IP Traffic Matrix Estimation Methods: Comparisons and Improvements

Md. Mostafizur Rahman, Subrata Saha, Usha Chengan, and Attahiru Sule Alfa
Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, MB, Canada R3T 5V6
Email: {mmrahman, subrata, uchengan, alfa}@EE.UManitoba.ca

***Abstract*— Determining point to point traffic matrix is essential for Internet service providers (ISPs) in carrying out traffic engineering tasks for network management and planning purposes. However, it is very difficult and costly to measure this traffic matrix directly. Hence, traffic matrices are inferred from link measurements through estimation, using different techniques. There are different techniques for this traffic matrix estimation and there is still a need for evaluating these existing techniques. Some of those techniques have been previously compared, but with new improved techniques recently developed there is a need to revisit the comparisons. In this paper, we have carried out studies to compare three very popular methods: the tomogravity, the entropy maximization and linear programming methods. We find that the tomogravity method best estimates the traffic matrix among the methods we tested. We then incorporate some enhancements which improve this method. Specifically we established that knowing some point to point traffic may improve the estimation but not necessarily, and this is counter-intuitive. We modify the existing entropy maximization method by adding more constraints and we find that our modified method outperforms the existing entropy maximization and tomogravity methods.**

Keywords: Network Tomography, Tomogravity, Entropy Maximization, Linear Programming

I. Introduction

Each element of a point to point traffic matrix represents the volume of traffic originating from an origin and intended to a possible destination in a telecommunication network. In other words, a traffic matrix gives the amount of data transferred from a origin node i to a destination node j for all possible origin-destination (OD) pairs (i, j). Traffic Matrix is important for Internet service providers (ISPs) for traffic engineering tasks such as designing or upgrading the network topology, performing capacity planning, load balancing and configuring network routing policies. However, it is impractical and costly to measure this matrix by monitoring the origin and destination addresses of Internet protocol (IP) packets. On the other hand, the amount of traffic in a link between routers, or between routers and end hosts can be easily measured by collecting aggregate traffic statistics using simple network management protocol(SNMP). These measured link data are used to infer the point to point traffic matrices. Inference of the traffic matrix is important for ISPs that use it to minimize their cost of providing Quality of Service (QoS) by not over provisioning the network resources. Traffic matrices can be of different granularity. For an ISP traffic matrices can be the volume of data between Points-of-Presence (PoPs) or backbone routers. On the other hand, for a local area network the traffic matrix can be the volume of data between the IP prefixes.

For a network with n routers there are $c = n(n-1)$ router to router traffic. However, the number of links, m, among these routers can be anything from $O(n)$ to $O(n^2)$ but in general $O(n)$. The number of links in a network depends on the topology of the network. Hence, traffic matrix estimation involves inferring the $n(n-1)$ router to router traffic from these m link measurements. If the column vector $Y = (Y_1, \ldots, Y_m)^T$ denotes the link measurements of m links and the column vector $X = (X_1, \ldots, X_c)^T$ represents the traffic between all ingress-egress pairs then the relationship between X and Y can be expressed as $Y = AX$. Here, A is an $m \times c$ matrix where the element $A_{i,j}$ represents the fraction of the traffic volume for the ingress-egress pair j crossing through link i. The network traffic matrix estimation problem is to determine the traffic intensity vector X given the traffic link data Y. The difficulty associated with estimating the traffic matrix is that the number of unknowns is far greater than the number of equations, i.e. $n(n-1) > m$. In other words the system is under-determined. There is no unique solution to the problem.

There are several methods developed for estimating traffic matrix from volume of link data measurement. We will discuss four of the techniques which we consider to be the major ones for estimating a traffic matrix. They are categorized as network tomography, network tomogravity method, entropy maximization and linear programming methods. We carry out extensive experiments on network tomography with weighted least squares method using different weights, network tomogravity with two MATLAB tools *pdsco* [5] and *maxent* [3] and linear programming method. We perform further experiments to assess whether direct measurement of a limited number of point to point traffic could reduce the error in the estimation of traffic matrix. One of our contributions is that we identify, through numerical experiments, that network tomogravity is best for traffic matrix estimation. Our second contribution is that we find that knowing some point to point traffic reduces the error in traffic matrix estimation as expected.

We organize the rest of the paper in the following way. In section II we explain network tomography; network to-

1-4244-0355-3/06/$20.00 (c) 2006 IEEE

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

mogravity is discussed in section III; entropy maximization in section IV; and linear programming in section V. We discuss our experiment in section VI and conclude the report in section VII.

II. Network Tomography

The problem of inferring point to point traffic in a network using measurements of data flowing through links is commonly referred to as network tomography which is analogous to the medical tomography problem of imaging a large number of cells of the internals of the human body from a smaller number of scans. In medical tests tomographic methods are widely used in Computer Aided Tomography (CAT) scans. CAT uses an optimization function to determine the solution of under-determined system in such a way that the objective function reaches its optimum value. On the other hand, network tomography approaches which are adapted for inferring point to point traffic use higher order moments of the link data to create additional constraints for the under-determined system.

In network tomography methods, it is assumed that traffic through each link in the network is known, averaged over some time interval. In these methods knowledge of how the traffic of each ingress-egress pair is routed is combined with measured aggregate link rates to infer the traffic matrix. The assumption of these methods is that the routes from every ingress-egress pair are stable and known over a period of time. Network tomography consists of the statistical framework and the associated computational methods to infer estimation of every ingress-egress pair from the link data measured over some time interval. In these methods traffic flows are assumed to be some parameterized stochastic process. Network tomography infers the missing spatial information about every ingress-egress pair from multiple measurements of link data over time.

Vardi [7] first used the idea of tomography to estimate traffic matrix. He assumed that every ingress-egress traffic intensity Y_i's are independent Poisson random variables with covariance $Cov(Y_i, Y_j) = \sum_k B_{ij,k} Y_k$ where $B_{ij,k} = A_{ik} A_{kj}$ The equation $Y = AX$ is supplemented with the equation $Z = BX$ where Z_{ij} denotes the measured covariance of the traffic across links i and j. For m links there are $\frac{m(m+1)}{2}$ number of Z_{ij}s and B is $\frac{m(m+1)}{2} \times c$ matrix. After adding these higher order moments, the system of equations can be compactly written as

$$\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} X. \tag{1}$$

It is problematic to directly solve equation 1. Vardi [7] used expectation maximization (EM) algorithm iteratively to solve equation 1.

Tebaldi and West [6] and Cao *et al.* [1] worked on network tomography. Tebaldi and West followed Bayesian approach by including prior information into the estimate. Cao *et al.* followed Vardi's approach but assumed the data rate for every ingress-egress pair to be Gaussian. In the model of Cao *et al.* variance of link data is assumed to be proportional to some power of mean link data rates. Cao *et al.* used a modified EM algorithm to find maximal likelihood parameters in the Gaussian model.

Medina *et al.* [4] showed that the Poisson and Gaussian assumptions of network tomography are not justified. They showed that network tomography is not successful when the assumptions of the approaches are violated. They also showed that the number of steps required to compute the matrix is computationally challenging. Moreover, they showed that the amount of sample data required to reach a sufficient precision is large. Therefore, network tomography is not applicable to a large-scale traffic estimation problem. However, this method can be used to obtain a "local" traffic matrix at a single router. We also contend that the assumption of Poisson or Gaussian distribution is unrealistic. In fact the assumption of any standard statistical distribution may be unrealistic.

III. Network Tomogravity

Gravity model is an alternative to network tomography model based on modified form of Newton's law of gravitation. It has been used to predict movement of people, information, and commodities between cities and even continents. When gravity model is combined with a tomographic estimation step the method is called tomogravity method.

Newton's law of gravitation involving two particles can be expressed in mathematically in the following way

$$F = G\frac{m_1 m_2}{d^2}, \tag{2}$$

where F represents the force of gravitation between two particles, G represents the universal gravitation constant, m_1 and m_2 are the masses of the two particles and d is the separating distance between two particles. In a similar fashion gravity model is adapted for estimation of traffic among areas. Gravity model for traffic estimation can be expressed by the following equation

$$X_{ij} = \frac{R_i A_j}{f_{ij}}, \tag{3}$$

where X_{ij} is the matrix element representing the force from i to j; R_i represents the repulsive factors that are associated with "leaving" from i; A_j represents the attractive factor associated "going" to j; and f_{ij} is a friction factor from i to j. Here, X_{ij} is the amount of traffic that enters the network at location i and leaves at location j which is equivalent to force of gravitation F. Similarly the repulsion factor R_i is the total amount of traffic leaving point i which is equivalent to m_1, the mass of the first particle. In the same way the attractivity factor A_j is the traffic volume exiting at location j which is equivalent to m_2, the mass of the second particle. The friction matrix (f_{ij}) is more like impedance for traffic to travel between two areas and is equivalent to the ratio $\frac{G}{d^2}$. The inference of $n \times (n-1)$ friction matrix is of same complexity of inference as that of the traffic matrix. Hence, simplifying assumptions are taken for friction matrix. Sometimes the elements of the friction matrix are taken as a common constant, for example either $f_{ij} = \sum_j A_j$ or $f_{ij} = \sum_i R_i$; the two give different results.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

Gravity model does not accurately model the traffic between all OD pairs. However, when combined with detailed knowledge of ISP routing policies, it can match the actual network traffic very well. Medina *et al.* [4] used additional assumptions to estimate the traffic matrix. Their methods are choice models. Their choice models made some practical assumptions that network traffic from a particular origin to a particular destination depends on a set of alternative routes, attributes of the decision makers as well as the alternative and decision rules that govern the decision process. They modeled their decision process of choosing a particular decision based on a utility maximization criterion.

Zhang *et al.* [8] used tomogravity model to solve network traffic estimation problem. In their model, traffic matrix X_g is computed using generalized gravity model. This traffic matrix X_g is then used in the subsequent tomography step to compute the traffic matrix X. The traffic matrix estimation problem then becomes a quadratic programming problem below.

$$\begin{gathered}\min ||X - X_g|| \\ \text{subject to } ||AX - Y|| \text{ is minimized}\end{gathered}$$

where $||.||$ is the L_2 norm of a vector.

In tomography step some techniques are used to reduce the number of unknown X_is. After reducing the number of unknowns in the traffic matrix, singular value decomposition is applied to solve the quadratic program to find a solution that minimizes its distance to X_g subject to the tomographic constraints. Zhang *et al.* [8] also investigated weighted least-squares (or wlse) solutions as the objective function to minimize the above quadratic program. The three weighting schemes used are — i) the weight for each term of the traffic matrix is constant, ii) linearly proportional to the terms in the gravity model traffic matrix and iii) proportional to the square root of the gravity model. After the quadratic programming is terminated negative values of X_{ij}s are replaced with zero values and iterative proportional fitting (IPF) is performed to obtain a non-negative solution that satisfies the constraints.

IV. Entropy Maximization

Zhang *et al.* [9], [10] recently proposed the use of information theoretic approach to estimate the traffic matrix. This information theoretic approach is a generalization of their tomogravity model. In this recent work they showed that their solution selects the traffic matrix that is information theoretically closest to a model where origin/destination pairs are independent. In this model entropy is used for estimating the traffic matrix. In information theory the Discrete Shannon Entropy of a discrete random variable X taking values x_i is defined as

$$H(X) = -\sum_i p(x_i) log_2(p(x_i))$$

where $p(x_i)$ is the probability of X taking the value x_i. $H(X)$ takes maximum value when X is uniformly distributed ($i.e.$ when the uncertainty about its value is the greatest).

Maximum entropy principle states that enumerating all the constraints that the unknown probability distribution obeys and searching the probability distribution space that maximizes the entropy subject to those constraints should estimate an unknown probability distribution. In other words, given some constraints C, the random variable X can be estimated by maximizing the entropy $H(X|C)$. If there are no constraints, *i.e.*, no prior information, then X can be estimated by maximizing $H(X)$.

Now, $H(Y|X) = -\sum_j p(x_j) \sum_i p(y_i|x_j) log_2(p(y_i|x_j))$ and the joint entropy of X and Y can be written as $H(X,Y) = -\sum_{i,j} p(x_i,y_j) log_2(p(x_i,y_j))$. Mathematically, we can write $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$. We can also define Shannon information $I(X,Y) = H(X) + H(Y) - H(X,Y) = \sum_{i,j} p(x_i,y_j) log_2 \frac{p(x_i,y_j)}{p(x_i)p(y_j)}$. Here, $I(X,Y)$ represents the decrease in uncertainty about Y from the measurement of X or the decrease in uncertainty about X from the measurement of Y. $I(X,Y)$ is also called mutual information. The mutual information can also be written in the following way

$$I(X,Y) = \sum_{x,y} p(x,y) log_2 \frac{p(x,y)}{p(x)p(y)} = K(p_{x,y}||p_x \times p_y) \quad (4)$$

where $K(f||g) = \sum_i f_i log_2 \frac{f_i}{g_i}$ is the Kullback-Leibler divergence of f with respect to g, a well-known measure of distance about probabilities.

The goal of entropy maximization results in minimizing mutual information (MMI). This goal is achieved by the following optimization problem

$$minimize \sum_l (T(l) - N \sum_{s,d} A(s,d;l) p(s,d))^2 + \lambda^2 I(S,D) \quad (5)$$

Here, $T(l)$ is the traffic though the link l. N is the total amount of traffic entering or exiting the network. $p(s)$ is the fraction of traffic exiting from the origin s. $p(d)$ is the fraction of traffic entering the destination d. $p(s,d)$ is the fraction of traffic exiting from the origin s and entering the destination d. The amount of traffic between origin s and destination d is computed from this fraction $p(s,d)$. $A(s,d;l)$ is 1 if the traffic between s and d uses the link l, otherwise it is 0. λ is a penalizing factor which depends on the level of noise. S is the set of origins. D is the set of destinations.

pdsco [5] and *maxent* [3], two different MATLAB packages, can be used to solve equation 5. The papers [9], [10] also discuss an approximation to calculate $I(S,D)$. The papers compared the performance of MMI, its approximation and the tomogravity model (SVD and IPF) and showed that MMI outperforms other methods with respect to accuracy.

V. Linear Programming

Conway and Li [2] developed methods for inferring OD traffic matrix from aggregate link and origin/sink measurements that are taken over disjoint time periods. They developed methods for estimating OD traffic proportions where OD traffic proportion p_{ij} ($1 \leq i,j \leq n$ where n is the number of nodes) means that p_{ij} fraction of traffic from the origin i destined to j. If in time-period t, the total measured traffic

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

arriving at node i is $\alpha_i^{(t)}$, the total measured traffic leaving node i is $\beta_i^{(t)}$, the total measured traffic on link k ($1 \leq k \leq m$ where m is the number of links) is $\gamma_k^{(t)}$ and r_{ijk} is the proportion of traffic originating at node i and destined to node j flowing through link k then

$$\beta_i^{(t)} = \sum_{j=1}^{n} \alpha_i^{(t)} p_{ji} : 1 \leq i \leq n, \tag{6}$$

$$\gamma_k^{(t)} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i^{(t)} p_{ij} r_{ijk} : 1 \leq k \leq m, \tag{7}$$

$$\sum_{j=1}^{n} p_{ij} = 1 : 1 \leq i \leq n. \tag{8}$$

We solve equations 6, 7 and 8 to obtain the point to point traffic matrix p_{ij}s. However, one measurement period does not provide sufficient number of independent equations. Hence, measurement of multiple time periods are used construct the required number of independent equations by using row-echelon technique. The next step is to solve this set of independent equations to infer the unknown p_{ij}s. There are errors generated at this step. Hence, the inference step formulates a linear programming to minimize this error. The positive and negative error terms are denoted by the positive column vectors ${\mathbf{e}^+}^T$ and ${\mathbf{e}^-}^T$ respectively. The linear programming formulation is as follows

We want to determine the vector $[[\mathbf{x}^{(1)}, ..., \mathbf{x}^{(n)}], [\mathbf{e}^+, \mathbf{e}^-]]^T$ which *minimizes* $\mathbf{1}(\mathbf{e}^+ + \mathbf{e}^-)^T$

subject to $\mathbf{A}[\mathbf{x}^{(1)}, ..., \mathbf{x}^{(n)}]^T = \mathbf{b}^T + {\mathbf{e}^+}^T - {\mathbf{e}^-}^T$,
$\mathbf{1}{\mathbf{x}^{(\mathbf{i})}}^T = 1$, for $1 \leq i \leq n$,
${\mathbf{x}^{(\mathbf{i})}}^T \geq \mathbf{0}^T$, for $1 \leq i \leq n$,
${\mathbf{e}^+}^T \geq \mathbf{0}^T$,
${\mathbf{e}^-}^T \geq \mathbf{0}^T$

where $\mathbf{1}$ and $\mathbf{0}$ are compatible vectors of ones and zeros respectively, ${\mathbf{x}^{(\mathbf{i})}}^T$ is the column vector of unknowns corresponding to $\{p_{ij}$: $1 \leq j \leq n$, $j \neq i\}$, $\mathbf{b}^T$ is a compatible column vector of constants and $\mathbf{A}$ is a matrix of constant coefficients.

Conway and Li [2] tested their method on two directed level graphs and achieved good performance. They also developed another approximation method which achieves quite comparable performance with less computation. However, the level graphs with which they tested their methods are not representative of any practical network. Furthermore, their method infers average point to point traffic and as the point to point traffic does not follow any particular statistical distribution this average point to point traffic cannot give any clear insight about the nature of the point to point traffic.

VI. Comparison Methodology

Since real traffic matrices are not available we conduct our experiment with artificially constructed traffic matrices. We compare tomogravity, entropy maximization and linear programming on a 4-node network. Then we study a 14-node network which was used by other researchers to test their models. We did not consider the delay in our experiments as the traffic matrix estimation is not performed online.

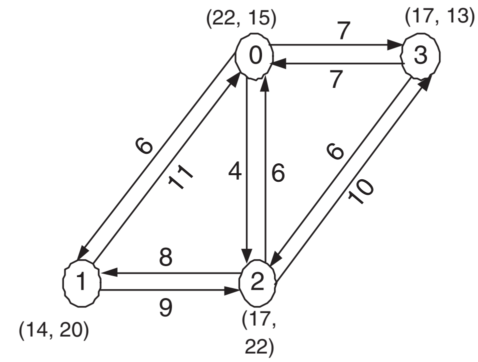

Fig. 1. 4-node Network

A. An Example Network – the 4-node Network

The first experiment we carried out was on a very simple network shown in Figure 1. There are four vertices and ten directed edges in the graph. Each vertex is associated with two numbers which are respectively its incoming and outgoing traffic (e.g. at vertex 1 incoming and outgoing traffic are 14 and 20 respectively). Each directed edge or link is associated with one number which is the amount of traffic passing through the link (e.g. 6 unit of traffic flows through the link 0-1). For the four vertices there are twelve OD pairs in the network.

In Table I, we show the original OD and estimated OD traffic using tomogravity [8] with weighted least squares *wlse*, entropy maximization [9], [10] with *maxent* [3] and *pdsco* [5] and linear programming [2]. We show the amount of calculated traffic for different weights for tomogravity method and compared the calculated and the original traffic matrices. Here, $w = 1$ implies that the weight for each OD pair is 1. On the other hand, $w = {t_g}^i$ implies that the weight for each OD pair is raised to the i-th power of OD traffic calculated from the gravity method.

From our experiment with 4-node network we find that tomogravity method with weight $w = {t_g}^4$ gives very good results. We discover that entropy maximization method [9], [10] using *pdsco* [5] is as good as tomogravity method. We can also see that entropy maximization method [9], [10] using *maxent* [3] is the worst among the tools used for estimating traffic matrix.

B. A Second Example Network – 14-node Network

After experimenting with 4-node network we experiment with a 14-node network which is shown in Figure 2. This 14-node network was used by other researchers [4]. In this network there are 26 undirected edges. Each undirected edge represents two directed edges. Therefore, there are 52 directed edges in the network. We constructed five different types of synthetic traffic matrices using the following distribution —

OD Pair	Original OD Traffic	Estimated OD Traffic							
		wlse							
		$w = 1$	$w = {t_g}^{0.5}$	$w = t_g$	$w = {t_g}^2$	$w = {t_g}^4$	*maxent*	*pdsco*	*linprog*
0-1	5	3.805	3.958	4.089	4.283	4.493	2.108	3.959	5
0-2	4	4.000	4.000	4.000	4.000	4.000	4.000	4.000	4
0-3	6	3.677	3.730	3.770	3.812	3.781	1.615	3.732	6
1-0	10	7.677	7.730	7.770	7.812	7.781	5.615	7.732	12.433
1-2	8	5.677	5.730	5.770	5.812	5.781	3.615	5.732	8.684
1-3	2	6.646	6.540	6.460	6.376	6.438	10.771	6.537	1
2-0	6	6.000	6.000	6.000	6.000	6.000	6.000	6.000	6
2-1	7	5.805	5.958	6.089	6.283	6.493	4.108	5.959	7
2-3	9	6.677	6.730	6.770	6.812	6.781	4.615	6.732	9
3-0	6	4.805	4.958	5.089	5.283	5.493	3.108	4.959	7.194
3-1	2	4.389	4.083	3.823	3.434	3.014	7.785	4.081	1
3-2	5	3.805	3.958	4.089	4.283	4.493	2.108	3.959	10.024

TABLE I

RESULT FROM WLSE FOR 4-NODE NETWORK

Distribution	Max. Err.			Avg. Err.			Std. Dev.		
	wlse	*pdsco*	*linprog*	*wlse*	*pdsco*	*linprog*	*wlse*	*pdsco*	*linprog*
Constant	4.67	4.66	209.01	1.92	1.92	41.12	1.37	1.37	32.91
Uniform	95.77	98.95	715.73	21.76	21.87	122.59	16.32	16.53	99.63
Poisson	82.41	83.58	65.32	21.21	21.1	22.68	15.09	15.2	17.21
Normal	84.05	84.76	608.66	22.73	22.59	51.8	17.39	17.55	58.44
Bimodal	133.47	136.5	342.47	33.95	32.53	40.77	23.15	23.26	44.29

TABLE III

ESTIMATION OF OD TRAFFIC FOR 14-NODE NETWORK

Distribution	Max. Err.			Avg. Err.			Std. Dev.		
	wlse	*pdsco*	*linprog*	*wlse*	*pdsco*	*linprog*	*wlse*	*pdsco*	*linprog*
Constant	4.69	45.97	602.13	1.69	5.93	101.88	1.39	7.47	94.93
Uniform	96.23	99.7	769.43	19.8	20.42	83.47	17.14	18.09	110.58
Poisson	66.29	66.29	151.89	19.14	19.44	28.56	15.14	15.76	24.92
Normal	89.12	113.07	676.95	21.05	23.49	61.68	17.47	20.81	71.6
Bimodal	125.2	110.82	442.61	31.27	30.24	45.97	24.53	25.07	55.13

TABLE IV

ESTIMATION OF OD TRAFFIC FOR 14-NODE NETWORK WITH SOME KNOWN VALUE

uniform, bimodal, normal, Poisson and constant. All the results we show are for the highest 80% point to point traffic.

Uniform point to point traffic is obtained by generating continuous uniform random values in the interval [100, 500]. Bimodal traffic is generated by a mixture of two Gaussian random variables, the first gaussian variable is $N(\mu_1 = 150; \sigma_1 = 20)$ with probability 0.8, and the second one is $N(\mu_2 = 400; \sigma_2 = 20)$ with probability 0.2. The normal distributed traffic is generated from $N(\mu = u; \sigma = 40)$ where u is a continuous uniform random variable in the interval [100, 400]. Poisson traffic is generated from $Poisson(\lambda)$ where λ is a continuous uniform random variable in the interval [100, 500]. Constant value is taken by assigning all the values of the point to point traffic as 300. All these distributions and the network of Figure 2 are taken from [4].

In Table II, we present the results of tomogravity method for the network of Figure 2 with different probability distributions for point to point traffic. The maximum relative absolute error percentage, average of relative absolute error percentage and standard deviation for relative absolute error percentage are showed. Here, relative absolute error is $|\frac{d_a - d_m}{d_a}|$ where d_a is the actual data and d_m is the measured data. We use five different weights for the computation — $w = 1$, $w = {t_g}^{0.5}$, $w = t_g$, $w = {t_g}^2$ and $w = {t_g}^4$. We can see from Table II that different values of weight gives different result for different distributions. How the choice of weights impacts the results is not conclusive. We can find from Table II that the estimation of point to point traffic for different probability distribution have different amount of errors. The amount of average percentage error for constant point to point traffic is the smallest among all the five cases. On the other hand, the average percentage error for the bimodal distributed point to point traffic is the largest among the differently distributed point to point traffic.

In Table III, we show the result of estimation for the network of Figure 2 with different probability distributions using *wlse* for tomogravity with constant weight, *pdsco* [5] for entropy maximization and linear programming. The average percentage error is the smallest for constant point to point

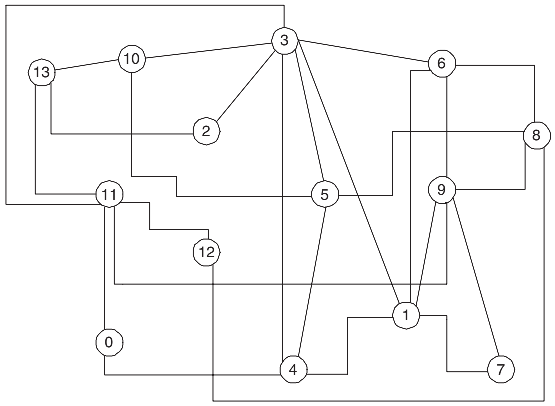

Fig. 2. 14-node Network [4]

Distribution	Weight	Max. Err.	Avg. Err.	Std. Dev.
Constant	$w = 1$	4.67	1.92	1.37
	$w = {t_g}^{0.5}$	4.67	1.92	1.37
	$w = t_g$	4.67	1.92	1.37
	$w = {t_g}^2$	4.67	1.92	1.37
	$w = {t_g}^4$	4.67	1.92	1.37
Uniform	$w = 1$	95.77	21.76	16.32
	$w = {t_g}^{0.5}$	99.11	21.82	16.56
	$w = t_g$	102.2	21.92	16.81
	$w = {t_g}^2$	107.35	22.28	17.39
	$w = {t_g}^4$	112.58	23.66	18.78
Poisson	$w = 1$	82.41	21.21	15.09
	$w = {t_g}^{0.5}$	82.9	21.01	15.2
	$w = t_g$	83.5	20.89	15.34
	$w = {t_g}^2$	85.33	20.87	15.7
	$w = {t_g}^4$	93.01	21.58	16.62
Normal	$w = 1$	84.05	22.73	17.39
	$w = {t_g}^{0.5}$	84.66	22.61	17.56
	$w = t_g$	84.89	22.61	17.66
	$w = {t_g}^2$	86.4	22.88	17.79
	$w = {t_g}^4$	103.98	23.94	18.77
Bimodal	$w = 1$	133.47	33.95	23.15
	$w = {t_g}^{0.5}$	137.41	32.64	23.26
	$w = t_g$	140.79	31.49	23.53
	$w = {t_g}^2$	146.07	30.5	23.78
	$w = {t_g}^4$	152.88	31.42	25.94

TABLE II
RESULT FROM WLSE FOR 14-NODE NETWORK

traffic and highest for bimodal distributed point to point traffic in the tomogravity and entropy maximization methods. We can see from Table III that the tomogravity and the entropy maximization methods do not differ significantly in average percentage of error in the estimation of point to point traffic. The average of absolute percentage error is 22.68% for Poisson distribution in linear programming which is comparable to results from *w*lse and *p*dsco tools. For all other point to point traffic distributions the results are worse for linear programming.

Later we provide the model with partial known point to point traffic data and then assess by how much the prediction of the unknown ones can be estimated. In Table IV, we present the result of tomogravity method with 13 (of the 182) known values for point to point traffic from node 0 to other nodes of the network of Figure 2 with *wlse*, *pdsco* and linear programming. We can see from Tables III and IV that there is some improvement in the estimation if some point to point traffic are directly measured. There is approximately 2% decrease in the average percentage error for knowing the values of point to point traffic from a node to other nodes in the network of Figure 2 for *wlse*. We can see from Tables III and IV that with some known point to point traffic average percentage error decreases in case of uniform, Poisson and bimodal point to point traffic for entropy maximization method. However, the average percentage error increases for constant and normally distributed point to point traffic in entropy maximization method. In entropy maximization method Equation 5 is used for minimizing mutual information and in case of constant and normally distributed traffic the minimization technique of *pdsco* stuck in a local minima which increase the average percentage error in estimation of point to point traffic. We can see from Tables III and IV that with some known point to point traffic average percentage error does not decrease for linear programming. This is also counter-intuitive like *p*dsco for some known OD traffic.

We augment the $AX = Y$ with the constraints for each node that the summation of its incoming (outgoing) traffic from other nodes is equal to its total incoming (outgoing) traffic for the entropy maximization method. We present the result of entropy maximization and the modified entropy maximization with added extra constraints in Table V for the network of Figure 2. We have found as expected that the addition of extra constraints reduces error in inferring OD traffic.

VII. CONCLUSION

In this paper, different methods of traffic matrix inference from link count data are described. We can see that tomogravity [8] and maximization of entropy [9], [10] are better methods, than linear programming, for inferring traffic matrix. We have carried out experiments of gravity method using *wlse* and entropy maximization with *pdsco* [5] and *maxent* [3] and linear programming [2] and currently these methods are being assessed in real traffic testbed. From the preliminary experiments we can comment that *wlse* and *pdsco* are better than *maxent* and linear programming in estimating OD traffic. We also find that *wlse* and *pdsco* do not differ much in average percentage error in estimating OD traffic. We have carried out experiment with the same network topology using different traffic for all the OD pairs. We have found as expected that change in amount of traffic in OD pair affects accuracy of calculation. Secondly we found that knowing some point to point traffic reduces the error in estimating the traffic matrix for *wlse* method. However, the error does not always decrease for *pdsco* when some point to point traffic are known. Tomogravity method using *wlse* results in different amount of error for different networks with different traffic distributions for different weight assignments. From these experiments we conclude that *wlse* is the best among the three existing tools we have used for estimating point to point

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

Distribution	Existing Entropy Maximization			Modified Technique		
	Max. Err.	Avg. Err.	Std. Dev.	Max. Err.	Avg. Err.	Std. Dev.
Constant	4.66	1.92	1.37	0.039	0.013	0.007
Uniform	98.95	21.87	16.53	90.292	21.044	16.246
Poisson	83.58	21.1	15.2	63.687	21.206	14.884
Normal	84.76	22.59	17.55	82.998	21.39	16.6
Bimodal	136.5	32.53	23.26	125.583	29.665	23.544

TABLE V

ESTIMATION OF OD TRAFFIC FOR 14-NODE NETWORK WITH ADDED CONSTRAINTS FOR ENTROPY MXIMIZATION

traffic. We add extra constraints for entropy maximization and our modified entropy maximization method outperforms both existing entropy maximization and tomogravity methods.

Our next direction for tomogravity model is to determine two different gravity model solutions using $f_{ij} = \sum_j A_j$ and $f_{ij} = \sum_i R_i$ and then using tomography steps to infer a point on $AX = Y$ hyper-plane which minimizes the summation of the square of distances from both the points estimated. We can augment the $AX = Y$ with the constraints for each node that the summation of its incoming (outgoing) traffic from other nodes is equal to its total incoming (outgoing) traffic for tomogravity.

Acknowledgements: This research is partially supported by grants from NSERC to A.S. Alfa.

REFERENCES

[1] J. Cao, D. Davis, S. V. Wiel, and B. Yu. Time-varying network tomography. *American Statistical Association*, 95:1063–1075, 2000.

[2] A. E. Conway and M. Li. Estimating the traffic distribution matrix of a packet-switched network using aggregate traffic measurements. *Telecommunication Systems*, 24(1):9–28, 2003.

[3] P. C. Hansen. Regularization tools (for matlab). http://www.imm.dtu.dk/ pch/Regutools/index.html.

[4] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot. Traffic matrix estimation: Existing techniques and new directions. In *SIGCOMM '02: Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 161–174, Pittsburg, USA, August 2002.

[5] M. A. Saunders and B. Kim. Pdsco: Primal-dual interior method for separable convex objectives. http://www.stanford.edu/group/SOL/software/pdsco.

[6] C. Tebaldi and M. West. Bayesian inference on network traffic using link count data. *American Statistical Association*, 93:557–576, 1998.

[7] Y. Vardi. Network tomography: estimating source-destination traffic intensities from link data. *American Statistical Association*, 91:365–377, 1996.

[8] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg. Fast accurate computation of large-scale ip traffic matrices from link loads. *ACM SIGMETRICS*, 31(1):206–217, 2003.

[9] Y. Zhang, M. Roughan, C. Lund, and D. Donoho. An information theoretic approach to traffic matrix estimation. In *SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 301–312, Karlsruhe, Germany, August 2003.

[10] Y. Zhang, M. Roughan, C. Lund, and D. Donoho. Estimating point-to-point and point-to-multipoint traffic matrices: an information-theoretic approach. *IEEE/ACM Transactions on Networking*, 13(5):947–960, 2005.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.