Mixed Poisson Traffic Rate Network Tomography

Yariv Ephraim, Joshua Coblenz, Brian L. Mark
Dept. of Electrical and Computer Engineering
George Mason University
Fairfax, VA 22030, U.S.A.
{ yephraim, jcoblen2, bmark }@gmu.edu

Hanoch Lev-Ari
Dept. of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115, U.S.A.
levari@ece.neu.edu

***Abstract*—We extend network tomography to traffic flows that are not necessarily Poisson random processes. This assumption has governed the field since its inception in 1996 by Y. Vardi. We allow the distribution of the packet count of each traffic flow in a given time interval to be a mixture of Poisson random variables. Both discrete as well as continuous mixtures are studied. For the latter case, we focus on mixed Poisson distributions with Gamma mixing distribution. As is well known, this mixed Poisson distribution is the negative binomial distribution. Other mixing distributions, such as Wald or the inverse Gaussian distribution can be used. Mixture distributions are overdispersed with variance larger than the mean. Thus, they are more suitable for Internet traffic than the Poisson model. We develop a second-order moment matching approach for estimating the mean traffic rate for each source-destination pair using least squares and the minimum I-divergence iterative procedure. We demonstrate the performance of the proposed approach by several numerical examples. The results show that the averaged normalized mean squared error in rate estimation is of the same order as in the classic Poisson based network tomography. Furthermore, no degradation in performance was observed when traffic rates are Poisson but Poisson mixtures are assumed.**

***Index Terms*—Network traffic, network tomography, inverse problem, mixed Poisson distribution.**

I. Introduction

In network tomography the rates of traffic flows on source-destination pairs are estimated from traffic flows over links of the networks. Traffic flow rates are measured, for example, by the number of packets or messages per second. Traffic flows are commonly assumed to be independent Poisson random processes. This network tomography problem was first formulated by Vardi [1]. A related formulation in which the rates of independent Poisson link traffic flows are estimated from input/output traffic flows was studied earlier by Vanderbei and Iannone [2]. The two formulations lead to the same set of underdetermined linear equations for the two inverse problems.

We focus here on Vardi's formulation as applied to networks operating under a deterministic routing regime and where link measurements are passive and require no probes. Under this regime, a fixed path is assigned to a traffic flow associated with each source-destination pair. The traffic flow over a link is the superposition of all source-destination traffic flows that share that link. Thus, link traffic flows are also Poisson but are not independent. The dependence of link traffic flows renders maximum likelihood estimation of the rates impractical. Instead, moment matching rate estimation has dominated the field. A thorough Bayesian approach to Vardi's problem was developed by Tebaldi and West [3]. Other closely related work to Vardi's approach appeared in [4], [5], [6] where maximum likelihood estimation of the source-destination rates from link data was implemented under a Gaussian rather than a Poisson traffic model.

The Poisson assumption simplifies the network tomography problem but it is considered unrealistic since, for example, it does not account for bursty traffic. In this paper we take a first step in relaxing this assumption and assume that the distribution of each source-destination traffic flow count is a mixture of Poisson distributions. Mixture distributions are overdispersed with variance larger than their mean. For Poisson traffic flows, the mean equals the variance. In this paper, both continuous and discrete mixtures are studied. For the continuous mixture we assume a Gamma mixing distribution. Thus, the distribution of the source-destination traffic flow count becomes negative binomial, see, e.g., [7]. A negative binomial road traffic flow model was advocated in [8] irrespective of any mixture model. A Bayesian approach for characterization of transportation origin-destination matrices using Poisson mixtures was developed in [9]. The goal of network tomography under mixed Poisson traffic flows is to estimate the *mean* source-destination traffic rate. We use second-order moment matching to estimate the mean traffic rate for each source-destination pair of interest. The least squares solution and the minimum I-divergence iterative solution of the moment matching equations are studied in this work [10], [11], [12], [1].

The plan for this paper is as follows. In Section II we present network tomography under the continuous mixed Poisson traffic flow count distribution. In Section III we discuss network tomography for discrete Poisson mixture traffic flows. We present our numerical results in Section IV. Concluding remarks are given in Section V.

II. Continuous Mixture of Poisson Distributions

In this section we present our moment matching approach when source-destination traffic flow counts are assumed to be characterized according to a continuous mixture of Poisson distributions.

This work was supported in part by the U.S. National Science Foundation under Grant CCF-1717033.

Let K be a scalar random variable with conditional Poisson probability mass function (pmf) given its mean $E\{K\} = \lambda$. Suppose that λ is a realization of a Gamma-distributed random variable Λ with shape parameter α and scale parameter $1/\rho$. That is, the conditional pmf of K is given by

$$p_{K|\Lambda}(k \mid \lambda) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots, \tag{1}$$

and the Gamma probability density function (pdf) is given by

$$f_\Lambda(\lambda) = \frac{\rho e^{-\rho\lambda}(\rho\lambda)^{\alpha-1}}{\Gamma(\alpha)} \tag{2}$$

where

$$\Gamma(\alpha) = \int_0^\infty e^{-\lambda}\lambda^{\alpha-1}d\lambda, \tag{3}$$

and it satisfies $\Gamma(\alpha) = \alpha\Gamma(\alpha-1)$, $\alpha \geq 1$, $\Gamma(0) = 1$. We also have

$$E\{\Lambda\} = \frac{\alpha}{\rho} \quad \text{and} \quad \text{var}(\Lambda) = \frac{\alpha}{\rho^2}. \tag{4}$$

We denote $\bar{\Lambda} = E\{\Lambda\}$. It is well known that the unconditional pmf of K is negative binomial with parameter (β, r) where

$$\beta = \frac{\rho}{1+\rho} \quad \text{and} \quad r = \alpha. \tag{5}$$

The negative binomial pmf is given by

$$\mathbf{P}(K = k) = \binom{k+r-1}{r-1}\beta^r(1-\beta)^k. \tag{6}$$

From (4) and (5),

$$\bar{\Lambda} = r\frac{1-\beta}{\beta}. \tag{7}$$

This is the quantity of interest in network tomography. It represents the *average* rate of a source-destination traffic under mixed Poisson regime with Gamma prior.

Define a scalar random variable $U = K + r$. The pmf of U is

$$\mathbf{P}(U = n) = \binom{n-1}{r-1}\beta^r(1-\beta)^{n-r}, n = r, r+1, \ldots. \tag{8}$$

The random variable U represents the time of occurrence of the rth success in a sequence of independent Bernoulli trials with probability of success β. For this random variable,

$$\begin{aligned} \xi_u &:= E\{U\} = \frac{r}{\beta} \\ \eta_u &:= \text{var}(U) = \frac{r}{\beta}\frac{1-\beta}{\beta}. \end{aligned} \tag{9}$$

Note that $\eta_u > \xi_u$ when $\beta < .5$. For this reason, the mixed Poisson model is refereed to as "overdispersed." In the Poisson model $\eta_u = \xi_u$. From (9) and (7),

$$\beta = \frac{\xi_u}{\xi_u + \eta_u} \tag{10}$$

and

$$\bar{\Lambda} = \frac{\xi_u \eta_u}{\xi_u + \eta_u}. \tag{11}$$

This expression allows us to estimate $\bar{\Lambda}$ from estimates of ξ_u and η_u.

Returning to the network tomography problem, consider a network with L source-destination pairs, M links, and a routing matrix $A = \{a_{ij}\}$ where $a_{ij} \in \{0, 1\}$, $i = 1, \ldots, M$, and $j = 1, \ldots, L$. Let X_j denote traffic flow over source-destination pair j, and assume that $\{X_1, \ldots, X_L\}$ are statistically independent mixed Poisson random variables with independent Gamma priors given by $\{f_{\Lambda_1}(\cdot), \ldots, f_{\Lambda_L}(\cdot)\}$, respectively. Thus, $\{X_1, \ldots, X_L\}$ are statistically independent negative binomial random variables. The parameter of X_j is defined similarly to the parameter of U and is denoted by (β_j, r_j). The mean rate of traffic flow on the jth source-destination pair is denoted by $\bar{\Lambda}_j = E\{\Lambda_j\}$. Let $\{Y_1, \ldots, Y_M\}$ denote traffic flows over the M links. We collectively describe source-destination traffic flows as an $L \times 1$ vector X and the link traffic flows as an $M \times 1$ vector Y. We have $Y = AX$.

Let $\xi = E\{X\}$ denote the mean of X. Let vec$[\cdot]$ denote vectorization by rows of a matrix. Let

$$\begin{aligned} \eta &= \text{vec}\left[E\{(X-\xi)(X-\xi)'\}\right] \\ &= E\{(X-\xi)\otimes(X-\xi)\} \end{aligned} \tag{12}$$

denote the vectorized covariance of X where $'$ denotes vector transpose and $\otimes$ denotes the Kronecker product. For the second line of (12) we have used the identity vec$[uv'] = v \otimes u$, which holds for any two column vectors u and v. Let $\mu_1(Y)$ and $\mu_2(Y)$ denote, respectively, the empirical mean and the vectorized empirical covariance of the centralized vector Y. These statistics are estimated from N realizations $\{y(1), \ldots, y(N)\}$ of the vector Y as follows:

$$\begin{aligned} \mu_1(Y) &= \frac{1}{N}\sum_{n=1}^{N} y(n) \\ \mu_2(Y) &= \frac{1}{N}\sum_{n=1}^{N}(y(n)-\mu_1(Y))\otimes(y(n)-\mu_1(Y)). \end{aligned} \tag{13}$$

Following the moment matching approach in [1], [13], we seek $\{\xi, \eta\}$ that satisfies the following moment matching equation:

$$\begin{pmatrix}\mu_1(Y)\\ \mu_2(Y)\end{pmatrix} = \begin{bmatrix} A & 0 \\ 0 & A\odot A\end{bmatrix}\begin{pmatrix}\xi \\ \eta\end{pmatrix} \tag{14}$$

where $\odot$ denotes a Khatri-Rao product and 0 denotes a null matrix of suitable dimensions. The RHS of (14) equals the concatenated mean and vectorized covariance of Y. Let $A = [a_1, \ldots, a_L]$ where $\{a_j\}$ denote the columns of A. The Khatri-Rao product is defined by

$$A \odot A := [a_1 \otimes a_1, a_2 \otimes a_2, \ldots, a_L \otimes a_L]. \tag{15}$$

Let $\hat{\xi}$ and $\hat{\eta}$ denote, respectively, estimates of ξ and η as obtained from (14). Using (11), we can estimate the mean rate for source-destination j from $\hat{\xi}$ and $\hat{\eta}$ as follows:

$$\hat{\bar{\Lambda}}_j = \frac{\hat{\xi}_j \hat{\eta}_j}{\hat{\xi}_j + \hat{\eta}_j}, \quad j = 1, \ldots, L. \tag{16}$$

Following [13], implementation of (14) requires removal of null and duplicate rows in $A \odot A$. Such rows may exist due to the symmetry of the covariance matrix and since A is a zero-one matrix. Also, rows with negative empirical covariance estimates in (13) ought to be removed. We denote the resulting row-reduced matrix in (14) by $\mathcal{A}_2$, and the row-reduced vector on the LHS of (14) by $\hat{\psi}(Y)$. Letting $\zeta = \text{col}\{\xi, \eta\}$, the moment matching matrix equation (14) becomes

$$\hat{\psi}_2(Y) = \mathcal{A}_2 \zeta. \tag{17}$$

We consider two approaches to estimate (ξ, η) from (17). These are the least squares approach and the minimum I-divergence iterative approach [11], [12], [1], [10]. Least squares estimation of ζ is accomplished here by applying Tikhonov's regularized least squares solution to the linear equation (17). The unique Tikhonov regularized least squares solution for the possibly inconsistent set of equations (14), when the matrix containing $\mathcal{A}_2$ is not necessarily full column rank is given by [14, p. 51]

$$\hat{\zeta} = (\mathcal{A}_2^* \mathcal{A}_2 + \gamma I)^{-1} \mathcal{A}_2^* \ \hat{\psi}_2(Y) \tag{18}$$

for some $\gamma > 0$. Note that the regularized estimator applies to a skinny as well as a fat matrix $\mathcal{A}_2$.

The least squares estimates of the components of $\text{col}\{\xi, \eta\}$ are not guaranteed to be non-negative. Hence negative rate estimates are possible with the least squares estimator. Non-negative estimates can be obtained by using non-negative least squares optimization [15]. This approach did not lead to good results in our earlier work [13]. Instead, we have substituted negative estimates in our numeric examples with the value of 0.001. This approach resulted in [13] in substantially lower MSE compared to using the constrained optimization algorithm of [15, p. 161]. The performance of the algorithm should not be affected by this substitution since usually negative estimates are rare at our working point.

An alternative approach is to use the minimum I-divergence estimator of ζ, which guarantees non-negative estimates of the mean rate. Suppose that there are M_a equations in (17). Let ζ_j^{old} denote a current estimate of the jth component of ζ, and let ζ_j^{new} denote the new estimate of that component at the conclusion of the iteration. Let $\mathcal{A}_2 = \{b_{ij}, i = 1, \ldots, M_a; j = 1, \ldots, 2L\}$. Let $(\mathcal{A}_2 \zeta^{\text{old}})_i$ denote the ith component of $\mathcal{A}_2 \zeta^{\text{old}}$. Similarly, let $\hat{\psi}_i(Y)$ denote the ith component of $\hat{\psi}(Y)$. The iteration is given by

$$\zeta_j^{\text{new}} = \zeta_j^{\text{old}} \sum_{i=1}^{M_a} \bar{b}_{ij} \frac{\hat{\psi}_i(Y)}{(\mathcal{A}_2 \zeta^{\text{old}})_i} \text{ where } \bar{b}_{ij} := \frac{b_{ij}}{\sum_{t=1}^{M_a} b_{tj}}, \tag{19}$$

for $j = 1, \ldots, L$. This iteration was shown to converge monotonically to the minimizer of Csiszár's I-divergence [10] between the two sides of (17) [12].

The computational effort in the moment matching approach consists of the effort to construct and solve the set of equations (17). The number of equations in this set is denoted by $n_2(M)$ and it satisfies $n_2(M) \ll M + M^2$. Construction of the right hand side of (17) requires $M^2 L$ operations. Construction of the left hand side of (17) requires $M^2 N$ operations where N is the number of vectors used to estimate the empirical moments. Solving the equations requires effort that depends only on $n_2(M)$ and L. The combined effort is dominated by $M^2 N$ since N must be large to produce meaningful empirical moment estimates. Thus the computational effort of the mean rate estimation approach is approximately linear in N when N is large, which is always the case.

III. Discrete Mixtures of Poisson Distributions

Let X_j be a scalar random variable with a discrete Poisson mixture pmf on the set of non-negative integers. Let Z_j denote the random index of the mixture component which takes values in $\{1, 2, \ldots, \kappa\}$ with probabilities $\{\beta_{j1}, \ldots, \beta_{j\kappa}\}$, respectively. Thus, $X_j \mid Z_j = i$ is a Poisson random variable with mean λ_{ji}. The pmf of X_j is given by

$$p_{X_j}(k) = \sum_{i=1}^{\kappa} \beta_{ji} \ p_{X_j|Z_j}(k \mid i) = \sum_{i=1}^{\kappa} \beta_{ji} e^{-\lambda_{ji}} \frac{\lambda_{ji}^k}{k!}. \tag{20}$$

The mean of X_j is given by

$$\xi_j := E\{X_j\} = \sum_{i=1}^{\kappa} \beta_{ji} \lambda_{ji} \tag{21}$$

and the variance follows from the total variance theorem and is given by

$$\begin{aligned} \eta_j &:= \text{var}(E\{X_j \mid Z_j\}) + E\{\text{var}(X_j \mid Z_j)\} \\ &= \sum_{i=1}^{\kappa} \beta_{ji} \lambda_{ji}^2 - \xi_j^2 + \xi_j. \end{aligned} \tag{22}$$

Letting

$$\upsilon_j = \sum_{i=1}^{\kappa} \beta_{ji} \lambda_{ji}^2 - \xi_j^2, \tag{23}$$

we can rewrite (22) as

$$\eta_j = \upsilon_j + \xi_j. \tag{24}$$

Proceeding as in the continuous mixture case, we define a column vector X of source-destination traffic flows whose jth component X_j is defined as above. The mean rate vector is $\xi = E\{X\}$. We also define the vector Y of link traffic flows and invoke the relation $Y = AX$. Finally, we define a vector υ whose jth component is given by υ_j. Then, the moment matching equation becomes

$$\begin{pmatrix} \mu_1(Y) \\ \mu_2(Y) \end{pmatrix} = \begin{bmatrix} A & 0 \\ A \odot A & A \odot A \end{bmatrix} \begin{pmatrix} \xi \\ \upsilon \end{pmatrix}. \tag{25}$$

The parameter of interest in this case is the mean rate vector ξ.

The moment matching equation (25) reduces to the moment matching equation in [13] when traffic flows are Poisson rather than mixtures of Poisson distributions. In [13] the rate vector ξ satisfies

$$\begin{pmatrix} \mu_1(Y) \\ \mu_2(Y) \end{pmatrix} = \begin{pmatrix} A \\ A \odot A \end{pmatrix} \xi. \tag{26}$$

Eq. (25) coincides with this equation when $\upsilon = 0$. We will elaborate further on this important point in the next section. Equations (25) and (26) undergo row reduction as in Section II.

IV. Numerical Examples

In this section we demonstrate the performance of the mean rate estimators for negative binomial sources and for discrete Poisson mixture sources. The proposed approach was applied to the NSFnet [16] whose topology is shown in Fig. 1. The network consists of 14 nodes and 21 bidirectional links. Hence, it contains $L = 14 \cdot 13/2 = 91$ source-destination pairs. A network of this size may represent a private network, a transportation network or a subnetwork of interest of a large-scale network. The link weights in Fig. 1 may be used to determine $k \geq 1$ shortest paths for each source-destination pair [17]. Otherwise, they play no role in the traffic rate estimation problem. We use here $k = 1$ and hence the number of source-destination paths equals the number of source-destination pairs.

For the continuous Poisson mixture traffic flows we have estimated the mean source-destination rates $\{\bar{\Lambda}_1, \ldots, \bar{\Lambda}_L\}$ in $T = 500$ independent runs. In each run we have generated a sequence of L real random numbers uniformly distributed on $[0, 4]$. This sequence constitutes the L source-destination mean rates. We also generated a sequence of L random integers $\{r_j, j = 1, \ldots, L\}$ uniformly distributed on $[2, 6]$ where each represents the time of the rth success in a source-destination pair. Using these sequences, we obtained from (7) the parameter β_j of the negative binomial pmf (8) for the jth source-destination pair as follows:

$$\beta_j = \frac{r_j}{r_j + \bar{\Lambda}_j}. \tag{27}$$

Now using L pairs $\{(r_j, \beta_j), j = 1, \ldots, L\}$ of the parameters of the negative binomial traffic flows we generated N statistically independent identically distributed vectors $\{X\}$ which were subsequently transformed into the vectors $\{Y = AX\}$ using the assumed known routing matrix A. These N statistically independent identically distributed link traffic vectors were used in (13) to generate the empirical moments. We experimented with $N = 5000$, $N = 10000$ and $N = 20000$ vectors. The empirical moments were used in (18) for least squares estimation, and in (19) for minimum I-divergence rate estimation. The estimates, say, $\{\hat{\xi}, \hat{\eta}\}$ were subsequently used in (16) and yielded mean rate estimates $\{\hat{\bar{\Lambda}}_j\}$ for the L source-destination pairs. The regularization parameter in (18) was set to $\gamma = .0005$ in this work. The iteration was initialized uniformly with all rates set to .1. It was terminated after 300 iterations.

The mean rate estimate $\hat{\bar{\Lambda}}_j$ was contrasted with the true mean rate $\bar{\Lambda}_j$ as follows. We denote by $\bar{\Lambda}_j(t)$ and $\hat{\bar{\Lambda}}_j(t)$, respectively, the mean rate of the jth source-destination pair and its estimate at the tth run where $j = 1, \ldots, L$ and $t = 1, \ldots, T$. For each estimate we evaluated the *normalized MSE* defined by

$$\nu_j^2 = \frac{\frac{1}{T}\sum_{t=1}^{T}(\bar{\Lambda}_j(t) - \hat{\bar{\Lambda}}_j(t))^2}{\frac{1}{T}\sum_{t=1}^{T}(\bar{\Lambda}_j(t))^2} \tag{28}$$

and the *averaged normalized MSE* defined by

$$\overline{\nu^2} = \frac{1}{L}\sum_{j=1}^{L}\nu_j^2. \tag{29}$$

The MSE in estimating $\bar{\Lambda}_j$ given by the numerator of (28) is approximately

$$\nu_j^2 \frac{1}{T}\sum_{t=1}^{T}(\bar{\Lambda}_j(t))^2 \approx \nu_j^2 \cdot E\{\bar{\Lambda}_j^2\} \tag{30}$$

when T is sufficiently large. A normalized MSE is appropriate here since new mean rates are used in each run.

A similar evaluation procedure was applied to the discrete Poisson mixture traffic flow. The least squares estimator (18) and the minimum I-divergence (19) were applied to the row-reduced version of (25) where now the mean rate estimates are given by the components of $\hat{\xi} = \{\hat{\xi}_1, \ldots, \hat{\xi}_L\}$ and $\hat{\xi}$ constitutes the first L components of $\hat{\zeta}$. We demonstrate the results for a mixture of four Poisson distributions with mixing probabilities drawn randomly from $[0, 1]$ and rates drawn randomly from $[0, 4]$ in each of the 500 runs. We also examined a mismatched scenario in which the mean rate of a Poisson source-destination traffic flow is estimated by the approach outlined in this paper under the false assumption that it actually is a discrete Poisson mixture traffic flow. This is an important case since we do not want to sacrifice performance when the actual distribution of the traffic flow is indeed Poisson.

We next present our numerical results. Table I provides the averaged normalized MSE $\overline{\nu^2}$ for the continuous mixed Poisson as a function of N for the least squares estimator (18) in conjunction with (16). The table also shows the percentage of negative least squares mean rate estimates. Table II provides similar results for the minimum I-divergence iterative approach. Tables III and IV present the results for the discrete

	$N = 5000$	$N = 10000$	$N = 20000$
Ave. Nor. MSE $\overline{\nu^2}$:	0.0841	0.0750	0.0703
Percent Neg. Est.:	4.3648	3.2066	2.3868

TABLE I
CONTINUOUS MIXED POISSON TRAFFIC FLOWS WITH MEAN RATE $\{\bar{\Lambda}_j\}$ ESTIMATED BY THE LEAST SQUARES ESTIMATOR (18) AND (16).

	$N = 5000$	$N = 10000$	$N = 20000$
Ave. Nor. MSE $\overline{\nu^2}$:	0.0362	0.0263	0.0214
Percent Neg. Est.:	0	0	0

TABLE II
CONTINUOUS MIXED POISSON TRAFFIC FLOWS WITH MEAN RATE $\{\bar{\Lambda}_j\}$ ESTIMATED ITERATIVELY USING (19) AND (16).

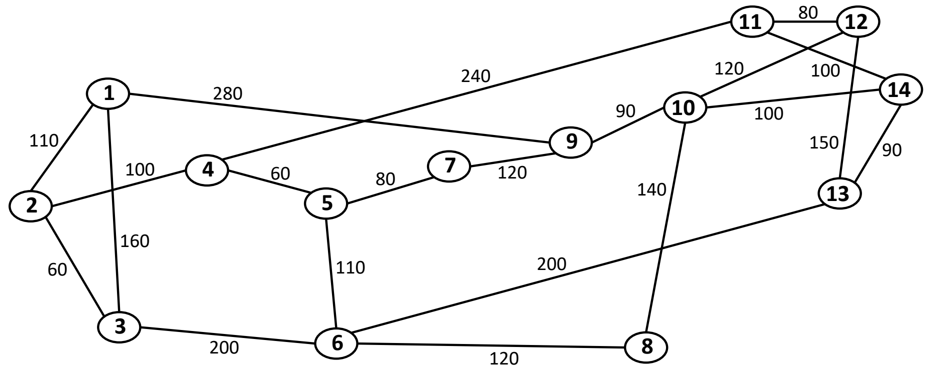

Fig. 1. NSFnet topology with link weights as in [16, Fig. 4].

	$N = 5000$	$N = 10000$	$N = 20000$
Ave. Nor. MSE $\overline{\nu^2}$:	0.0917	0.0828	0.0782
Percent Neg. Est.:	10.6484	9.3648	8.7451

TABLE III
DISCRETE (4) MIXED POISSON TRAFFIC FLOWS WITH MEAN RATE $\{\bar{\Lambda}_j\}$ ESTIMATED BY THE LEAST SQUARES ESTIMATOR (18).

	$N = 5000$	$N = 10000$	$N = 20000$
Ave. Nor. MSE $\overline{\nu^2}$:	0.1051	0.0903	0.0826
Percent Neg. Est.:	0	0	0

TABLE IV
DISCRETE (4) MIXED POISSON TRAFFIC FLOWS WITH MEAN RATE $\{\bar{\Lambda}_j\}$ ESTIMATED ITERATIVELY USING (19).

	$N = 5000$	$N = 10000$	$N = 20000$
Ave. Nor. MSE $\overline{\nu^2}$ (25):	0.0819	0.0771	0.0750
Ave. Nor. MSE $\overline{\nu^2}$ (26):	0.0380	0.0211	0.0109
Percent Neg. Est. in (25):	44.7956	44.4681	44.0945
Percent Neg. Est. in (26):	3.0813	2.1868	1.5868

TABLE V
AVERAGED NORMALIZED MSE $\overline{\nu^2}$ AND PERCENT OF NEGATIVE ESTIMATES RESULTING FROM LEAST SQUARES ESTIMATION AS APPLIED TO (25) AND (26) FOR POISSON TRAFFIC FLOWS.

	$N = 5000$	$N = 10000$	$N = 20000$
Ave. Nor. MSE $\overline{\nu^2}$ (25):	0.0331	0.0202	0.0138
Ave. Nor. MSE $\overline{\nu^2}$ (26):	0.0363	0.0208	0.0111
Percent Neg. Est. in (25):	0	0	0
Percent Neg. Est. in (26):	0	0	0

TABLE VI
AVERAGED NORMALIZED MSE $\overline{\nu^2}$ AND PERCENT OF NEGATIVE ESTIMATES RESULTING FROM APPLICATION OF THE ITERATION (19) TO (25) AND (26) FOR POISSON TRAFFIC FLOWS.

Poisson mixture traffic flows in the same format as in Tables I and II, respectively.

The numerical results in Tables I and II for the continuous mixed traffic flows show that the iterative approach outperforms the least squares approach with its significantly lower averaged normalized MSE. Furthermore, the rate estimates in the iterative approach are always non-negative. The results in Tables III and IV show for the discrete Poisson mixtures that the two estimation approaches provide similar averaged normalized MSE while the iterative approach is guaranteed to provide non-negative rate estimates. In all cases, the performance improves as N increases. This improves the accuracy of the empirical estimates and results in lower averaged normalized MSE and less frequent negative estimates by the least squares estimator.

The next two tables examine the performance of the proposed approach when applied to Poisson traffic flows. The studies of network tomography to date have exclusively focused on Poisson traffic flows following the seminal work of Vardi [1]. This case corresponds to a single component Poisson mixture. The moment matching equation for a single component Poisson mixture is given by (26) while the moment matching equation for the Poisson mixture is given in (25). Here we compare the rate estimates of the Poisson traffic flows as obtained from each of the two equations. This comparison will show if there is a degradation in performance when the data correspond to Poisson traffic flows but the estimator is geared towards mixed Poisson traffic flows. In Table V the rate estimation is based on least squares while in Table VI the iterative rate estimation from (19) is used. From Tables V and VI it is evident that the iterative approach of (19) is superior to the least squares approach in this mismatched estimation problem. It provides significantly lower averaged normalized MSE and all rate estimates are non-negative. Furthermore, the performance of the iterative approach is close to that of the matched approach when the linear equations (26) that are suitable for the input Poisson traffic flow counts are used. Thus, in summary, we conclude from Tables I-VI that the iterative estimator (19) performs better than the least squares estimator.

V. CONCLUDING REMARKS

We have extended network tomography to include non-Poissonian traffic flows in the forms of continuous and discrete mixtures of Poisson distributions. To date, all work on network tomography has assumed Poisson traffic flows. We have developed a second-order moment matching approach for estimating the mean rate of each source-destination traffic flow. The specific continuous mixture we studied was the mixed Poisson with Gamma mixing distribution. This model yields negative binomial traffic flows. The discrete Poisson mixture comprises a mixture of several Poisson distributions. We have employed two approaches to estimate the mean rate from the moment matching equations. We used a least squares approach and the well known minimum I-divergence iteration [11], [1], [12], [10]. The approaches were numerically studied and compared. We have demonstrated the superiority of the minimum I-divergence iteration over the least squares approach. We have also examined network tomography under a mismatched condition where the traffic flows in the networks were Poisson but network tomography for mixtures of Poisson distributions was applied. We have demonstrated that there is essentially no loss in performance in this mismatch and hence network tomography can only become more accurate if indeed traffic flows do not obey the Poisson law. While we have focused on networks with deterministic routing, the approach is applicable to networks with random routing using the "super-network" approach of Tebaldi and West [3].

The key property that makes the mixed Poisson distribution useful is its overdispersion, meaning that its variance is larger than its mean. This property may be enhanced by using mixing distributions other than the Gamma. A commonly used mixing distribution is the inverse Gaussian or the Wald distribution given by

$$f_\Lambda(\lambda) = \frac{1}{(2\pi\sigma^2\lambda^3)^{1/2}} e^{-\frac{(\lambda-1)^2}{2\sigma^2\lambda}}, \quad \lambda > 0, \tag{31}$$

where $\sigma^2 = \text{var}(\Lambda)$. The mixed Poisson pmf in this case is given by

$$p_{X_1}(k) = \int_0^\infty e^{-\lambda\mu} \frac{(\lambda\mu)^k}{k!} f_\Lambda(\lambda) d\lambda \tag{32}$$

where the conditional mean of the Poisson distribution is $\lambda\mu$. It follows that [18]

$$E\{X_1\} = \mu \quad \text{and} \quad \text{var}(X_1) = \mu + \sigma^2\mu^2. \tag{33}$$

Furthermore, μ and σ^2 can be made functions of a regression parameter. This model provides heavier tails than the negative binomial mixed Poisson. It was found particularly useful in characterization of transportation origin-destination matrices in [9]. Maximum likelihood estimation of the mixed Poisson inverse Gaussian regression parameters was developed in [18] where $\{X\}$ is observable. This is not the case here since we only observe $Y = AX$. Hence application of this model would require moment matching as was done in this paper for the negative binomial mixed Poisson model. Other examples of mixed Poisson distributions may be found in [19].

REFERENCES

[1] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *J. Amer. Stat. Assoc.*, vol. 91, no. 433, pp. 365–377, 1996.

[2] R. J. Vanderbei and J. Iannone, "An EM approach to OD matrix estimation," Princeton University, Tech. Rep. SOR 94-04, 1994.

[3] C. Tebaldi and M. West, "Bayesian inference on network traffic using link count data (with discussion)," *J. Amer. Stat. Assoc.*, vol. 93, no. 442, pp. 557–576, 1998.

[4] J. Cao, D. Davis, S. Vander Wiel, and B. Yu, "Time-varying network tomography: Router link data," *J. Amer. Statist. Assoc.*, vol. 95, pp. 1063–1075, 2004.

[5] G. Liang and B. Yu, "Maximum pseudo likelihood estimation in network tomography," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2043–2053, Aug. 2003.

[6] R. Castro, M. Coates, G. Liang, R. Nowak, and B. Yu, "Network tomography: Recent developments," *J. Stat. Sci.*, vol. 19, no. 3, pp. 499–517, 2004.

[7] J. Grandell, *Mixed Poisson Processes*, ser. CRC: Statistics and Mathematics. New York: Chapman and Hall, 1997.

[8] M. L. Hazelton, "Network tomography for integer-valued traffic," *The Annals of Applied Statistics*, vol. 9, no. 1, pp. 474–506, 2015.

[9] K. Perrakis, D. Karlis, M. Cools, and D. Janssens, "Bayesian inference for transportation origin-destination matrices: the Poisson-inverse Gaussian and other Poisson mixtures," *J. Royal Stat. Society. Series A*, vol. 178, no. 1, pp. 271–296, Jan 2015.

[10] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *The Annals of Statistics*, vol. 19, no. 4, pp. 2032–2066, Dec. 1991.

[11] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, pp. 55–59, Jan 1972.

[12] D. L. Snyder, T. J. Schulz, and J. A. O'Sullivan, "Deblurring subject to nonnegativity constraints," *IEEE Trans. Sig. Proc.*, vol. 40, 1992.

[13] H. Lev-Ari, Y. Ephraim, and B. L. Mark, "Traffic rate network tomography with higher-order cumulants," arXiv, Tech. Rep. arXiv:2012.07224 [cs.PF], Dec. 2020. [Online]. Available: https://arxiv.org/abs/2012.07224

[14] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, 1st ed. New Jersey: Prentice Hall, 2000.

[15] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. New Jersey: Prentice-Hall, 1974.

[16] L. H. Bonani, "Modeling an optical network operating with hybrid-switching paradigms," *J. Microw. Optoelectron. Electromagn. Appl.*, vol. 15, no. 4, pp. 275–292, 2016.

[17] J. Y. Yen, "Finding the k shortest loopless paths in a network," *Management Science*, vol. 17, no. 11, pp. 712–716, Jul. 1971.

[18] C. Dean, J. F. Lawless, and G. E. Willmot, "A mixed Poisson-inverse-Gaussian regression model," *The Canadian Journal of Statistics*, vol. 17, no. 2, 1989.

[19] S. Laurent, "Some Poisson mixtures distributions with a hyperscale parameter," *Brazilian Journal of Probability and Statistics*, vol. 26, no. 3, pp. 265–278, Aug. 2012.