Traffic Matrix Estimation on a Large IP Backbone – A Comparison on Real Data

Anders Gunnar
Swedish Institute of Computer Science
P.O. Box 1263
SE-164 29 Kista, Sweden
anders.gunnar@sics.se

Mikael Johansson
Department of Signals, Sensors and Systems, KTH
SE-100 44 Stockholm, Sweden
mikaelj@s3.kth.se

Thomas Telkamp
Global Crossing, Ltd
Croeselaan 148
NL-3521 CG Utrecht, The Netherlands
telkamp@gblx.net

ABSTRACT

This paper considers the problem of estimating the point-to-point traffic matrix in an operational IP backbone. Contrary to previous studies, that have used a partial traffic matrix or demands estimated from aggregated Netflow traces, we use a unique data set of complete traffic matrices from a global IP network measured over five-minute intervals. This allows us to do an accurate data analysis on the time-scale of typical link-load measurements and enables us to make a balanced evaluation of different traffic matrix estimation techniques. We describe the data collection infrastructure, present spatial and temporal demand distributions, investigate the stability of fan-out factors, and analyze the mean-variance relationships between demands. We perform a critical evaluation of existing and novel methods for traffic matrix estimation, including recursive fanout estimation, worst-case bounds, regularized estimation techniques, and methods that rely on mean-variance relationships. We discuss the weaknesses and strengths of the various methods, and highlight differences in the results for the European and American subnetworks.

Categories and Subject Descriptors

C.2.3 [**Computer Communications Networks**]: Network Operations—*Network Management, Network Monitoring*

General Terms

Measurement, Performance

Keywords

Traffic matrix estimation, Optimization, SNMP, MPLS

IMC'04 October 25–27, 2004, Taormina, Sicily, Italy.

1. INTRODUCTION

Many of the decisions that IP network operators make depend on how the traffic flows in their network. A *traffic matrix* describes the amount of data traffic transmitted between every pair of ingress and egress points in a network. When used together with routing information, the traffic matrix gives the network operator valuable information about the current network state and is instrumental in traffic engineering, network management and provisioning (see, e.g., [1, 2, 3, 4]).

Despite the importance of knowing the traffic matrix, the support in routers for measuring traffic matrices is poor and operators are often forced to estimate the traffic matrix from other available data, typically link load measurements and routing configurations. In its simplest form, the estimation problem then reduces to finding a non-negative vector s that satisfies $Rs = t$, where R is a matrix reflecting the routing, t is a vector of measured link loads and s is a vectorized version of the (unknown) traffic matrix. The link loads are readily obtained using the Simple Network Management Protocol (SNMP). This approach leads to an under-constrained problem since the number of links in a network is typically much smaller than the number of node pairs. Some sort of side information or assumptions must then be added to make the estimation problem well-posed.

To evaluate how well different approaches to traffic matrix estimation will work in an operational IP network, and how reasonable various assumptions are, one needs access to a measured traffic matrix on the time-scale of standard link-load measurements. Previous studies have used NetFlow data to measure the traffic matrix in 5-minute increments on a single router [5] or one-hour traffic matrices on a partial network [6]. However, since NetFlow data is unable to capture traffic variability within flows, this is not very accurate for validating estimation methods that use a time-series of link-load measurements. Our study provides new results in the sense that it uses a complete network traffic matrix, based on direct measurements at 5-minute intervals. The data set is collected from Global Crossing's global backbone and consists of routing configuration and the number of bytes transfered in MPLS-tunnels during 5-minute intervals over a 24-hour period.

To make the analysis more transparent, we extract traffic matrices and routing information for the American and European subnetworks. We present temporal and spatial demand distributions and analyze some statistical properties

of the demands. In particular, we find that there is a surprisingly strong relationship between the mean and variance of demands, and that fanout factors tend to be relatively more stable over time compared to the demand themselves. We then evaluate a selection of existing methods for traffic matrix estimation, including gravity models, regularized methods (such as Bayesian and maximum entropy approaches), and methods that exploit mean-variance relationships. In addition, we investigate the use of worst-case bounds and estimation of fanout factors based on a time-series of link load measurements. We find that the regularized methods work very well provided that we choose the regularization parameter, *i.e.*, the tradeoff between prior information and link measurement, appropriately. Somewhat surprisingly, we fail to achieve good results using methods that exploit mean-variance relationship. We argue that the failure stems from the problem of accurately estimating the covariance matrix of link loads, and present a study on synthetic data to support our claim.

One can note that many classes of traffic matrices occur in the literature (see [7] for a thorough classification). In this paper, we only study the performance of the estimation methods on PoP-to-PoP traffic matrices. This choice is solely based on properties of the data we have obtained, and we make no statement on which class of traffic matrices is more important than the other.

The remaining parts of this paper are organized as follows. In Section 2, we present related work in this area. Section 3 introduces the problem and notation. The estimation methods that we evaluate are introduced in Section 4, while data collection, data analysis and benchmarking of the methods is presented in Section 5. Finally, some concluding remarks are collected in Section 6.

2. RELATED WORK

The origin-destination estimation problem for telephone traffic is a well-studied problem in the telecom world. For instance, already in 1937, Kruithof [8] suggested a method for estimation of point-to-point traffic demands in a telephone network based on a prior traffic matrix and measurements of incoming and outgoing traffic. However, it appears that it was not until 1996 that the problem was addressed specifically for IP networks. In order to handle the difficulties of an under-constrained problem, Vardi [9] assumes a Poisson model for the traffic demands and covariances of the link loads is used as additional constraints. The traffic demands are estimated by Maximum Likelihood estimation. Related to Vardi's approach is Cao *et al.* [5] that propose to use a more general scaling law between means and variances of demands. The Poisson model is also used by Tebaldi and West [10], but rather than using ML estimation, they use a Bayesian approach. Since posterior distributions are hard to calculate, the authors use a Markov Chain Monte Carlo simulation to simulate the posterior distribution. The Bayesian approach is refined by Vaton *et al.* [11], who propose an iterative method to improve the prior distribution of the traffic matrix elements. The estimated traffic matrix from one measurement of link loads is used in the next estimation using new measurements of link loads. The process is repeated until no significant change is made in the estimated traffic matrix. An evaluation of the methods in [10, 9] together with a linear programming model is performed by Medina *et al.* [12]. A novel approach based on choice models is also suggested in the article. The choice model tries to estimate the probability of an origin node to send a packet to a destination node in the network. Similar to the choice model is the gravity model introduced by Zhang *et al.* [6]. In its simplest form the gravity model assumes a proportionality relation between the traffic entering the network at node i and destined to node j and the total amount of traffic entering at node i and the total amount of traffic leaving the network at node j. The authors of the paper use additional information about the structure and configuration of the network such as peering agreements and customer agreements to improve performance of the method. An information-theoretic approach is used by Zhang *et al.* in [13] to estimate the traffic demands. Here, the Kullback-Leibler distance is used to minimize the mutual information between source and destination. In all papers mentioned above, the routing is considered to be constant. In a paper by Nucci *et al.* [14] the routing is changed and shifting of link load is used to infer the traffic demands. Feldmann *et al.* [15] uses a somewhat different approach to calculate the traffic demands. Instead of estimating from link counts they collect flow measurements from routers using Cisco's NetFlow tool and derive point-to-multipoint traffic demands using routing information from inter- and intra-domain routing protocols.

3. PRELIMINARIES

3.1 Notation and Problem Statement

We consider a network with N nodes and L directed links. Such a network has $P = N(N-1)$ pair of distinct nodes that may communicate with each other. The aggregate communication rate (in bits/second) between any pair (n, m) of nodes is called the *point-to-point demand* between the nodes, and we will use s_{nm} to denote the rate of the aggregate data traffic that enters the network at node n and exits the network at node m. The matrix $S = [s_{nm}]$ is called the *traffic matrix*. It is usually more convenient to represent the traffic matrix in vector form. We then enumerate all P source-destination pairs, and let s_p denote the point-to-point demand of node pair p.

For simplicity, we will assume that each point-to-point demand is routed on a single path. The paths are represented by a *routing matrix* $R \in \mathbb{R}^{L \times P}$ whose entries r_{lp} are defined as

$$r_{lp} = \begin{cases} 1 & \text{if the demand of node pair } p \text{ is routed across link } l \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Note that the routing matrix may easily be transformed to reflect a situation where traffic demands are routed on more than one path from source to destination by allowing fractional values in the routing matrix. Let t_l denote the aggregate data rate on link l, $t = [t_l] \in \mathbb{R}^L$ be the vector of link rates, and $s \in \mathbb{R}^P$ be the vector of demands for all source-destination pairs. Then, s and t are related via

$$Rs = t \tag{2}$$

The *traffic matrix estimation problem* is simply the one of estimating the non-negative vector s based on knowledge of R and t.

The challenge in this problem comes from the fact that this system of equations tends to be highly underdetermined: there are typically many more source-destination pairs than links in a network, and (2) has many more unknowns than equations.

It is important to note that in an IP network setting, not all links are interior links connecting the core routers in the network: some of the links are access and peering links that supply data to and receive data from the edge nodes. To make this more explicit, we introduce the notation $e(n)$ for the link over which demand enters at node n, and $x(m)$ for the link over which demand exits at node m. For ease of notation, we assume that each edge node is either an access or a peering point (if this is not the case we can always introduce artificial nodes in our network representation so that this holds). Under these assumptions, $t_{e(n)}$ is the total traffic entering the network at node n and $t_{x(m)}$ is the total traffic exiting the network at node m. Finally, we let $\mathcal{A}$ be the set of nodes acting as access points, and $\mathcal{P}$ the set of nodes acting as peering points.

3.2 Alternative formulations of traffic estimation problems

3.2.1 The traffic matrix as a demand distribution

Since demands are non-negative, it is natural to normalize s with the total network traffic

$$s_{\text{tot}} = \sum_i \sum_j s_{ij} = \sum_n t_{e(n)}$$

and view $\tilde{s} = s/s_{\text{tot}}$ as a probability distribution. We may then interpret $\tilde{s}_p$ as the probability that a random packet in the network is sent between node pair p. Introducing $\tilde{t} = t/s_{\text{tot}}$, we can re-write (2) as

$$\begin{cases} R\tilde{s} = \tilde{t} \\ \mathbf{1}^T \tilde{s} = 1, \quad \tilde{s} \succeq 0 \end{cases} \tag{3}$$

The traffic estimation problem then becomes the one of estimating a vector $\tilde{s}$ that satisfies (3) based on knowledge of R and $\tilde{t}$ (cf. [8, 16]).

3.2.2 Fanout formulations

Another alternative is to normalize the demands by the total aggregate traffic entering the source node, i.e., to write

$$s_{nm} = \alpha_{nm} \sum_m s_{nm} = \alpha_{nm} t_{e(n)} \qquad \sum_m \alpha_{nm} = 1 \tag{4}$$

Rather than estimating s_{nm}, one can now focus on estimating the *fanouts* $\alpha_{nm} = s_{nm}/t_{e(n)}$. Also the fanouts can be interpreted as probability distributions: α_{nm} is the probability that a random packet entering the network at node n will exit the network at node m (cf. [12, 17]).

4. METHODS FOR TRAFFIC MATRIX ESTIMATION

4.1 Gravity Models

A simple method for estimating the traffic matrix is to use a so-called *gravity model*. Although these models have a long history in the social sciences [18] and in telephony networks [19], the first application to demand estimation in IP networks appears to be [20]. In our notation, the basic version of the gravity model predicts the demand between node n and node m as

$$s_{nm}^{(p)} = C t_{e(n)} t_{x(m)} \tag{5}$$

where C is a normalization constant that makes the sum of estimated demands equal to the measured total network traffic. With the choice $C = 1/\sum_m t_{x(m)}$, the gravity model reduces to

$$s_{nm}^{(p)} = \frac{t_{x(m)}}{\sum_m t_{x(m)}} t_{e(n)}$$

and a comparison with (4) reveals that this is equivalent to the fanout model

$$\alpha_{nm} = \frac{t_{x(m)}}{\sum_m t_{x(m)}}$$

i.e., that the amount of data that node n sends to node m is proportional to the fraction of the total network traffic that exits at node m. Such a model makes sense if the user populations served by different nodes are relatively uniform. However, as pointed out in [6], traffic transit between peering networks behaves very differently. This has led to the generalized gravity model, where traffic between peers is forced to be zero, i.e.,

$$s_{nm}^{(p)} = \begin{cases} 0 & \text{if } n \in \mathcal{P} \text{ and } m \in \mathcal{P} \\ C t_{e(n)} t_{x(m)} & \text{otherwise} \end{cases}$$

Once again, C is a normalization constant that makes, for example, the estimated total traffic equal to the measured total network traffic. In this study, however, we focus on the simple gravity model and leave the generalized gravity model without further reference. It should be noted that the gravity model does not use any information about the traffic on links interior to the network, and that the estimates are typically not consistent with the link load measurements (in fact, the model may not even produce consistent estimates of the total traffic exiting each node). Thus, gravity models are often not used in isolation, but in combination with some statistical approach that accounts for measured link loads. Such methods will be described next.

4.2 Statistical Approaches

4.2.1 Kruithof's Projection Method

One of the oldest methods for estimating traffic matrices is the iterative method due to Kruithof [8]. The original formulation considers the problem of estimating point-to-point traffic in a telephony network based on a known prior traffic matrix and measurements of total incoming and total outgoing traffic to each node in the network. Thus, Kruithof's method can, for example, be used to adjust the gravity model estimate to be consistent with measurement of total incoming and outgoing traffic at edge nodes.

Kruithof's method was first analyzed by Krupp [16], who showed that that the approach can be interpreted from an information theoretic point-of-view: it minimizes the Kullback-Leibler distance from the prior traffic matrix $\left[s_{ij}^{(p)}\right]$ (interpreted as a demand distribution). Krupp also extended Kruithof's basic method to general linear constraints,

$$\begin{array}{ll} \text{minimize} & D(s\|s^{(p)}) \\ \text{subject to} & Rs = t, \qquad s \succeq 0 \end{array}$$

and showed that the extended iterative method convergences to the unique optimal solution. It is interesting to note that Kruithof's method appears to be the first iterative scaling method in statistics, and that these methods are closely related to the celebrated EM-algorithm [21].

Recently [13], Zhang *et al.* have suggested to use the related criterion

$$\begin{array}{ll} \text{minimize} & \|Rs - t\|_2^2 + \sigma^{-2} D(s\|s^{(p)}) \\ \text{subject to} & s \succeq 0 \end{array} \tag{6}$$

for estimating traffic matrices for backbone IP traffic. The practical advantage of this formulation, which we will refer to as the Entropy approach, is that the optimization problem admits a solution even if the system of linear constraints is inconsistent. We will comment on possible choices of prior matrices at the end of this section.

4.2.2 Estimation under Poissonian and Generalized Linear Modeling Assumptions

Vardi [9] suggested to use a Poissonian model for the traffic, i.e., to assume that

$$s_p \sim \text{Poisson}(\lambda_p)$$

and showed that the mean and covariance matrix of the link loads are given by

$$\mathbf{E}\{t\} = R\lambda \qquad \mathbf{Cov}\{t\} = R\,\text{diag}(\lambda)\,R^T$$

A key observation is that the Poissonian model provides an explicit link between the mean and covariance matrix of the traffic. Based on a time-series of link load measurements, we compute the sample mean and covariance,

$$\hat{t} = \frac{1}{K}\sum_{k=1}^{K} t[k] \qquad \hat{\Sigma} = \frac{1}{K}\sum_{k=1}^{K}(t[k]-\hat{t})(t[k]-\hat{t})^T$$

and then match the measured moments with the theoretical, *i.e.*, solve

$$R\lambda = \hat{t} \qquad R\,\text{diag}(\lambda)\,R^T = \hat{\Sigma}$$

for the vector λ of mean traffic rates. By accounting for the model of the covariance matrix we get $L(L+1)/2$ additional relations, and Vardi proves that the combined information makes the vector λ statistically identifiable. In practice, however, there will typically be no vector λ that attains equality in the moment matching conditions (this may for example be due to lack of data, outliers, or violated modeling assumptions). Vardi suggests to use the EM algorithm to minimize the Kullback-Leibler distance between the observed sample moments and their theoretical values. However, as pointed out in [22], when the observed values are not guaranteed to be non-negative, it is more reasonable to use a least squares fit. To this end, we find the estimate λ by solving the non-negative least-squares problem

$$\begin{array}{ll} \text{minimize} & \|R\lambda - \hat{t}\|_2^2 + \sigma^{-2}\|R\text{diag}(\lambda)R - \hat{\Sigma}\|_2^2 \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

The parameter $\sigma^{-2} \in [0,1]$ reflects our faith in the Poissonian modeling assumption (compare [9, Section 4]): if σ^{-2} tends to zero, then we base our estimate solely on the first moments, while $\sigma^{-2} = 1$ is natural if we believe in the Poisson assumption.

Cao *et al.* [5] have extended the Vardi's approach by considering a generalized linear modeling assumption

$$s_p \sim \mathcal{N}\left(\lambda_p, \phi\lambda_p^c\right)$$

and assumes that all source-destination flows are independent. The additional scaling parameters ϕ and c give somewhat more freedom than the strict Poissonian assumption. However, even for fixed scaling constants ϕ and c, the estimation procedure is more complex (the associated optimization problem is non-convex), and Cao *et al.* propose a pseudo-EM method for estimation under fixed value of c. An interesting aspect of the paper by Cao *et al.* is that they also try to account for time-variations in the OD flows in order to use more measurements than the 12 link count vectors logged during a busy hour.

4.2.3 Regularized and Bayesian Methods

A related class of methods can be motivated from Bayesian statistics [10]. For example, by modeling our prior knowledge of the traffic matrix as

$$s \sim \mathcal{N}(s^{(p)}, \sigma^2 I)$$

and assuming that the traffic measurements are subject to white noise with unit variance, i.e.

$$t = Rs + v$$

with $\mathbf{E}\{v\} = 0$, $\mathbf{Cov}\{v\} = I$, the maximum a posteriori (MAP) estimate is found by solving

$$\text{minimize} \quad \|Rs - t\|_2^2 + \sigma^{-2}\|s - s^{(p)}\|_2^2 \tag{7}$$

Once again, the optimal estimate can be computed by minimizing a weighted distance of the errors between theoretical and observed means and the distance between the estimated demands and a prior "guesstimate". The variance σ^2 in the prior model is typically used as a tuning parameter to weigh the relative importance that we should put on the two criteria. The formulation (7) has been used in, for example, [6], where the prior is computed using a gravity model.

4.2.4 Fanout Estimation

Although fanout estimation does not simplify the estimation problem if we only use a single snapshot of the link loads, it can be useful when we have a time-series of link load measurements. As discussed in Section 3.2.2, the fanout formulation of (2) is the one of finding a non-negative vector $\alpha[k]$ such as

$$RS[k]\alpha[k] = t[k], \qquad \sum_m \alpha_{nm}[k] = 1, \quad n = 1,\ldots,N$$

where $S[k]$ is a diagonal scaling matrix such that $s[k] = S[k]\alpha[k]$.

Given a time series of link load measurements, we may assume that the fanouts are constant (i.e., that all link load fluctuations are due to changes in the total traffic generated by each node) and try to find $\alpha \succeq 0$ satisfying

$$\begin{aligned} RS[k]\alpha &= t[k], & k &= 1,\ldots,K, \\ \sum_m \alpha_{nm} &= 1, & n &= 1,\ldots,N \end{aligned}$$

Even if the routing matrix itself does not have full rank, the above system of equations will quickly become overdetermined, and there is a unique vector α that minimizes the

errors (in a given norm) between the observed link counts and the ones predicted by the constant-fanout model. These can be found by solving the optimization problem

$$\begin{array}{ll} \text{minimize} & \sum_{k=1}^{K} \|RS[k]\alpha - t[k]\|_2^2 \\ \text{subject to} & \sum_{n=1}^{N} \alpha_{nm} = 1, \qquad m = 1, \ldots, N \end{array}$$

which is simply an equality-constrained quadratic programming problem.

4.3 Deterministic Approaches

4.3.1 Worst-case bounds on demands

In addition to statistical estimates, it is also interesting to find upper and lower bounds on the demands. Making no underlying statistical assumptions on the demands, we note that a single measurement $t[k]$ of the link loads could be generated by the set of possible communication rates,

$$\mathcal{S} = \{s \succeq 0 \mid Rs = t[k]\}$$

Thus, an upper bound on demand p can be computed by solving the linear programming problem

$$\begin{array}{ll} \text{maximize} & s_p \\ \text{subject to} & Rs = t[k], \qquad s \succeq 0 \end{array}$$

The associated lower bound is found by minimizing s_p subject to the constraints. Obviously, this approach is only interesting when it finds an upper bound smaller than the trivial $\max_{l \in \mathcal{L}(p)} t_l[k]$ and a lower bound greater than zero. Also note that the method is computationally expensive, as it requires solving two linear programs for each point-to-point demand.

5. BENCHMARKING THE METHODS ON REAL DATA

A major contribution of this paper is to study the traffic in the backbone of a commercial Internet operator, and to benchmark the existing traffic matrix estimation methods on this data. A complete traffic matrix is measured using the operator's MPLS-enabled network.

Previous work also validated estimation methods on real data, but they instead used NetFlow data to measure the traffic matrix on single router or on a partial network. [6] validates the tomogravity method with NetFlow measurements of 2/3 of a tier-1 IP backbone, using hourly traffic matrices. In [5] NetFlow data from a single router is used to create traffic matrices in 5 minute increments, for validating time-varying network tomography.

NetFlow exports flow information from the routers to a collector system. The exported information contains the start and end time of every flow, and the number of bytes transmitted during that interval. The collector calculates the average rate during the lifetime of the flow, and adds that to the traffic matrix. For validating time-varying tomography, this is not a very accurate methodology. The variability within a flow is lost because of the NetFlow aggregation. This might affect the variance-mean relationship this method is based on.

Our study provides new results in the sense that it uses a *full* network traffic matrix, based on the *direct measurements* (rather than analysis of NetFlow traces) of all demands at *5 minute intervals*.

In the remaining parts of this section, we describe how a complete traffic matrix is measured using Global Crossing's MPLS-enabled network, investigate some basic properties of the demands, and evaluate the existing methods for traffic matrix estimation on the data.

5.1 Data Collection and Evaluation Data Set

5.1.1 Network

Global Crossing is using MPLS for Traffic Engineering on its global IP backbone. A mesh of Label Switched Paths (LSPs, a.k.a. "tunnels") has been established between all the core routers in the network. Every LSP has a bandwidth value associated with it, and the core router originating the LSP (head-end) will use a constraint based routing algorithm (CSPF) to find the shortest path that has the required bandwidth available. RSVP is then used to setup the actual path across the network. This architecture is described in detail in [23].

By measuring the utilization of every LSP in 5 minute intervals using SNMP, we can create a full and accurate traffic matrix of the network. This is an additional, but important, benefit of running an MPLS-enabled network.

5.1.2 Data Collection

To collect SNMP data from the network, a geographically distributed system of "pollers" has been set up. Each poller retrieves SNMP information from a dedicated set of routers in its area, and also functions as a backup for neighboring pollers. SNMP uses the unreliable UDP protocol for communications between the routers and monitoring systems, and hence there is the risk of losing data during transmission. A distributed system with the pollers located close to the routers being monitored increases the reliability in the case of network performance issues or outages, and keeps the load per poller manageable.

The link and LSP utilizations are collected every 5 minutes, at fixed timestamps (e.g. 9:00:00, 9:05:00, 9:10:00, etc.). There will be some variation in the exact polling time, as it is impossible to query every router and interface at exactly the same time. The exact response time of the routers is recorded, and the corresponding utilization rate data is adjusted for the length of the real measurement interval (e.g. 5 minutes and 3 seconds). The impact of this on the measurements is only minimal, and it provides uniform time series of link and LSP utilization data.

The pollers transfer their data to a central database at fixed intervals, using a reliable transport protocol (TCP).

5.1.3 Routing Matrix

The routing matrix in the form described by equations (1) and (2) is created using a simulation of the network. Although the routing of the LSPs in the network could be retrieved from the routers, it proves to be more practical to simulate the constraint based routing protocol (CSPF) as used by the routers, using the same constraints data (i.e. LSP bandwidth values).

We use the tool MATE from Cariden [24] to perform this routing simulation, and export this information in a text file. The data is then converted to a routing matrix according to equation (1).

Although the routing in the network is often in a state of flux because of link and/or equipment outages, this is not of much relevance to our study. These routing changes only have a minor effect on the point-to-point demands (i.e., the traffic matrix).

5.1.4 Evaluation Data Set

In order to perform a scientific evaluation of the estimation methods, we need the measurements of routing, traffic matrix elements and link loads to be consistent. By consistent we mean measurements which satisfies equation (2).

By using equation (2) we are able to compute the link loads needed as input to the estimation methods, from the measured point-to-point demands and the simulated routing matrix. The above mentioned procedure enable us to evaluate the accuracy of the methods on real data without the errors incurred by errors in the measurement of the link loads.

From Global Crossing's network, we have extracted routing information and traffic matrices for the European and American subnetworks. The reason for this is that we wanted to study networks of manageable size that still accommodate large traffic demands. It also allows us to study if there are any significant differences in the demand patterns on the two continents. To create these separate traffic and routing matrices, we simply exclude all links and demands that do not have both source and destination inside the specific region.

Further, core routers located in the same city were aggregated to form a point of presence (PoP), and we study the PoP-to-PoP traffic matrix. Many PoPs contain routers who only transit traffic. We have in this study included links between these transit routers since we focus on estimation in real networks where transit routers are present. Because not necessarily all the original demands between two PoPs were following the same path, we decided to route the aggregated demand according to the routing of the largest original demand. In practice though, most parallel demands already followed the same path.

Using this approach, the European network has 12 PoPs (thus 132 point-to-point demands) and 72 links, while the American network has 25 PoPs (600 demands) and 284 links.

Since the precise details of the traffic are considered proprietary, we scale all plots by the maximum value of the total traffic during the measurement period. It might, however, be interesting to know that the largest traffic demands are on the order of 1200 Mbps.

5.2 Preliminary Data Analysis

5.2.1 Busy hours and demand distributions

Figure 1 shows how the normalized total traffic in the two subnetworks vary with time. The solid and dashed lines represent the European and American networks, respectively. There is a clear diurnal cycle, and both subnetworks have a pronounced busy periods. The busy periods overlap partly around 18:00 GMT, and the time period shaded in Figure 1. We will focus our data analysis to this interval.

Figure 2 shows the cumulative demand distribution for the subnetworks. The figure shows that the top 20 percent of demands account for approximately 80 percent of the traffic in both networks.

A similar insight can be obtained from the spatial traffic distributions illustrated in Figure 3, where we see that a limited subset of nodes account for the majority of network traffic.

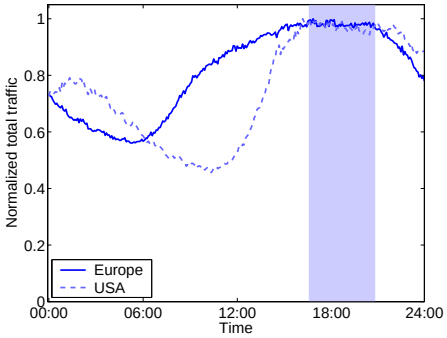

Figure 1: Total network traffic over time. The solid line represents the European network, while the dashed line represents the American subnetwork.

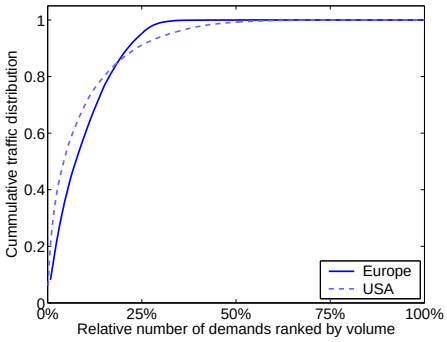

Figure 2: Cumulative demand distributions for the European network (solid) and the American subnetwork (dashed).

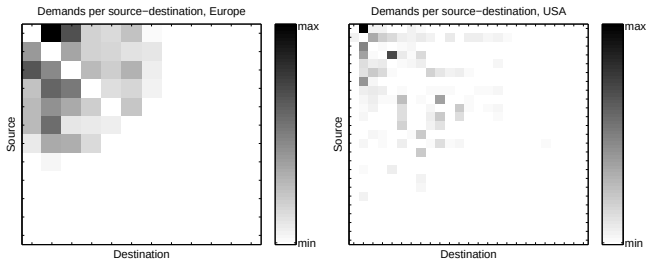

Figure 3: Spatial distribution of traffic in the two subnetworks.

5.2.2 On the stability of fanout factors

As we have seen in Section 3, there are several possible formulations of the traffic estimation problem: we may estimate the demands directly, focus on the relative demands (viewing the traffic matrix as a demand distribution) or the fanout factors. While the total network traffic changes with the number of active users, one may conjecture that fanouts

would be stable as long as the average user behavior does not change. In this section, we investigate whether fanouts are more stable over time than demands themselves. If this is the case, fanout estimation may be easier than demand estimation since we do not have to rely on data logged only during the stationary busy hour. Furthermore, if fanouts are stable, it is a worthwhile idea to develop models for fanout factors based on node characteristics (cf. [17]).

Figure 4 shows how the demands from the four largest PoPs in the American network fluctuate over the 24-hour measurement period, while Figure 5 shows the associated fanouts. We can see that the fanouts are much more stable than the demand themselves during this measurement period. The same qualitative relationship holds for all large demands in the network; for the smaller demands, however, the fanouts sometimes fluctuate more than the demands themselves.

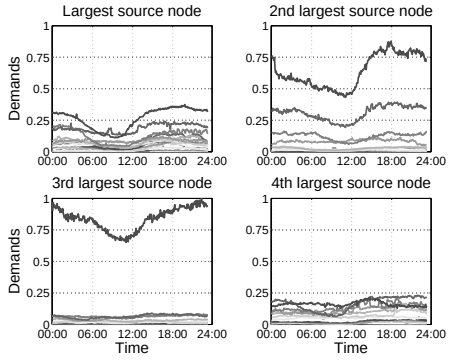

Figure 4: The four largest outgoing demands from the four largest PoPs in the American network.

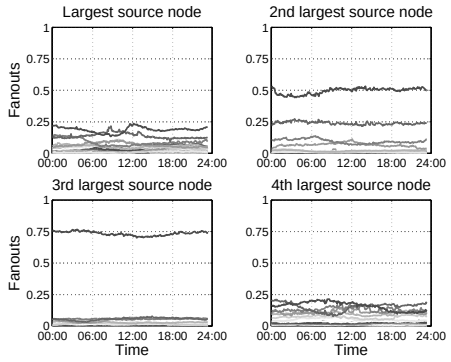

Figure 5: The associated fanouts for the four largest outgoing demands from the four largest PoPs in the American network.

5.2.3 On the Poissonian Modeling Assumption

The assumption that demands are Poissonian, or that they follow a generalized scaling law, provides an explicit link between mean and covariances of link load measurements. Such a link allows us, at least in theory, to statistically identify the demands based on a time series of link load measurements. It is therefore interesting to investigate how well our data satisfies the generalized scaling law [5]

$$\mathbf{Var}\{s_p\} = \phi\lambda_p^c$$

In particular, if the traffic is Poissonian, then $\phi = c = 1$. Figure 6 shows the relationship between the 5-minute averages of mean and variance for the demands in our subnetworks during busy hour. The plots show a remarkably strong relation between mean and variance and that the generalized scaling law is able to capture the mean-variance relationship for the demands in both subnetworks. The parameters $\phi = 0.82, c = 1.6$ gives the best fit for the European demands, and $\phi = 2.44, c = 1.5$ results in the best fit for the American network.

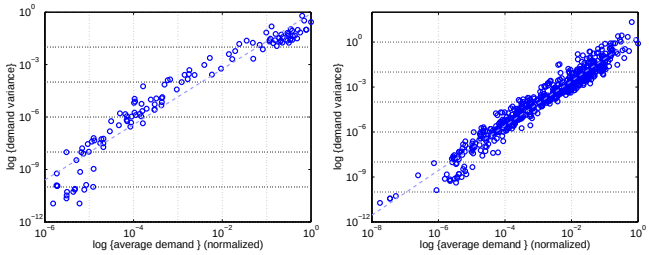

Figure 6: Relation between mean and variance for the demands in the European (left) and American (right) subnetworks.

Similar mean-variance relationships have been established for web-traffic in [25] and for IP traffic demands in [5, 12]. Our observations are consistent with the measurements on a single LAN router in [5] (which suggest that $c = 2$ is more reasonable than the Poissonian assumption $c = 1$), but differs from the measurements on the Sprint backbone reported in [12] (which finds that c varies uniformly over the interval $[0.5, 4.0]$). This difference could be explained from the fact that [12] calculates the 1-second mean-variance relationship *per demand* over 400 intervals of 100 seconds each. The variation of the per demand mean over these 400 intervals (a little more than 11 hours) is not going to be very large. In our analysis, we use the 5-minute mean-variance numbers from all demands during a single interval, like the busy hour for which we want to estimate the traffic matrix. This way we fit the data over an average demand range of 6 magnitudes or more, based on the same measurement intervals that will be used for the estimation procedures.

5.2.4 On the Gravity model assumption

Finally, we investigate to what extent the gravity model provides a good estimate of the demands. We focus our analysis on the simple gravity model although the generalized gravity model potentially yield more accurate results since the latter model requires information we do not have access to. Figure 7 shows the actual traffic matrix elements against the gravity model estimates. While the gravity model is reasonably accurate for the European network, it significantly underestimates the large demands in the American network. With our knowledge about the spatial distribution of demands shown in Figure 3 we could have foreseen this result. Contrary to the gravity model assumption that all PoPs send the same fraction of their total traffic to each destination, PoPs tend to have a few dominating destinations that differ from PoP to PoP.

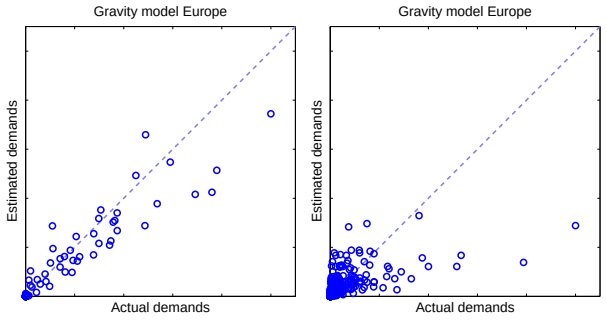

Figure 7: Real demands vs. gravity model estimates for European (left) and American (right) subnetworks.

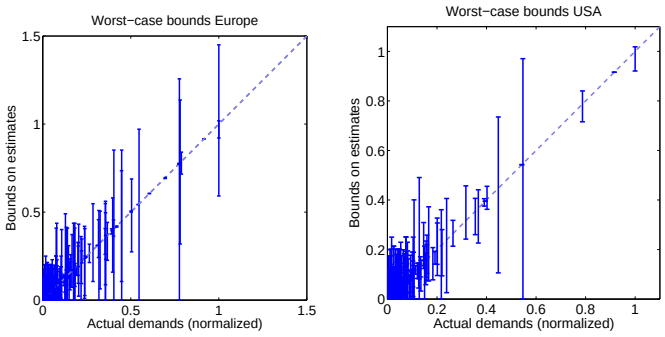

Figure 8: Worst-case bounds on demands in European (left) and American (right) subnetworks.

5.3 Evaluation of Traffic Matrix Estimation Methods

In this section, we evaluate the methods for traffic matrix estimation described in Section 4. Since fanout estimation and the Vardi approach both use a time-series of measurements rather than a snapshot, they are analyzed separately from the other methods.

5.3.1 Performance Metrics

To evaluate the methods, we must first determine an appropriate performance measure. Although many aspects could potentially be included in the evaluation, we focus on the potential impact of performance errors on traffic engineering tasks such as load balancing or failure analysis. For these applications, it is most important to have accurate estimation of the largest demands since the small demands have little influence on the link utilizations in the backbone. We will thus focus our performance analysis on how well the methods are able to estimate the large demands. In order to quantify performance of the estimation and compare results from different estimation methods we introduce the mean relative error (MRE):

$$MRE = \frac{1}{N_T} \sum_{i:s_i>s_T} \left| \frac{\hat{s}_i - s_i}{s_i} \right| \quad (8)$$

Here, s_i denotes the true traffic matrix element and $\hat{s}_i$ denotes the corresponding estimate. The sum is taken over the elements in s larger than s_T and N_T is the number of elements in s larger than the threshold. In our analysis, we have chosen the threshold so that the demands under consideration carry approximately 90% of the total traffic. This corresponds to including the 29 largest demands in the European subnetwork, and the 155 largest demands in the American network.

5.3.2 Evaluation of Worst-Case Bounds

To get a feel for how difficult it is to estimate different demands, it is useful to compute worst-case bounds for the demands using the approach described in Section 4.3.1. The resulting bounds for are shown in Figure 8.

Although most bounds are non-trivial, they tend to be relatively loose and only very few bounds can be measured exactly. Still, as shown in Figure 9 the average of the upper and lower bound for each flow gives a relatively accurate estimate of the demands. We can observe that many of the largest demands in the European subnetwork have relatively large worst-case bounds, indicating potentially large uncertainty in the estimates.

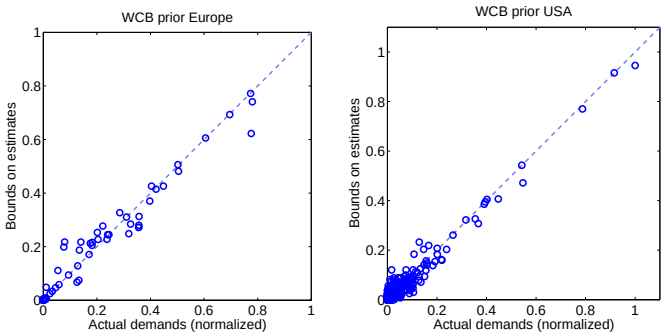

Figure 9: Priors obtained from worst-case bounds.

5.3.3 Evaluation of Fanout Estimation

Figure 10 shows the results of the fanout-based estimation scheme on the American subnetwork. Since the approach uses a time-series of link load measurements, we have the average demands over the time window on the x-axis against the estimated average demand on the y-axis. Although the system of equations becomes overdetermined already for a window length of 3, the actual performance only improves marginally as we include more data.

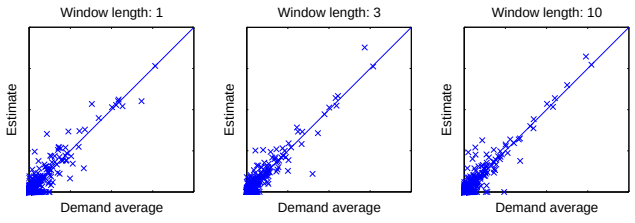

Figure 10: Average demands over time window vs. estimates for the fanout estimation procedure using actual data from American subnetwork.

To quantify the error we plot the MRE as a function of the window length as shown in Figure 11. The figure shows that the error decreases for short time-series of measurements, but levels out for larger window sizes.

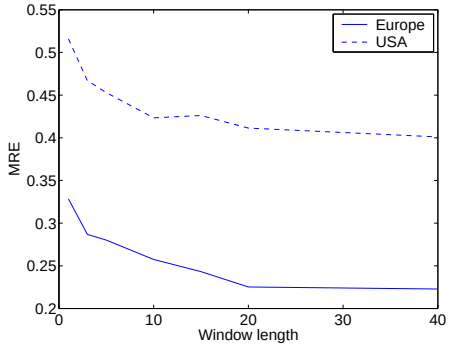

Figure 11: MRE as a function of window length.

	Europe	America
$\sigma^{-2} = 0.01$	0.47	0.98
$\sigma^{-2} = 1$	302	1183

Table 1: MRE for the Vardi approach, $K = 50$

5.3.4 *Evaluation of the Vardi approach*

In the analysis of the Vardi approach, we apply the method on the busy period of respective network (*i.e.*, the shaded interval in Figure 1). The busy period is 250 minutes, or 50 samples long, and we use the sample mean of the traffic demands over the busy period as the reference value in the MRE calculations.

Table 1 shows MRE for $\sigma^{-2} = 0.01$ and $\sigma^{-2} = 1$. The value $\sigma^{-2} = 1$, which corresponds to strong faith in the Poisson assumption, gives unacceptable performance; some estimates are several orders of magnitude larger than the true demands while other elements are set to zero despite that the corresponding demand is non-zero. Smaller values of σ give better performance, but are still not very convincing. We believe there are two reasons for the poor performance. First, although there is a strong mean-variance relationship, the analysis in Section 5.2.3 has shown that the demands are not Poissonian. Second, the convergence of the covariance matrix estimation is slow and one needs a large set of samples to have an accurate estimate. To support this argument, we calculate the mean of the elements of the traffic matrix over the busy period and generate a time-series of synthetic traffic matrices with Poisson distributed elements with the calculated mean. Figure 12 shows MRE as a function of window size for synthetically generated traffic matrices. The solid line shows the error for the European network and the dashed line the error for the American network. To have errors in the estimation less than 20% we need a window size of 100 for the American network. Hence, even when the Poisson assumption is valid, a large window size is needed in order to achieve an acceptable level of the estimation error.

5.3.5 *Comparison of Bayesian and Entropy models*

In this section, we evaluate the methods that use a single snapshot measurement from the network. We use the simple gravity model as prior. As before, the threshold value of the MRE method is adjusted so that approximately 90% of the total traffic in the network is included in the study.

Relying on regularization, the results of both the Bayesian (7) and the Entropy (6) approach depend on the choice of

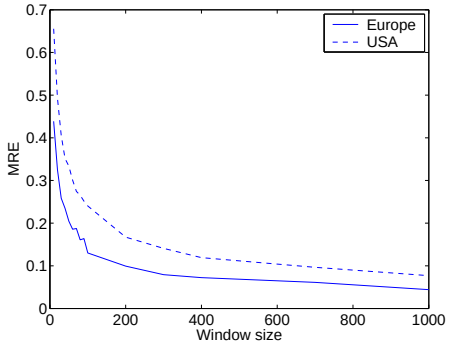

Figure 12: MRE as a function of window size for a synthetic traffic matrix, $\sigma^{-2} = 1$

regularization parameter. For a small values of σ we make little use of the measurement and focus on finding a solution that is close to the prior. For very large values of σ, on the other hand, we put a strong emphasis on the measurements, and only use prior to select the most plausible solutions of the demand estimates that satisfy $Rs = t$. This is clearly shown in Figure 13, where we have computed the MRE values for both methods as function of the regularization parameter. The leftmost values should be compared with the MRE of the gravity prior, which is 0.26 in European and 0.8 in the American subnetwork. As the plots show, we get the best results for large values of the regularization parameter. We can also see that there is no single best method; the Bayesian performs better in Europe while the Entropy approach works better in the American subnetwork.

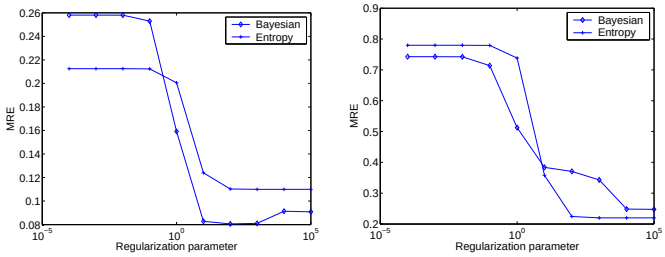

Figure 13: Mean relative error (MRE) as a function of the regularization parameter for the European network (left) and the American network (right).

To gain intuition about the performance of the estimation we have plotted the actual traffic matrix elements against the estimated for the American network. Figure 14 shows the plot for Bayesian (left) and Entropy (right) estimation. The regularization parameter was set to 1000 producing the best possible estimation for both Bayesian and Entropy estimation. The plots show that the estimation manage to capture the traffic demands for the whole spectrum of traffic demands.

Finally, we have demonstrated that using the mean of the upper and lower worst-case bound for each demand resulted in an estimate which is significantly better than the gravity model, and is thus natural to use this as an alternative prior in the regularized approaches. Figure 15 shows the MRE for the Bayesian approach as function of regularization pa-

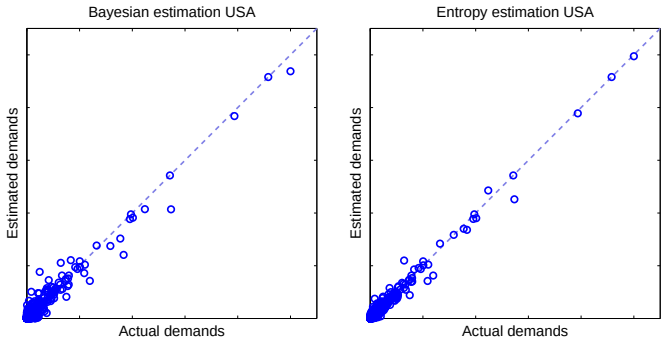

Figure 14: Real vs. estimated traffic demands for the American subnetwork using the Bayesian approach (left) and Entropy estimation (right).

rameter for the gravity and worst-case bound prior on the European (left) subnetwork and the American (right) subnetwork. We can see that the worst-case bound prior gives significantly better results for small values of the regularization constant (*i.e.* when we put large emphasis on the prior). For large values of the regularization parameter, however, the performance of the two priors is practically equal.

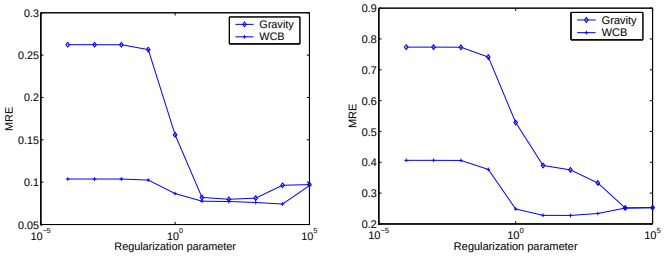

Figure 15: Mean relative error (MRE) as a function of the regularization parameter for the European (left) network and the American (right) network using gravity and worst-case bound priors.

5.3.6 *Combining Tomography with Direct Measurements*

As a final exercise, we investigate the usefulness of combining traffic matrix estimation based on link-loads with direct measurements of specific demands. To get correspondence with the rest of this paper, we focus on the problem of adding measurements that allow us to decrease the MRE of the Entropy method.

Figure 16 shows how the MRE for the Entropy approach decreases with the number of measured demands for the European subnetwork. We can see that it is sufficient to measure six demands in order for the MRE to drop from the initial 11% to below 1%. For the American network, on the other hand, we need to measure 17 demands for the MRE to decrease from the initial 23% to below 10%. These results are generated by finding, by exhaustive search in each step, the demand that when measured gives the largest decrease in MRE. They indicate that significant performance improvements can be achieved by measuring only a handful demands.

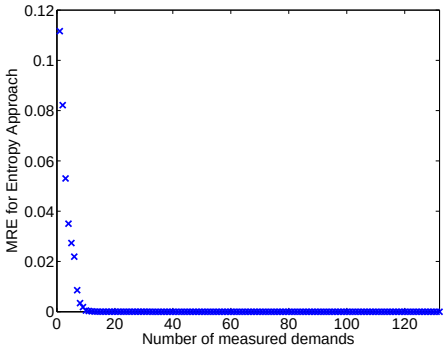

Figure 16: The MRE versus number of demands that we measure exactly in the European network.

	Europe	America
Worst-case bound prior	0.10	0.39
Simple gravity prior	0.26	0.78
Entropy w. gravity prior	0.11	0.22
Bayes w. gravity prior	0.08	0.25
Bayes w. WCB prior	0.07	0.23
Fanout	0.22	0.40
Vardi	0.47	0.98

Table 2: Performance comparison of the various methods. The table shows the best MRE values that we have been able to achieve for the various methods on the two subnetworks.

In practice, however, one would also need an approach for choosing the best demand. Comparing Figures 16 and 1, one is easily led to believe that they are nothing but each others' inverses, and it would be sufficient to measure the largest demands. In passing, we note that most estimation methods are very accurate in ranking the size of demands, so identifying the largest demands and measuring them is indeed a viable practical approach. However, the MRE measures the relative error, and in our data set, it is not the largest demands that have the largest relative estimation errors. In Europe, one would need to measure the 19 largest demands to have a MRE less than 1%, and in the American network, one would need to measure 74 demands to force the MRE below 10%.

5.3.7 *Evaluation in summary*

To allow an easy performance comparison of the methods, Table 2 summarizes the best MRE values that we have been able to achieve for the different approaches. The table demonstrates that the Bayesian and Entropy methods gave the best performance, followed by the fanout and Vardi approaches. The worst-case bounds provide a better prior than the simple gravity model on our dataset, and both methods provide better MRE values than the Vardi approach. Note, however, that the fanout and Vardi approaches use, and are evaluated on, a sequence of link load measurements.

Since our experiences of other aspects of the methods, such as ease-of-use and computational complexity, are not easily summarized in a single numbers, we have omitted a direct comparison and refer to the discussions above.

6. CONCLUSION AND FUTURE WORK

This paper has presented an evaluation of traffic matrix estimation techniques on data from a large IP backbone. In contrast to previous studies that used partial traffic matrices or demands estimated from aggregated NetFlow traces, we have used a unique data set of complete traffic matrices measured over five-minute intervals. The data set has allowed us to do accurate data analysis on the time-scale of standard link-load measurements and enabled us to evaluate both methods that use a time-series of link-loads and methods that rely on snapshot measurements.

We have shown that the demands in our data set have a remarkably strong mean-variance relationship, yet we have been unable to achieve good estimation performance using methods that try to exploit this fact. We have argued that this failure is due the problem of accurate estimation of covariance matrices and presented a study on synthetic data to support this claim.

Based on our observation that fanout factors tend to be much more stable over time than the demands themselves, we have proposed a novel method for estimating fanouts based on a time-series of link load measurements. We have also proposed to estimate worst-case bounds on the demands. Although these bounds are not always very tight, they turned out to be useful for constructing a prior for use in other estimation schemes. We have illustrated that the gravity model fails to construct a good prior in one of our subnetworks due to violations of underlying assumptions in the traffic patterns. The regularized methods, such as Bayesian and Entropy approaches, were found to be simple and provide the best results, if the regularization parameter was chosen appropriately. Finally, we noted that by measuring only a handful of demands directly, it was possible to obtain significant decreases in the MRE of the Entropy approach.

This study has focused on analyzing key properties of the demand data set and evaluating the performance of traffic matrix estimation techniques in terms of their estimation error. Although we have covered most methods from the literature, we have not implemented and evaluated the approach by Cao *et al.* [5]. Clearly, a more complete evaluation should include also this method. It would also be useful to complement the evaluation by a more rigorous theoretical analysis to bring a better understanding of our observations. Our study also leaves many important issues unexplored. For example, our data set does not contain measurement errors or component failures and we have not evaluated the effect of such events on the estimation. Furthermore, we have not considered how sensitive traffic engineering tasks are to estimation errors in different demands, and how such information could be incorporated in the estimation procedures. Another interesting topic for future work would be to understand the nature of the worst-case bounds, and see if they could be exploited in other ways.

Acknowledgments

This work is supported in part by the Swedish Foundation for Strategic Research (SSF), the Swedish Research Council (VR), the Swedish Agency for Innovation Systems (VINNOVA) and the European Commission.

7. REFERENCES

[1] H. Abrahamsson, J. Alonso, B. Ahlgren, A. Andersson, and P. Kreuger, "A multi path routing algorithm for IP networks based on flow optimisation," in *From QoS Provisioning to QoS Charging – Third COST 263 International Workshop on Quality of Future Internet Services, QoFIS 2002 and Second Interntational Workshop on Internet Charging and QoS Technologies, ICQT 2002*, B. Stiller, M. Smirnow, M. Karsten, and P. Reichl, Eds., Zürich, Switzerland, Oct. 2002, pp. 135–144, Springer, LNCS 2511.

[2] A. Sridarhan, R. Guerin, and C. Diot, "Achieving near-optimal traffic engineering solutions for current OSPF/IS-IS networks," in *Proc. of IEEE INFOCOM 2003*, San Francisco, USA , November 2003.

[3] D. Applegate and E. Cohen, "Making intra-domain routing robust to changing and uncertain traffic demands: Understanding fundamental tradeoffs," in *Proc. ACM SIGCOMM*, Karlsruhe, Germany, August 2003.

[4] M. Roughan, Mikkel Thorup, and Yin Zhang, "Traffic engineering with estimated traffic matrices," in *Proc. ACM Internet Measurement Conference*, Miami Beach, Florida, USA, October 2003.

[5] J. Cao, D. Davis, S. Vander Wiel, and B. Yu, "Time-varying network tomography: router link data," *Journal of Americal Statistical Association*, vol. 95, pp. 1063–1075, 2000.

[6] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg, "Fast accurate computation of large-scale IP traffic matrices from link loads," in *Proc. ACM Sigmetrics*, San Diego, CA, June 2003.

[7] A. Medina, C. Fraleigh, N. Taft, S. Bhattacharyya, and C. Diot, "A Taxonomy of IP Traffic Matrices," in *SPIE ITCOM: Scalability and Traffic Control in IP Networks II*, Boston, Aug. 2002.

[8] J. Kruithof, "Telefoonverkeersrekening," *De Ingenieur*, vol. 52, no. 8, pp. E15–E25, 1937.

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

[10] C. Tebaldi and M. West, "Bayesian inference on network traffic using link count data," *Journal of the American Statistical Association*, vol. 93, no. 442, pp. 557–576, June 1998.

[11] S. Vaton and A. Gravey, "Network tomography : an iterative bayesian analysis," in *Proc. ITC 18*, Berlin, Germany, August 2003.

[12] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic matrix estimation: Existing techniques and new directions," in *Proc. ACM SIGCOMM*, Pittsburg, USA, August 2002.

[13] Y. Zhang, M. Roughan, C. Lund, and D. Donoho, "An information-theoretic approach to traffic matrix estimation," in *Proc. ACM SIGCOMM*, Karlsruhe, Germany, August 2003.

[14] A. Nucci, R. Cruz, N. Taft, and C. Diot, "Design of IGP link weight changes for estimation of traffic matrices," in *Proc. IEEE INFOCOM*, Hong Kong, March 2004.

[15] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, J. Rexford, and F. True, "Deriving traffic demands for operational IP networks: Methodology and experience.," in *Proc. ACM SIGCOMM*, Stockholm, Sweden, August 2000.

[16] R. S. Krupp, "Properties of Kruithof's projection method," *The Bell System Technical Journal*, vol. 58, no. 2, pp. 517–538, February 1979.

[17] A. Medina, K. Salamatian, N. Taft, I. Matta, Y. Tsang, and C. Diot, "On the convergence of statistical techniques for inferring network traffic demands," Tech. Rep. BUCS-2003-003, Boston University, Computer Science, USA, February 2003.

[18] G. K. Zipf, "Somde determinants of the circulation of information," *American Journal of Psychology*, vol. 59, pp. 401–421, 1946.

[19] J. Kowalski and B. Warfield, "Modeling traffic demand between nodes in a telecommunications network," in *Australian Telecommunications and Networks Conference*, Sydney, Australia, December 1995.

[20] M. Roughan, A. Greenberg, C. Kalmanek, M. Rumsewicz, J. Yates, and Y. Zhang, "Experience in modeling backbone traffic variability: models, metrics, measurements and meaning," in *Proc. ACM SIGCOMM Internet Measurement Workshop*, Marseille, France, November 2002.

[21] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statistics and Decisions, Suppl. 1*, vol. Supplement Issue No. 1, pp. 205–237, 1984.

[22] I. Csiszár, "Why least squares and maximum entropy? – an axiomatic approach to inverse problems," *The Annals of Statistics*, vol. 19, pp. 2033–2066, December 1991.

[23] X. Xiao, A. Hannan, B. Bailey, and L. M. Ni, "Traffic engineering with MPLS in the internet," *IEEE Network*, vol. 14, no. 2, pp. 28–33, March–April 2000.

[24] Cariden, Inc., Mountain View, CA, *MATE*, 2004, http://www.cariden.com.

[25] R. Morris and D. Lin, "Variance of aggregated web traffic," in *Proc. IEEE INFOCOM*, Tel Aviv, Israel, March 2000, pp. 360–366.

[26] M. Coates, A. Hero, R. Nowak, and B. Yu, "Internet tomography," *Signal Processing Magazine*, vol. 19, no. 3, pp. 47–65, May 2002.

[27] S. Bhattacharyya and C. Diot and J. Jetcheva and N. Taft, "Pop-Level and Access-Link-Level Traffic Dynamics in a Tier-1 PoP," in *Proc. ACM SIGCOMM Internet Measurement Workshop*, San Francisco, USA , November 2001.