ which is a descendant of k, and let $\gamma_k = \gamma_k(\alpha) = \mathbf{P}_\alpha[\Omega(k)]$. An estimate of γ_k is

$$\widehat{\gamma}_k = \sum_{x \in \Omega(k)} \widehat{p}(x), \tag{3}$$

where $\widehat{p}(x) := n(x)/n$ is the observed proportion of trials with outcome x. We will show that α can be calculated from $\gamma = (\gamma_k)_{k \in V}$, and that the MLE

$$\breve{\alpha} = \arg\max_{\alpha \in [0,1]^{\#L}} \mathcal{L}(\alpha) \tag{4}$$

can be calculated in the same manner from the estimates $\widehat{\gamma}$. The relation between α and γ is as follows. We use U to denote the set of nodes other than the root.

Theorem 1 *Let $\mathcal{A} = \{(\alpha_k)_{k \in U} : \alpha_k > 0\}$, and $\mathcal{G} = \{(\gamma_k)_{k \in U} : \gamma_k > 0 \, \forall k; \, \gamma_k < \sum_{j \in d(k)} \gamma_j \, \forall k \in U \setminus R\}$. There is a bijection Γ from $\mathcal{A}$ to $\mathcal{G}$ which is differentiable and has a differentiable inverse.*

Candidates for the MLE are solutions of the *likelihood equation*:

$$\frac{\partial \mathcal{L}}{\partial \alpha_k}(\alpha) = 0, \quad k \in U. \tag{5}$$

Theorem 2 *When $\widehat{\gamma} \in \mathcal{G}$, the likelihood equation has the unique solution $\widehat{\alpha} := \Gamma^{-1}(\widehat{\gamma})$.*

The proof depends on a detailed analysis of $\mathcal{L}$ on the sets $\Omega(k)$.

We complete the picture by showing that the solution of the likelihood equation actually maximizes the likelihood function under some additional conditions. The set $\mathcal{A}$ contains all positive α_k, including the possibility $\alpha_k > 1$. Let us now restrict our attention to link probabilities $\alpha \in \mathcal{B} = (0,1)^{\#R} \subset \mathcal{A}$. Being a solution of the likelihood equation does not preclude $\widehat{\alpha}$ from being either a minimum or a saddlepoint for the likelihood function, with the maximum falling on the boundary of $\mathcal{B}$. For some simple topologies we are able to establish directly that $\mathcal{L}(\alpha)$ is (jointly) concave in the parameters at $\alpha = \widehat{\alpha}$, which is hence the MLE $\breve{\alpha}$. For more general topologies we use general results on maximum likelihood to show that $\widehat{\alpha} = \breve{\alpha}$ for all sufficiently large n.

Theorem 3

(i) The model is identifiable in $\mathcal{B}$, i.e., $\alpha, \alpha' \in \mathcal{B}$ and $\mathbf{P}_\alpha = \mathbf{P}_{\alpha'}$ implies $\alpha = \alpha'$. Thus, distinct link probabilities α produce distinct statistical behavior of the $\widehat{\gamma}$ as $n \to \infty$.

(ii) As $n \to \infty$, $\breve{\alpha} \to \alpha$, with $\mathbf{P}_\alpha$- probability 1, i.e., the MLE is strongly consistent.

(iii) With probability 1, for sufficiently large n, $\breve{\alpha} = \widehat{\alpha}$, i.e., the solution of the likelihood equation maximizes the likelihood.

This is proven using large sample theory for MLE, such as in [8]. Finally we have a result on asymptotic normality of the MLE. The *Fisher Information Matrix* at α based on $X_{(R)}$ is the matrix $\mathcal{I}_{jk}(\alpha) := \mathrm{Cov}\left(\frac{\partial \mathcal{L}}{\partial \alpha_j}(\alpha), \frac{\partial \mathcal{L}}{\partial \alpha_j}(\alpha)\right)$.

Theorem 4 *When $\mathcal{I}(\alpha)$ is non-singular, then as $n \to \infty$, under $\mathbf{P}_\alpha$, $\sqrt{n}(\widehat{\alpha} - \alpha)$ converges in distribution to a multivariate normal random vector with mean vector 0 and covariance matrix $\mathcal{I}^{-1}(\alpha)$.*

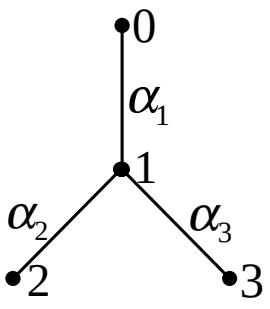

Figure 1: A two-leaf logical multicast tree

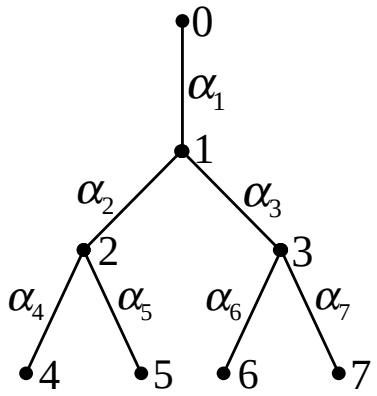

Figure 2: A four-leaf logical multicast tree

This also involves verifying the standard conditions for asymptotic normality. We have shown that the information matrix is nonsingular in several special cases, and naturally conjecture that it is always so.

As an example, we exhibit the MLEs for the tree, consisting of a root node ($k = 0$), its one child ($k = 1$), and two grandchildren ($k = 2, 3$) corresponding to two receivers; see Figure 1 The state space is now written as $\Omega = \{00, 01, 10, 11\}$. Then

$$\widehat{\gamma}_1 = \widehat{p}(11) + \widehat{p}(10) + \widehat{p}(01), \quad \widehat{\gamma}_2 = \widehat{p}(11) + \widehat{p}(10), \quad \widehat{\gamma}_3 = \widehat{p}(11) + \widehat{p}(01), \tag{6}$$

and

$$\widehat{\alpha}_1 = \frac{\widehat{\gamma}_2\widehat{\gamma}_3}{\widehat{\gamma}_2 + \widehat{\gamma}_3 - \widehat{\gamma}_1} = \frac{(\widehat{p}(01) + \widehat{p}(11))(\widehat{p}(10) + \widehat{p}(11))}{\widehat{p}(11)} \tag{7}$$

$$\widehat{\alpha}_2 = \frac{\widehat{\gamma}_2 + \widehat{\gamma}_3 - \widehat{\gamma}_1}{\widehat{\gamma}_3} = \frac{\widehat{p}(11)}{\widehat{p}(01) + \widehat{p}(11)} \tag{8}$$

$$\widehat{\alpha}_3 = \frac{\widehat{\gamma}_2 + \widehat{\gamma}_3 - \widehat{\gamma}_1}{\widehat{\gamma}_2} = \frac{\widehat{p}(11)}{\widehat{p}(10) + \widehat{p}(11)} \tag{9}$$

Note that $\widehat{\alpha}_1$ can be greater than 1 for a fixed n, but not as n tends to infinity.

4 Simulation Results

We illustrate our results through simulations on two types of examples. The first is simply the basic model of section 2, and these show that the MLEs behave as expected. The second type consists of simulations in which losses of probes are caused by queue overflows as probe traffic competes with other traffic generated by sources that use TCP (Transmission Control Protocol), which is the dominant transport protocol in the Internet [11]. Here the model assumptions are not satisfied and a bias is introduced into the MLEs. This is discussed further in the next section.

4.1 Basic Model Simulations

We simulated a simple binary trees with two and four receivers, under the basic model assumptions. In the 4 receiver tree, each node, other than the root, has exactly two children, and the leaves are the great-grandchildren of the root; see Figure 2. In all the simulation runs, the estimated link probabilities converged to within 0.01 of the actual probabilities within 2,000 observations; see Figure 3(Left). Note that a stream of two thousand 200-byte packets, one every 20 ms., moves at a rate of 10 Kb/s and takes 40 seconds to transmit, the equivalent of a single compressed audio transfer. There already exist a number of MBone "radio" stations that send long-lived streams of sequenced multicast packets. In some cases these could be used as measurement probes without additional cost.

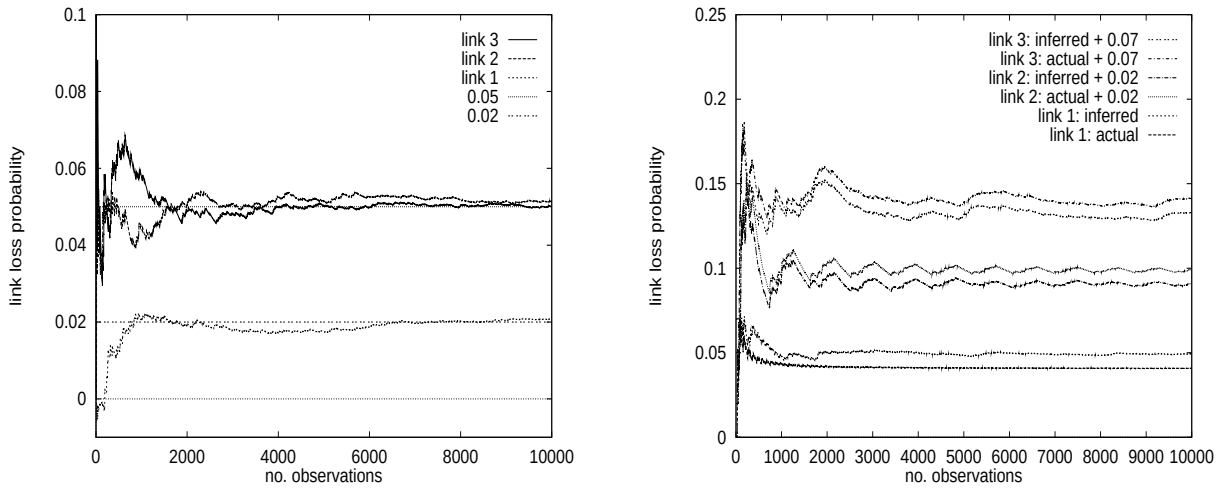

Figure 3: CONVERGENCE OF INFERRED LOSS RATES TO ACTUAL LOSS RATES IN SIMULATIONS. For 2-leaf tree of Figure 1. LEFT: model simulations from Section 4.1; RIGHT: TCP simulations of Section 4.2. Pairs of inferred and actual loss rates for the 3 links, some pairs offset for clarity. The inferred loss rates closely track the actual loss rates over 10,000 observations.

4.2 TCP Simulations

For the TCP simulations we used the ns network simulator [6] on 4 receiver tree. All links had 1.5 Mb/s bandwidth, 10 ms of propagation delay, and were served by a FIFO queue with a 4-packet limit. Thus, a packet arriving at a link was dropped when it found four packets already queued at the link.

Each node maintained TCP connections to its children, sending 1,000-byte packets driven by an infinite data source. Links to left children carried one such TCP stream, while links to right children carried two TCP streams. The link between the root 0 and its only child also carried one TCP stream. Probe packets were generated by a Constant Bit Rate (CBR) source with 200-byte packets and interpacket times chosen randomly between 2.5 and 7.5 msec.

The actual loss rates on individual links were compared with rates estimated using the basic model MLEs, the purpose being to stress the model. Although the estimated rates tracked the actual loss rates over a span of 10,000 observations, there was an asymptotic bias of between 4 and 18% estimated on the basis of 100 simulation runs; see Figure 3(right). This is due to spatial dependence of the losses: the same probe is lost on sibling links more often than the independence assumption dictates. This causes the basic model MLE to underestimate losses on the child links and to overestimate them on the parent link.

It is possible to reduce the bias in the presence of such spatial dependence. Using the method discussed in the next section, it was reduced to about 1%. We believe, however, that strong spatial dependence is unlikely in real networks like the Internet because of their traffic and link diversity. Elsewhere we consider the question of temporal dependence.

5 Spatial Dependence

If we expand the basic model slightly, we can allow for dependence among the losses on sibling links and can correct for the bias in the MLEs of section 3 if we have a prior estimate of the correlation between losses on sibling links. In this expanded model, it can be shown that the MLEs are continuous functions of a parameter that governs the degree of spatial dependence.

For simplicity we consider only the 2 receiver tree. We introduce an additional parameter ν and probabilities on Ω as follows:

$$p(11;\alpha,\nu) = \alpha_1(\alpha_2\alpha_3 + \nu\alpha_2\alpha_3) \tag{10}$$

$$p(10;\alpha,\nu) = \alpha_1(\alpha_2\overline{\alpha}_3 - \nu\alpha_2\alpha_3) \tag{11}$$

$$p(01;\alpha,\nu) = \alpha_1(\overline{\alpha}_2\alpha_3 - \nu\alpha_2\alpha_3) \tag{12}$$

$$p(00;\alpha,\nu) = \overline{\alpha}_1 + \alpha_1(\overline{\alpha}_2\overline{\alpha}_3 + \nu\alpha_2\alpha_3) \tag{13}$$

When $\nu > 0$, the correlation between the losses at the leaves is positive.

Now consider sending n probes through a network with losses described by these probabilities. As $n \to \infty$, the proportion $\widehat{p}(x)$ converges to $p(x;\alpha,\nu)$. To see how the estimator $\widehat{\alpha}$ interprets this, we insert these limit values in place of $\widehat{p}(x;\alpha)$ in eqs. (7)–(9). The resulting estimates are

$$\alpha_1(\nu) = \frac{\alpha_1}{1+\nu}, \quad \alpha_2(\nu) = (1+\nu)\alpha_2, \quad \alpha_3(\nu) = (1+\nu)\alpha_3. \tag{14}$$

For $\nu \to 0$ we obtain the original estimates.

If prior knowledge of the loss correlations is available, we can adjust the estimates in the following way. Suppose that κ, the conditional correlation between X_2 and X_3, given that $X_1 = 1$ is known. A calculation shows that

$$\nu = \kappa\sqrt{\frac{\overline{\alpha}_2\overline{\alpha}_3}{\alpha_2\alpha_3}}. \tag{15}$$

The MLEs from section 3 will yield estimates $\widehat{\alpha}(\widehat{\nu})$, where $\widehat{\nu} = \kappa\sqrt{(1-\widehat{\alpha}_2)(1-\widehat{\alpha}_3)/\widehat{\alpha}_2\widehat{\alpha}_3}$, following (15). Inserting $\widehat{\alpha}(\widehat{\nu})$ and $\widehat{\nu}$ into (14) in place of $\alpha(\nu)$ and ν yields an estimate of the true link probabilities, the α of (14). As mentioned above, this method reduced the bias in our simulations from between 8 and 15% to about 1%. We intend to undertake experiments on real networks to ascertain the magnitude of these correlations.

6 Discussion

We have introduced the use of end-to-end measurements of multicast traffic to infer internal link loss probabilities, and have shown that maximum likelihood estimation is feasible when the losses are independent. We also presented evidence that our techniques yield accurate results even in the presence of moderate spatial dependence. A corresponding discussion, not included here, pertains to temporal dependence.

We are extending our work in several directions. First, we are applying multicast-based inference to other network characteristics. For instance, we have developed estimators for link delays, and are investigating inference on link bandwidth and network topology using multicast probes.

Further, we plan to experiment with multicast-based inference on the Internet. We also plan to deploy our inference tools in multicast-enabled portions of the Internet, including the MBone, to test our techniques on a real network, and eventually integrate them with one of the large-scale measurement infrastructures under construction.

References

[1] Felix: Independent Monitoring for Network Survivability. For more information see ftp://ftp.bellcore.com/pub/mwg/felix/index.html

[2] IPMA: Internet Performance Measurement and Analysis. For more information see http://www.merit.edu/ipma

[3] V. Jacobson, Pathchar - A Tool to Infer Characteristics of Internet paths. For more information see ftp://ftp.ee.lbl.gov/pathchar

[4] J. Mahdavi, V. Paxson, A. Adams, M. Mathis, "Creating a Scalable Architecture for Internet Measurement," *to appear in Proc. INET '98*.

[5] mtrace – Print multicast path from a source to a receiver. For more information see ftp://ftp.parc.xerox.com/pub/net-research/ipmulti

[6] ns – Network Simulator. For more information see http://www-mash.cs.berkeley.edu/ns/ns.html

[7] V. Paxson, "End-to-End Routing Behavior in the Internet," *Proc. SIGCOMM '96*, Stanford, Aug. 1996.

[8] M.J. Schervish, "Theory of Statistics", Springer, New York, 1995.

[9] J. Postel, "Transmission Control Protocol," RFC 793, September 1981.

[10] Surveyor. For more information see http://io.advanced.org/surveyor/

[11] K. Thompson, G.J. Miller and R. Wilder, "Wide-Area Internet Traffic Patterns and Characteristics," *IEEE Network*, 11(6), November/December 1997.

[12] R.J. Vanderbei and J. Iannone, "An EM approach to OD matrix estimation," Technical Report, Princeton University, 1994

[13] Y. Vardi, "Network Tomography: estimating source-destination traffic intensities from link data," J. Am. Statist. Assoc., 91: 365–377, 1996.