Inverse Problems in Bandwidth Sharing Networks

Bruno Kauffmann
Orange Labs, France
bruno.kauffmann@orange.com

***Abstract*—Numerous probing and tomography techniques have been developed for the Internet. They all either flood the network with probing TCP connections or require to send additional probing packets which delays must be precisely measured. In this paper, we propose a new approach, on the basis of existing TCP connections and reaching therefore a zero probing overhead. The foundation of the proposed technique lies in the theory of inverse problems in bandwidth sharing networks, and the approximation of Internet and TCP behavior by that of a bandwidth sharing network. The field of this kind of inverse problems is explored, and we give a few application to toy networks, either with fixed population or with elastic traffic.**

I. Introduction

A. *Internet Tomography and Network Probing*

Tomography is an indirect measurement technique, where one aims at reconstructing the different components of a system through the use of any kind of measurement performed on that system. More specifically, assume that we are given a set of equations depending on some unknown parameter (possibly a vector) θ, and governing the behavior of a complex system. Observing the outputs of the system for some specific inputs then allows us to infer parameter θ. This technique is used, for example, in medical imaging: electro-magnetic waves are sent into a human body, and on the basis of both Maxwell equations for the propagation of electro-magnetic waves and the anatomy of a human body, it is possible to determine the precise condition of the patient.

This paper focuses on *network active probing and tomography*. The main motivation for network tomography is that the Internet is split in a large number of autonomous systems, and that a global observation of the network is difficult. The information obtained by network tomography techniques then allows both end-users to measure the performance of (part of) the Internet, and operators to manage and upgrade their infrastructure with detailed information. Network tomography is an active research subject, which is addressed for example of the ACM Sigcomm's Internet Measurement Conference (IMC) and the *IPPM* Working Group of IETF. We will review existing results in section I-B.

However, most of the known techniques use packet-level tomography, meaning here that the direct problem predicts packet level statistics (*e.g.*, the packet delay distribution), and belong explicitly or implicitly to the field of queueing theory. Although fruitful, this approach of tomography as 'inverse problems in queueing theory' has two main weaknesses:

1) it requires to measure these packet level statistics, which (depending on the statistics) can be difficult or require specific equipment (*e.g.*, DAG cards);
2) it is difficult to introduce the natural feedback from TCP protocol into queueing theory: queueing theory mostly deals with exogenous arrivals, without internal feedback process. This means in particular that the sending of probes can not follow the TCP protocol, and the probes therefore must be packets dedicated to the measurement process, which carry no useful data.

From a practical point of view, it would be ideal to perform network tomography based on TCP flow measurements. It would enable us in some cases to use already existing TCP flows, hence reaching the least intrusive measurement process that is feasible. In the other case, the advantage of TCP flows is that they are easy to set up, and that flow statistics are easier to collect than packet level statistics. This paper aims at being a first step towards this flow-based tomography.

In the next two sections, we review the existing literature on network tomography and recall the basics of bandwidth sharing networks. Section II introduce the main concepts of inverse problems in bandwidth sharing networks. The paper is then structured in three sections, with increasing complexity. Section III and IV study tomography of static toy networks, respectively with a single path and a "triangle network". Section V then extends the case of the single path network to a dynamical number of elastic traffic flows.

B. *Related Work*

There is a vast literature on Internet tomography. The reader may consult [1]–[3] as a few classical entry points. In particular, the determination of link capacities has received considerable interest, using packet pair techniques [4] or the relation between round-trip time, probe size and link capacities [5]. Residual bandwidth and achievable throughput can also be deduced with statistical techniques [6]. All these techniques measure packet-level loss and delay time series, with carefully constructed sequences of probes. Netalyzr [7] and Samknows [8] use flow-based flooding techniques to measure the achievable throughput on specific paths. Their output however is limited to the bottleneck of the path.

A few paper also aim at building solid mathematical foundations for network tomography. The choice of optimal probing sequence for example is discussed in [9], [10], and [11], [12] present results about theoretically-proven delay tomography. The author's PhD [13] contains the results of section III and IV. Other results in this article are new.

C. Bandwidth Sharing Networks

1) Static bandwidth sharing networks: We give now a short introduction to bandwidth sharing networks, as introduced in [14]. Such networks share the bandwidth between competing flows by maximizing a given utility function among the feasible allocation. In this paper, we consider the class of α-fairnesses. Group flows that share the same path and have an identical weight into a class, the $(\boldsymbol{w}, \alpha)$ fairness utility function[1] $U_{\boldsymbol{w},\alpha}$ reads

$$U_{\boldsymbol{w},\alpha}(\boldsymbol{\gamma}) = \sum_{i\in\mathcal{I}} n_i w_i \frac{\gamma_i^{1-\alpha}}{1-\alpha} \,,$$

where $\mathcal{I}$ is the set of classes, n_i (resp. w_i) the number of flows in (resp. the non-negative weight of) in class i and α a non-negative (possibly infinite) real number. Note that the max-min (resp. maximum-throughput) allocation is the limit of (w, α)-fair allocation when α goes to infinity (resp. α goes to zero). A classical and important result [14], [15] about bandwidth sharing networks is that the TCP protocol leads to a proportionally fair bandwidth allocation (*i.e.* $\alpha = 1$).

2) Extension to dynamical network: We follow the elastic traffic model presented in [16]. For each class i, we assume that new customers arrive according to a Poisson process with intensity λ_i. Each customer of class i requires an amount of service which is exponentially distributed with mean $\frac{1}{\mu_i}$, and then leaves the network. At any time, given the number of customers $\boldsymbol{n} = (n_1, \ldots, n_K)$, we consider that the network allocates bandwidth as a static bandwidth sharing network would do, *i.e.* that it maximizes a utility function. Let then $\gamma_i(\boldsymbol{n})$ denote the bandwidth allocated to each flow of class i.

Let T_i denote the operator which adds 1 to the i^{th} coordinate of a vector. The vector $\boldsymbol{n}(t), t \geq 0$ defines a Markov process [16], with transition rates

$$q\left(\boldsymbol{n}, T_i(\boldsymbol{n})\right) = \lambda_i \quad \text{and} \quad q\left(\boldsymbol{n}, T_i^{-1}(\boldsymbol{n})\right) = n_i \mu_i \gamma_i(\boldsymbol{n}) \quad (1)$$

It is also possible to add non-elastic (or streaming) flows to the model [17]. However, for simplicity of presentation, this extension will not be considered here, as the results stay qualitatively the same.

II. Tomography of Bandwidth Sharing Networks

What is an inverse problem in bandwidth sharing networks? Let us start with an example. Consider one user that starts downloading a file. This user can measure its long-term bandwidth, which is proportionally fair. The user may also start several TCP connections in parallel, and measure the long-term bandwidth allocated to these connections, as a function of the number of parallel connections. Is it possible from these long-term bandwidths to estimate the server capacities along the path and the number of competitors, *i.e.* the number of competing flows? We will show in section III that for some classical utility function, the inversion is possible if the user knows the topology along the path (*e.g.*, using Traceroute).

Tomography problems can be described depending on a few criteria. These criteria have been identified in [18] for inverse problems in queueing theory. For clarity, we will summarize them quickly in this section, and refer the reader to section 2 of [18] for a more complete description.

A. Direct problem

Bandwidth Sharing Network theory describes the bandwidth allocation to different flows in a network, as a function of several parameters. These parameters can be related to the structure of the network (the number of stations, the topology and the link capacities) or to the nature of the traffic (the number of flows in each class, the routes for a static problem, or arrival processes and service requirement / sojourn time for elastic/streaming traffic.). The solution of the direct problem might be either an explicit closed-form allocation, some equation that is verified by the allocation, or asymptotic properties of the allocation in some specific settings.

B. Noise

Deviation from ideal assumptions, which we will call hereafter 'noise', can happen for several reasons:

- the bandwidth allocation algorithm might be unperfect;
- there may be actual measurement noise in the data;
- in the dynamical case, time series obtained from a measurement experiment are by nature finite, and any estimator contains thus a statistical error when the bandwidth is a random process oscillating around the theoretical bandwidth allocation (*e.g.*, TCP);

C. Probing actions

The observables are generated through several actions of the network prober. These actions include the *choice of the probing topology* (including *point-to-point* or *multipoint* probing). One can also distinguish between *passive probing* where the prober just observes actual flows without perturbing the network, and *active probing* where the prober actively set up some probing flows, and thus interact with the network.

D. Observables and unknown parameters

Observables are the raw data available to the prober and deriving from the above actions. For bandwidth sharing networks, they consist of the times series of measured bandwidth and the number of probing flows for each class.

Unknown parameters are the parameters of the direct equation which are not specified, even indirectly, by the model.

E. Intrusiveness, bias and restitution

Probing flows are not identified as such by the network, and get their fair bandwidth share. They are therefore intrusive to the system, and a general recommendation is usually to limit the intrusiveness of the tomography. Intrusiveness might seem at first glance to make the tomography problem more difficult. In fact, we will see that it may be useful and can be leveraged in many cases (see section III-B).

[1] We will abusively write $U_1(\boldsymbol{\gamma}) = \sum_{s\in\mathcal{S}} \frac{\gamma_s^{1-\alpha}}{1-\alpha}$ when $\alpha = 1$, instead of $U_1(\boldsymbol{\gamma}) = \sum_{s\in\mathcal{S}} \log \gamma_s$. This is not formally correct, but all computations remain valid in the limit $\alpha \to 1$, as well as the final results.

An important effect of intrusiveness is that it affects the measured performance metrics. For example, the bandwidth allocated to a single probe flow is not equal to the bandwidth received by other flows following the same route in absence of the probe flow. Hence, in order to estimate *ground truth* (*i.e.* the system parameters in absence of probes), there is a need for a *restitution* phase, where one reconstructs the ground truth metrics on the basis of observations of the perturbed system.

F. Identifiability and ambiguity

The information contained in the observables might not be sufficient to determine a unique ground truth, meaning that the same observables can stem from different parameters. This raises a host of identifiability questions.

G. Estimation problems and design of experiment

We mentioned in section II-B that in some cases, the observables are finite series of a random dynamical system. The inversion step is then no longer a deterministic problem, but a problem of statistical estimation. The design of such estimators, as well as the characterization of their properties (*e.g.*, variance, asymptotic consistency) is an important issue.

For active probing, the degrees of freedom in how to set up probe connections opens another level of optimization, which is called *design of experiment*.

III. The static single path case

Consider a single static path, as depicted in Fig. 1. The path consists of K servers (S_1, …, S_k) in series, where the server S_j has capacity C_j. There are $K+1$ class of users: users of class 0 use the whole path, and users of class i ($1 \leq i \leq K$) cross only server S_i. Each class i has n_i users, and each of these users receives a bandwidth equal to γ_i. Additionally, there are x probes of class 0, using the whole path.

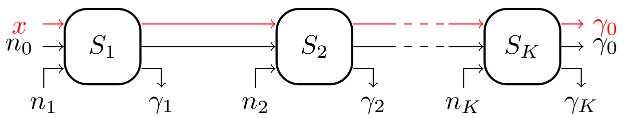

Fig. 1. An example of path.

A. Direct problem

We first describe the *direct problem*, *i.e.* the evolution of the system. The α-fairness here reads

$$U_{\boldsymbol{w},\alpha}(\boldsymbol{\gamma}) = \frac{1}{1-\alpha}\left[w_0(n_0+x)\gamma_0^{1-\alpha} + \sum_{i=1}^{K} w_i n_i \left(\frac{C_i-(n_0+x)\gamma_0}{n_i}\right)^{1-\alpha}\right], \quad (2)$$

where for notation simplicity, when $n_i = 0$, we set

$$n_i\left(\frac{C_i-(n_0+x)\gamma_0}{n_i}\right)^{1-\alpha} = \begin{cases} 0 & \text{if } C_i-(n_0+x)\gamma_0 \geq 0\,, \\ -\infty & \text{otherwise}\,. \end{cases}$$

1) Maximum throughput: If $\sum_{i=1}^{K} w_i \mathbb{1}_{n_i>0} > w_0$, the maximum utility allocation in such a case is $\gamma_0 = 0$ and $\gamma_i = \mathbb{1}_{n_i>0}\frac{C_i}{n_i}$.

Otherwise, it equals

$$\gamma_0 = \min_{1\leq j\leq K} \frac{C_j}{n_0+x}, \qquad \gamma_i = \frac{C_i - \min_{1\leq j\leq K} C_j}{n_i}\,. \quad (3)$$

2) Max-min allocation: The only max-min allocation for this network is

$$\gamma_0 = \min_{1\leq i\leq K} f_i(x), \qquad \gamma_i = \frac{C_i-(n_0+x)\gamma_0}{n_i}\,, \quad (4)$$

where we define $f_i(x) = \frac{C_i}{n_0+x+n_i}$ for all $1 \leq i \leq K$.

3) Other α-fair allocations: The parameter α is now strictly positive and finite. The $(\boldsymbol{w},\alpha)$-fair allocation then verifies the stationary equation

$$\frac{w_0}{\gamma_0^{\alpha}} = \sum_{i=1}^{K} w_i \mathbb{1}_{n_i>0}\left(\frac{n_i}{C_i-(n_0+x)\gamma_0}\right)^{\alpha}, \quad (5)$$

which has no closed-form solution in a general case. However, a solution can be found when all servers have the same capacity C. The $(\boldsymbol{w},\alpha)$-fair allocation then reads

$$\gamma_0 = \frac{C}{n_0+x+\tilde{n}} \quad \text{and} \quad \gamma_i = \frac{\tilde{n}}{\tilde{n}+n_0+x}\frac{C}{n_i} \quad (6)$$

where $\tilde{n} = \left(\frac{\sum_{i=1}^{K} w_i n_i^{\alpha}}{w_0}\right)^{\frac{1}{\alpha}}$ is the “weighted α-sum”.

B. Tomography

1) Maximum throughput allocation: If $w_0 < \sum_{1\leq i\leq K} w_i \mathbb{1}_{n_i>0}$, the inverse problem is ill-posed. The bandwidth allocated to the probing path is zero, independently of the probing intensity. It is therefore easy to identify such an allocation policy, as it is the only one among the considered policies that allows starvation of the probing flows. Unfortunately, the fact that the probing path gets a zero bandwidth allocation does not allow us to deduce any parameter. When observing the bandwidth allocated to users of class i, we can deduce the ratio $\frac{C_i}{n_i}$, but we can not identify individually C_i and n_i.

In the other case, the allocation is as in (3), and two observation points $(x_i, \gamma_0(x_i))$ are sufficient to determine n_0 and $\min_{1\leq j\leq K} C_j$ as follows (assuming $x_1 < x_2$):

$$n_0 = \frac{x_2\gamma_0(x_2) - x_1\gamma_0(x_1)}{\gamma_0(x_1)-\gamma_0(x_2)}$$

$$\min_{1\leq j\leq K} C_j = \frac{(x_2-x_1)\gamma_0(x_1)\gamma_0(x_2)}{\gamma_0(x_2)-\gamma_0(x_1)}\,.$$

2) Max-min allocation: Observe that $\gamma_0(x) = \min_{1\leq i\leq K} f_i(x)$. Several quantities are therefore immediately non-identifiable. First, one can identify the sums $n_0+n_i, i \geq 1$, and but not individually each $n_i, i \geq 0$. Second, only the servers which are bottlenecks for some probing intensity can have their capacity and cross-traffic intensity identified.

We now state the three following preliminary lemmas.

Lemma 1. *Two functions $f_i(x)$ and $f_j(x)$ that are equal at more than 2 points on the real positive line are identical, and $C_i = C_j$ and $n_0 + n_i = n_0 + n_j$.*

Proof: Note that $f_i(x) - f_j(x) = \frac{C_i(n_0+n_j)-C_j(n_0+n_i)+x(C_i-C_j)}{(n_0+n_i+x)(n_0+n_j+x)}$, hence $f_i(x) = f_j(x)$ is equivalent to $C_i(n_0+n_j)-C_j(n_0+n_i)+x(C_i-C_j) = 0$. This equation admits at most one solution unless it is degenerate, which means $C_i = C_j$ and $C_i(n_0 + n_j) - C_j(n_0 + n_i)$. The last equality is easily deduced from these 2 last equations. ■

Lemma 2. *Two functions $f_i(x)$ and $f_j(x)$ are equal and have equal derivatives at some point on the real positive line if and only if they are identical.*

Proof: Assume that f_i and f_j are equal and have equal derivatives at x. Then $f_i(x) = f_j(x)$ and $f_i^{'}(x) = f_j^{'}(x)$, which we can rewrite as $\frac{C_i}{n_0+n_i+x} = \frac{C_j}{n_0+n_j+x}$ and $\frac{-C_i}{(n_0+n_i+x)^2} = \frac{-C_j}{(n_0+n_j+x)^2}$. It follows directly that $n_0 + n_i + x = n_0 + n_j + x$, hence $n_0 + n_i = n_0 + n_j$, $C_i = C_j$ and both functions are identical. ■

Assume now that server j is at bottleneck when probing at the intensities belonging to the "minimum set" $X_j = \{x > 0, \gamma_0(x) = f_j(x)\}$. Then straightforward computation leads for all (x_1, x_2) to

$$(x_1, x_2) \in X_j^2 \Leftrightarrow \begin{cases} \frac{x_2\gamma_0(x_2)-x_1\gamma_0(x_1)}{\gamma_0(x_1)-\gamma_0(x_2)} & = n_0 + n_j \ , \\ \frac{(x_2-x_1)\gamma_0(x_1)\gamma_0(x_2)}{\gamma_0(x_1)-\gamma_0(x2)} & = C_j \ . \end{cases} \quad (7)$$

Lemma 3. *The minimum sets X_j are convex.*

Proof: By contradiction, let $x_1 < x_3$ be two elements of X_j, $x_2 \in [x_1, x_3]$ and assume that $x_2 \notin X_j$. Then $\exists k \neq j$ s.t.

$$f_j(x_1) \leq f_k(x_1) \ , \quad f_j(x_2) > f_k(x_2) \text{ and } \quad f_j(x_3) \leq f_k(x_3) \ .$$

Functions $f_j(.)$ and $f_k(.)$ are continuous, which implies from the intermediate value theorem that $\exists\ x_1 \leq x_4 \leq x_2$ and $x_2 \leq x_5 \leq x_3$ such that $f_j(x_4) = f_k(x_4)$ and $f_j(x_5) = f_k(x_5)$. This means that they intersect in two points and must be identical functions according to Lemma 1, and hence a contradiction and $x_2 \in X_j$. ■

The following theorem finally allows us to proceed with the inversion step

Theorem 4. *Consider a set of observation points $(x_i, \gamma_0(x_i)), i = 1, \ldots, N$, stemming from a max-min bandwidth sharing path, with capacities $(C_1, \ldots, C_K)$ and number of flows $(n_0, n_1, \ldots, n_K)$. If there is a subset Y of X such that $|Y| \geq 3$ and $\exists\ (C, n) \in \mathbb{R}^2, \forall x \in Y, \gamma_0(x) = \frac{C}{n+x}$, then there exists a server S_j such that $Y \subset X_j$ (hence $C_j = C$ and $n_0 + n_j = n$) and any observation point $(x_i, \gamma_0(x_i))$ such that $\min Y \leq x_i \leq \max Y$ also verifies $\gamma_0(x_i) = f_j(x)$.*

Proof: Let $x_1 < x_2 < x_3$ be three points of Y. By assumption of the max-min allocation, there exists a server j with capacity C_j and number of flows $n_0 + n_j$ such that $\gamma_0(x_2) = f_j(x_2)$ and $\forall i \neq j, f_i(x_2) \geq \gamma_0(x_2)$. By Lemma 2, if functions $f_j(.)$ and $\frac{C}{n+x}$ are equal and have equal derivatives, the result is immediate. Otherwise, there exists a such that $f_j(x_2 + a) < \frac{C}{n+x_2+a}$. Assume for simplicity that $a > 0$. Since x_2 and x_3 belongs to Y, we have by definition that $f_j(x_2) = \gamma_0(x_2) = \frac{C}{n+x_2}$ and $f_j(x_3) \geq \gamma_0(x_3) = \frac{C}{n+x_3}$. It follows from the intermediate value theorem that $\exists x_4 \in [x_2 + a, x_3]$ s.t. $f_j(x_4) = \frac{C}{n+x_4}$, and $f_j(.)$ and $\gamma_0(.)$ are equal at x_2 and x_4. Lemma 1 then concludes that $C = C_j$ and $n = n_0 + n_j$. The last part follows immediately from the fact that the minimum sets are convex. ■

The tomography algorithm then proceed as follows:

1) Given a pair of "unassigned" consecutive measurement intensities (x_1, x_2), determine a possible candidate pair (C, n) using (7);
2) Find the maximal subset X of measurement intensities such that $\forall x \in X$, $\gamma_0(x) = \frac{C}{n+x}$; from Theorem 4, this subset is convex, and one can sequentially check every measurement intensity;
3) if $|X| \geq 3$, then assign (C, n) to a server S_j, and X as its minimum set X_j;
4) If there exist at least 3 consecutive unassigned and unchecked measurement intensities, go back to 1.
5) Else, either stop and return the assigned pairs (C_j, n_j), or add measurement points in order to have 3 consecutive unassigned measurement intensities.

The algorithm relies on equation (7) and Theorem 4 to identify the minimum sets X_j. Note, however, that we identify only a subset of the pairs set $\{(C_j, n_0 + n_j)\}_{1\leq j\leq K}$, and that we are not able to give the order of the servers.

3) Other α-fair allocations: In this section, we analyze the case when the allocation is $\alpha-$fair, α being a positive integer.

First consider the case where all servers have the same capacity C. Recall solution (6): $\gamma_0(x) = \frac{C}{n_0+\tilde{n}+x}$, where $\tilde{n}$ is defined as $\tilde{n} = \left(\sum_{i=1}^{K} w_i n_i^\alpha\right)^{\frac{1}{\alpha}}$. Using (7), we can fully inverse the problem if and only if we have two different measure intensities x_1 and x_2, leading to:

$$\begin{cases} n_0 + \tilde{n} & = \frac{x_2\gamma_0(x_2)-x_1\gamma_0(x_1)}{\gamma_0(x_1)-\gamma_0(x_2)} \ , \\ C & = \frac{(x_2-x_1)\gamma_0(x_1)\gamma_0(x_2)}{\gamma_0(x_1)-\gamma_0(x2)} \ . \end{cases} \quad (8)$$

In the case of general capacities, we are not able to predict the bandwidth allocation. However, the identifiability problems disappear, and can infer all capacities and flow numbers with mild assumptions. In fact, recall that (5) is valid in the present case; multiplying each side by $\gamma_0(x)^\alpha \prod_{i=1}^{K} (C_i - (n_0 + x)\gamma_0(x))^\alpha$, we can rewrite it as

$$w_0 \prod_{i=1}^{K} (C_i - (n_0 + x)\gamma_0(x))^\alpha = \gamma_0(x)^\alpha \sum_{i=1}^{K} w_i n_i^\alpha \prod_{j\neq i} (C_j - (n_0 + x)\gamma_0(x))^\alpha \ . \quad (9)$$

From (9), it is obvious that the flow numbers n_i and weights w_i are not independently identifiable. They appear only within their product $w_i n_i^\alpha$, and as neither n_i, w_i nor α can be changed by the prober, the best that can be identified is the product

$w_i n_i^\alpha$. Remark that we can assume without loss of generality that $w_0 = 1$.

Assume that α is an integer, and define constant values $(a_{i,j})_{0\le i\le j\le \alpha K}$, which depend only on the capacities $(C_1, \dots, C_K)$, the flow numbers $(n_0, n_1, \dots, n_K)$ and the weights $\boldsymbol{w}$, such that equation (9) now reads:

$$\sum_{0\le i\le j\le \alpha K} a_{i,j} x^i \gamma_0(x)^j = 0\,. \tag{10}$$

The interest of the polynomial coefficients $a_{i,j}$ lies in the two following proposition:

Proposition 5. *(Proof in Appendix) Given the polynomial coefficients $a_{i,j}$, there is a unique set of capacities $\{C_1, \dots, C_K\}$ and associated "weighted-population" $(n_0^\alpha, w_1 n_1^\alpha, \dots, w_K n_K^\alpha)$ s.t. (9) and (10) are equivalent.*

Proposition 6. *(Proof in Appendix) Consider a set of N measurement points $(x_i, \gamma_0(x_i))_{i=1,\dots,N}$ for N different probing flow numbers $x_1, \dots, x_N$. For large enough N, there is a unique solution of coefficients $a_{i,j}$ such that (10) is verified for all x_i, $1 \le i \le N$.*

4) Summary: The inversion is possible when enough measurement points are available:

1) If $K = 1$ or if all capacities are known to be identical, use (8) to identify the capacity and the aggregated flow number. Individual flow numbers cannot be determined.
2) If $\alpha > 1$, use (17) to determine the polynomial coefficient $a_{i,j}$. Find $(C_1, \dots, C_k)$ as the roots of the polynomial $\sum_{i=0}^{\alpha K} a_{i,i} x^i \gamma_0(x)^i$. If all capacities are pairwise different, use (15) to estimate the individual flow numbers. Otherwise, estimate the "aggregated" flow numbers $\sum_{i:C_i=C} n_i^\alpha$ for identical capacities servers, and individual flows with (15).
3) If $\alpha = 1$, use (17) to determine the polynomial coefficient $a_{i,j}$. Find $(C_1, \dots, C_k)$ as the roots of the polynomial $\sum_{i=0}^{\alpha K} a_{i,i} x^i \gamma_0(x)^i$. Use then (15) to express all cross-traffic flow numbers $(n_1, \dots, n_K)$ (or their aggregated sum in case of identical capacities) as affine functions of n_0. Use these functions and (16) to obtain a second-degree equation, solve it to get n_0, hence the other flow numbers as well.

C. Numerical application

In this section, we focus on the case of general integer α, which seems the most interesting for applications and the simplest from a numerical point of view. To keep notation simple, we assume that all weights w_i are equal to 1.

The simulations were performed using Matlab. For any number of flows, the bandwidth allocation is computed using convex optimization tools. Based on these measurement points $(x_k, \gamma_0(x_k))$, we then estimate the coefficient $a_{k,(i,j)}$ (with a special care for the matrix inversion step), then use the Matlab "root" routine to find the roots $(C_1, \dots, C_K)$ of the polynomial $\sum_{i=0}^{K} a_{i,i} X^i$. For $k > 1$, the estimation of $\boldsymbol{n}$ is performed with the Matlab right-side division, a routine designed for solving matrices equations of the type $AX = B$. If $\alpha = 1$, the system is not linear, and we use the "Fsolve" function, where the objective function is the vector of the equations (15) and (16).

C	**n**	# add.	est. A	est. **C**	est. **n**
$\begin{pmatrix}2\\1\end{pmatrix}$	$\begin{pmatrix}0\\3\\1\end{pmatrix}$	0	$\begin{bmatrix}2.0001 & -4.9996 & -0.0015\\0 & -3.0001 & 3.9997\\0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}2.0001\\1\end{pmatrix}$	$\begin{pmatrix}-0.0004\\3.0001\\1.0002\end{pmatrix}$
$\begin{pmatrix}30\\1\end{pmatrix}$	$\begin{pmatrix}2\\5\\1\end{pmatrix}$	0	$\begin{bmatrix}-1.1101 & 0.4909 & 8.8867\\0 & 0.1102 & 5.8391\\0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}-1.11\\1\end{pmatrix}$	$\begin{pmatrix}2.97\\-0.08\\0.08\end{pmatrix}$
$\begin{pmatrix}30\\20\end{pmatrix}$	$\begin{pmatrix}2\\5\\1\end{pmatrix}$	0	$\begin{bmatrix}600.03 & -230.27 & 16.0772\\0 & -49.999 & 10.0095\\0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}29.99\\20.00\end{pmatrix}$	$\begin{pmatrix}2.01\\4.99\\0.997\end{pmatrix}$
$\begin{pmatrix}30\\1\end{pmatrix}$	$\begin{pmatrix}0\\200\\0\end{pmatrix}$	0	$\begin{bmatrix}51.82 & -568.6 & 1487\\0 & -31.256 & 204.2\\0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}29.49\\1.76\end{pmatrix}$	$\begin{pmatrix}7.56\\189\\-0.01\end{pmatrix}$
$\begin{pmatrix}30\\1\end{pmatrix}$	$\begin{pmatrix}0\\200\\0\end{pmatrix}$	5	$\begin{bmatrix}29.86 & -199.16 & 0.25\\0 & -30.86 & 199.06\\0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}29.86\\1.0001\end{pmatrix}$	$\begin{pmatrix}0.0013\\199.06\\-0.0001\end{pmatrix}$
$\begin{pmatrix}30\\1\end{pmatrix}$	$\begin{pmatrix}0\\50\\0\end{pmatrix}$	5	$\begin{bmatrix}0.3085 & 0.4756 & 0.0\\0 & -1.309 & -0.4756\\0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}1.0\\0.31\end{pmatrix}$	$\begin{pmatrix}-0.476\\0.476\\0.0000\end{pmatrix}$

TABLE I
PROPORTIONAL FAIRNESS: TWO-SERVER CASES.
THE TWO FIRST COLUMNS SHOW THE GROUND TRUTH. THE THIRD COLUMN INDICATES HOW MANY ADDITIONAL MEASUREMENT POINTS HAVE BEEN USED ON TOP ON THE MINIMUM 5 REQUIRED. THE COLUMN "EST. A" IS A MATRIX WHOSE COEFFICIENT (I,J) IS THE ESTIMATED $a_{i,j}$. THE FIFTH OR SIXTH COLUMNS SHOW THE GROUND TRUTH ESTIMATION.

1) Proportional fairness: The case of proportional fairness is slightly different, because (15) and (16) have to be used to determine n_0. Table I presents a few numerical results.

The estimation is reasonable in the cases 1, 3, 4 and 5. In the cases 2 and 6, the estimation of the bottleneck capacity (which, arguably, is the most important value to estimate) is precise, but the second value is meaningless. Comparing cases 4 and 5, we can see that additional measurement points allow a better precision in some cases. In both cases with poor precision, the matrix P of equation (17) was nearly singular. As a possible explanation, these were also the cases where a single server is the clear bottleneck for the probe path, meaning that the bandwidth is shared almost as $\gamma_0(x) = \frac{C}{n+x}$, where C is the capacity bottleneck server and n the number of flows (outside probes) which cross that server. As x grows large, the coefficient $k, (i,j)$ of P is $x_k^i \gamma_0(x_k)^j \to \frac{C^j}{x^{i-j}}$, and the coefficients $a_k, (i,i) \to C^i$ are almost independent of the line index k. This means that the columns (i,i) are almost all equivalent, and the matrix P is near singular.

As shown in table II, the instability is worse for the three-server case. It is still possible to get "correct" estimation as in the line 2, 3 or 4, but additional measurement points are now required in order to correctly estimate the coefficient $a_{i,j}$. The fact that the estimation is harder when one server is the clear bottleneck or is clearly over-provisioned, still remains.

2) Other fairnesses: The method is numerically unstable and leads to meaningless (complex) results. The reason is that it requires at least $\frac{(\alpha K+1)(\alpha K+2)}{2} - 1$ measurement points, and the inversion of a square matrix of the same size. Even for the smallest case with $K = 2$ and $\alpha = 2$, the minimum size is

C	**n**	# add.	est. A	est. **C**	est. **n**
$\begin{pmatrix}2\\1\\3\end{pmatrix}$	$\begin{pmatrix}0\\3\\1\\7\end{pmatrix}$	0	$\begin{bmatrix}-1.74 & 8.28 & 5.24 & -22.7\\0 & 5.04 & -19.6 & -3.7\\0 & 0 & -4.3 & 10.7\\0 & 0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}1.\\0.66\\2.64\end{pmatrix}$	$\begin{pmatrix}-0.28\\0.89\\-0.29\\9.22\end{pmatrix}$
$\begin{pmatrix}2\\1\\3\end{pmatrix}$	$\begin{pmatrix}0\\3\\1\\7\end{pmatrix}$	5	$\begin{bmatrix}-5.7 & 25.5 & 10.4 & -0.706\\0 & 10.6 & -33.9 & -9.12\\0 & 0 & -5.9 & 10.2\\0 & 0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}1.93\\1\\2.97\end{pmatrix}$	$\begin{pmatrix}-0.32\\1.70\\0.68\\6.80\end{pmatrix}$
$\begin{pmatrix}2\\1\\3\end{pmatrix}$	$\begin{pmatrix}0\\3\\1\\7\end{pmatrix}$	10	$\begin{bmatrix}-6 & 26.9 & 9.63 & -0.005\\0 & 11 & -35.15 & -8.46\\0 & 0 & -6 & 10.23\\0 & 0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}1.98\\1\\3.02\end{pmatrix}$	$\begin{pmatrix}-0.29\\2.25\\0.7\\6.4\end{pmatrix}$
$\begin{pmatrix}2\\5\\3\end{pmatrix}$	$\begin{pmatrix}0\\3\\1\\7\end{pmatrix}$	10	$\begin{bmatrix}-25.7 & 99.8 & 18.52 & -13.1\\0 & 27.5 & -67 & -9.4\\0 & 0 & -69.3 & 10.2\\0 & 0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}2.01\\4.39\\2.91\end{pmatrix}$	$\begin{pmatrix}-0.29\\2.71\\0.5\\6.15\end{pmatrix}$
$\begin{pmatrix}2\\20\\3\end{pmatrix}$	$\begin{pmatrix}0\\3\\1\\7\end{pmatrix}$	10	$\begin{bmatrix}-13.1 & 99.2 & -147.9 & -150\\0 & 17.0 & -85.15 & 62.8\\0 & 0 & -7.24 & 17.8\\0 & 0 & 0 & 1\end{bmatrix}$	$\begin{pmatrix}2.00\\3.16\\2.08\end{pmatrix}$	$\begin{pmatrix}1.53\\2.11\\7.8\\12.49\end{pmatrix}$

TABLE II
PROPORTIONAL FAIRNESS: THREE-SERVER CASES. THE COLUMNS ARE ORGANIZED AS IN TABLE I.

already 14. In all cases we have considered, the matrix has a few eigenvalues close to zero, and the inversion leads to unexploitable estimation of the coefficients $a_{i,j}$.

IV. THE STATIC TRIANGLE NETWORK

Previous section focused on the single source-destination path, which is a very specific network. We extend here the results for the general α-fairness to a non-trivial (but small) network topology: a "triangle network", as depicted in Fig. 2.

The networks consists of 3 servers, each pair being connected to each other. Server S_i has a capacity C_i. k_i flows cross the server S_i, and each of them gets an allocated bandwidth η_i. There are also flows using 2 servers: n_3 (resp. n_2 and n_1) flows use the route (S_1, S_2) (resp. (S_1, S_3) and (S_2, S_3)) and each of them gets an allocated bandwidth γ_3 (resp. γ_2 and γ_1). The prober can add x_1 (resp. x_2 and x_3) flows on the route (S_2, S_3) (resp. (S_1, S_3) and (S_1, S_2)). Each of these flows get each the same bandwidth allocated to the route.

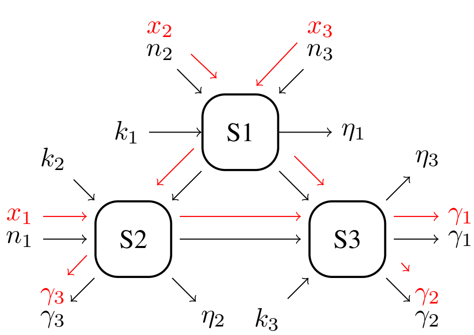

Fig. 2. The triangle network.

Differentiating the fairness with respect to each γ_i, we get the following equations:

$$\begin{aligned}(C_2 - (x_1+n_1)\gamma_1 - (x_3+n_3)\gamma_3)^\alpha \\ (C_3 - (x_1+n_1)\gamma_1 - (x_2+n_2)\gamma_2)^\alpha - \\ \gamma_1^\alpha \times [k_2^\alpha (C_3 - (x_1+n_1)\gamma_1 - (x_2+n_2)\gamma_2)^\alpha \\ + k_3^\alpha (C_2 - (x_1+n_1)\gamma_1 - (x_3+n_3)\gamma_3)^\alpha] = 0\end{aligned}$$

$$\begin{aligned}(C_1 - (x_2+n_2)\gamma_2 - (x_3+n_3)\gamma_3)^\alpha \\ (C_3 - (x_1+n_1)\gamma_1 - (x_2+n_2)\gamma_2)^\alpha \\ - \gamma_2^\alpha \times [k_1^\alpha (C_3 - (x_1+n_1)\gamma_1 - (x_2+n_2)\gamma_2)^\alpha \\ + k_3^\alpha (C_1 - (x_2+n_2)\gamma_2 - (x_3+n_3)\gamma_3)^\alpha] = 0 \quad (11)\end{aligned}$$

$$\begin{aligned}(C_1 - (x_2+n_2)\gamma_2 - (x_3+n_3)\gamma_3)^\alpha \\ (C_2 - (x_1+n_1)\gamma_2 - (x_3+n_3)\gamma_3)^\alpha \\ - \gamma_3^\alpha \times [k_1^\alpha (C_2 - (x_1+n_1)\gamma_1 - (x_3+n_3)\gamma_3)^\alpha \\ + k_2^\alpha (C_1 - (x_2+n_2)\gamma_2 - (x_3+n_3)\gamma_3)^\alpha] = 0\end{aligned}$$

From these equations, we can identify the server capacities C_1, C_2 and C_3, and the class population k_1, k_2,k_3 and n_1, n_2 and n_3. Indeed, similarly to (10), it is possible to define constant values $(b_{i,j,k,l,m,n})$ (resp. $(c_{i,j,k,l,m,n})$ and $(d_{i,j,k,l,m,n})$), depending only on the flow numbers n_i and k_i, the weights w_i and v_i and the capacities C_i, such that the first (resp. second and third) equation reads

$$\sum_{i+j+k\leq 2\alpha} \sum_{l\leq i} \sum_{m\leq j} \sum_{n\leq k} b_{i,j,k,l,m,n} \gamma_1^i \gamma_2^j \gamma_3^k x_1^l x_2^m x_3^n = 0\,. \quad (12)$$

The knowledge of these coefficients is sufficient to estimate the server capacities (C_1, C_2, C_3), the flow numbers (n_1, n_2, n_3) and $(k_1^\alpha, k_2^\alpha, k_3^\alpha)$, using carefully chosen combination of these coefficients. Due to lack of space, we just give hints of a possible inversion algorithm here:

- the capacities (C_1, C_2, C_3) can be identified from coefficients $\{b_{0,0,i,0,0,i}\}$, $\{b_{0,i,0,0,i,0}\}$ and $\{c_{0,0,i,0,0,i}\}$;
- (n_1, n_2, n_3) can be estimated based on coefficients $b_{0,1,0,0,0,0}$, $b_{0,0,1,0,0,0}$ and $c_{1,0,0,0,0,0}$;
- coefficients $b_{\alpha,0,0,0,0,0}$, $c_{0,\alpha,0,0,0,0}$ and $d_{0,0,\alpha,0,0,0}$ are linear in k_1^α, k_2^α and k_3^α.

Finally, it remains to show how one can estimate the polynomial coefficient values $(b_{i,j,k,l,m,n})_{j\leq\alpha, k\leq\alpha, i+j+k\leq\alpha, l\leq i, m\leq j, n\leq k}$. There are $M = \frac{5\alpha^6+51\alpha^5+209\alpha^4+441\alpha^3+506\alpha^2+300\alpha+72}{72}$ such $b_{i,j,k,l,m,n}$ coefficients. For example, we have $M = 22$ for $\alpha = 1$ and $M = 160$ for $\alpha = 2$. Assuming that the prober has access to $N \geq M-1$ measurement points $(x_1(p), x_2(p), x_3(p), \gamma_1(p), \gamma_2(p), \gamma_3(p))_{1\leq p\leq N}$ (where we abusively write $\gamma_i(k)$ for $\gamma_i(x_1(k), x_2(k), x_3(k))$), it is possible to estimate these coefficients using a linear system similar to (17). In all cases that we simulated, we were able to add enough points such that the system is full-rank, and we conjecture that it is possible in all non-degenerate cases.

V. Extension to the dynamical case

The two previous sections dealt with static networks. In this section, we finally address an example of dynamical case, where users (or flows) join and leave the network, such as presented in section I-C2.

A. The direct equation

Consider once again the single path network. Let $\boldsymbol{n}(t) = (n_0(t), \dots, n_K(t))$ denote the number of flows for each class at time t, and λ_i (resp. $\frac{1}{\mu_i}$) denote the arrival rate (resp. the mean required service) of flows of class i. Let also $x(t)$ denote the number of end-to-end probing connections at time t (they belong to class 0).

We assume that the bandwidth allocation is proportionally-fair (*i.e.* $\alpha = 1$), that all capacities are equal to C, and for notation simplicity, that all weights w_i are identical. At any time t, the bandwidth allocation is then described by (6) and we can explicit the rates of the Markov chain in (1) as

$$\begin{aligned} q\left(\boldsymbol{n}, T_i(\boldsymbol{n})\right) &= \lambda_i \, , \\ q\left(\boldsymbol{n}, T_0^{-1}(\boldsymbol{n})\right) &= \mu_0 \frac{n_0}{x + \sum_{i=0}^K n_i} C \, , \qquad (13) \\ q\left(\boldsymbol{n}, T_i^{-1}(\boldsymbol{n})\right) &= \mathbb{1}_{n_i > 0} \mu_i \frac{\sum_{j=1}^K n_j}{x + n_0 + \sum_{j=1}^K n_j} C \, . \end{aligned}$$

Let $\rho_i = \frac{\lambda_i}{C\mu_i}$ denote the load of class i. When the number of probing connection is fixed, we can generalize Theorem 1 in [16] and state that if $\sup_{1 \le i \le K} \rho_0 + \rho_i < 1$, the process $\boldsymbol{n}(t)$ is reversible with equilibrium distribution given by

$$\pi\left(\boldsymbol{n}\right) = A \prod_{i=0}^K \rho_i^{n_i} \binom{x + \sum_{i=0}^K n_i}{\sum_{j=1}^K n_j} \, , \qquad (14)$$

where the normalizing constant A equals

$$A = \frac{\rho_0^x \prod_{j=1}^K (1 - \rho_0 - \rho_j)}{(1 - \rho_0)^{K-1}} \, .$$

B. Tomography

Proposition 7. *Consider a dynamical bandwidth sharing network, as described in section V-A. Assume that one can observe the rate $\gamma_0(t)$ allocated to the probing flows. It is then possible to identify the capacity C of all nodes, as well as the load ρ_0 of the end-to-end class, the set $\{\rho_j\}_{1 \le j \le K}$ of the cross-traffic class loads and the total flow arrival intensity $\Lambda = \sum_{i=0}^K \lambda_i$.*

Proof: In fact, the tomography algorithm reads as follows:

1) Changing x when $\boldsymbol{n}(t)$ remains in the same state, use equation (8) to determine the value of C and the total number of flows $N(t) = \sum_{i=0}^K n_i(t)$ in the network. Note that given C, it is then possible at any time using (6) to measure $N(t)$, on the basis of C and $\gamma_0(t)$.
2) For different numbers x of probing connections, measure the stationary probability $P_x(N(t) = 0)$ that there are no other connection in the system. We know that (14),
$$P_x(N(t) = 0) = A^{-1} = \rho_0^x B$$
where B is independent of x. It is thus possible to estimate the load ρ_0.
3) Simultaneously, using (13), the total flow arrival rate Λ is equal to the rate of transitions from one state with N flows to one state with $N+1$ flows. Since $N(t)$ can be deduced at any time, the rate of such transitions, consequently Λ, can be measured.
4) Finally, for a fixed value of x, the prober can also observe the stationary probability $P(N(t) = l)$ that there are l other flows in the network, for all $1 \le l \le K$. We know that $P(N(t) = 1) = A^{-1}\left(\rho_0 + \sum_{j=1}^K \binom{x+1}{1} \rho_j\right)$, hence $\sum_{j=1}^K \rho_j$ can be estimated. Similarly, $P(N(t) = 2)) = A^{-1}\left(\rho_0^2 + \rho_0 \binom{x+2}{1} \sum_{j=1}^K \rho_j + \binom{x+2}{2} \sum_{j=1}^K \rho_j^2 + \binom{x+2}{2} \sum_{1 \le i \le j \le K} \rho_i \rho_j\right)$, and the last two sums can thus be estimated using a linear system (*e.g.*, using $\left(\sum_{j=1}^K \rho_j\right)^2 = \sum_{j=1}^K \rho_j^2 + 2\sum_{1 \le i \le j \le K} \rho_i \rho_j$ as a second equation). Recursively on l, for all decomposition of $l = n_1 + \dots + n_m$ in m positive integers, it is possible to estimate the sum
$$\sum_{\substack{(i_1, \dots, i_m) \\ i_j \ne i_k \\ n_j = n_k \Rightarrow i_j < i_k}} \rho_{i_1}^{n_1} \dots \rho_{i_m}^{n_m} \, .$$
$\{\rho_1, \dots, \rho_K\}$ are then the roots of the polynomial in variable X
$$\sum_{j=0}^K \binom{K}{j} X^{K-j} \sum_{1 \le i_1 < \dots < i_j \le K} \rho_{i_1} \dots \rho_{i_K} \, .$$

■

VI. Conclusion

This paper studied inverse problems in bandwidth sharing networks as a mathematical foundation for a flow-based Internet tomography. One fundamental result is that measuring the bandwidth allocated by an (idealized) TCP to probing connections allows us theoretically to infer each identifiable parameter of a static network. Since bandwidth allocation takes into account the topology of the network, specific equations need to be developed for a given topology, but the proposed method can be easily adapted to new topologies. Numerical examples illustrate the soundness of the proposed method, as well as the intrinsic numerical difficulty of the problem.

An example also show how this approach can deal with dynamical networks, where users join and leave the network. The observation of bandwidth allocation stationary distribution is sufficient to determine the link capacity, the set of server loads and the total arrival rate. This promising result needs to be generalized to other fairnesses and topologies.

Acknowledgment

This work was performed when the author was with INRIA, 23 avenue d'Italie, Paris. The author wants to thank Franois Baccelli for the research topic suggestion and support, and Alain Simonian for the remarks about the manuscript.

Appendix

Proof of Proposition 5

It is easy to check in (9) that

$$\prod_{i=1}^{K}(C_i - x\gamma_0(x))^\alpha = \sum_{i=0}^{\alpha K} a_{i,i}x^i\gamma_0(x)^i\,,$$

and $(C_1,\ldots,C_n)$ are the roots of the polynomial $\sum_{i=1}^{\alpha K} a_{i,i}X^i$.

Similar careful but standard computations then give

$$a_{l,l} = (-1)^l \sum_{\substack{j_1+j_2+\cdots+j_K=l\\ j_k\le\alpha}} \prod_{k=1}^{K} \binom{\alpha}{j_k} C_k^{\alpha-j_k}$$

$$a_{l,l+\alpha} = n_0^\alpha \binom{l+\alpha}{\alpha} a_{l+\alpha,l+\alpha} + (-1)^{1+l}$$

$$\sum_{i=1}^{K} w_i n_i^\alpha \sum_{\substack{j_1+\cdots+j_{i-1}+j_{i+1}+\cdots+j_K=l\\ j_k\le\alpha}} \prod_{k\neq i} \binom{\alpha}{j_k} C_k^{\alpha-j_k}. \quad (15)$$

The set of equations (15) consists of $\alpha K - \alpha + 1$ linear equations with $K+1$ values $(n_0^\alpha, w_1 n_1^\alpha, \ldots, w_K n_K^\alpha)$. For $\alpha > 1$ and $K > 1$, it is thus possible to determine the values of the flow numbers from these equations by solving the associated linear system. It is easy to see that the system will be regular if all capacities are pairwise different. When capacities of server S_i and S_j are equal, the only identifiable information is the "weighted α-power sum" $\left(w_i n_i^\alpha + w_j n_j^\alpha\right)$, which is similar to the identical capacities case. The case $K = 1$ falls into the "identical capacity" case. Finally, in the case when $\alpha = 1$, we have K equations with $K + 1$ unknowns. It allows us to express all the flow capacities as an affine function of n_0. As

$$a_{0,2} = n_0^2 \sum_{\{i,j\}} \prod_{\substack{k\neq i\\ k\neq j}} C_k + \sum_{i=1}^{K} w_i n_i \sum_{j\neq i} n_0 \prod_{\substack{k\neq i\\ k\neq j}} C_k\,. \quad (16)$$

we can rewrite (16) as a quadratic polynomial, which can be solved to determine n_0, and consequently all the flow numbers.

Proof of Proposition 6

Recall that equation (10) holds, for all probing intensities. We can rewrite it (10) in a vector form as

$$P \cdot A = (-1)^{1+\alpha K} Y\,, \quad (17)$$

where P is the $N \times \left(\frac{(\alpha K+1)(\alpha K+2)}{2} - 1\right)$ matrix, whose element of line k and column (i,j) [2] is $x_k^i\gamma_0(x_k)^j$, A the $\left(\frac{(\alpha K+1)(\alpha K+2)}{2} - 1\right) \times 1$ column matrix whose element of line (i,j) is $a_{i,j}$ (except $a_{K,K}$), and Y the $N \times 1$ column matrix whose k^{th} element is $x_k^{\alpha K}\gamma_0(x_k)^{\alpha K}$ (we force here the normalization $a_{K,K} = 1$). If $N = \frac{(\alpha K+1)(\alpha K+2)}{2} - 1$ and P is full-rank, there is unique solution A satisfying (17). We do not have any proof that P is full-rank; however, in all practical case we simulated, P was regular, though with some eigenvalues close to zero. If $N > \frac{(\alpha K+1)(\alpha K+2)}{2} - 1$ and P is full-rank, one can have a more robust estimation of A by minimizing the error $\|P \times A + (-1)^{\alpha K} Y\|$ for some well-chosen norm $\|.\|$ (a classical choice is the L^2 norm, which minimizes the quadratic error and leads to linear regression).

[2]For notation simplicity, we allow some columns or line to be numbered with a pair of integer, but one can obviously formally use a mapping between $\mathbb{N}^2$ and $\mathbb{N}$.

References

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

[2] N. Duffield, J. Horowitz, F. L. Presti, and D. Towsley, "Multicast Topology Inference from Measured End-to-End Loss," *IEEE Transactions in Information Theory*, vol. 48, no. 1, pp. 26–45, 2002.

[3] R. Caceres, N. Duffield, J. Horowitz, and D. Towsley, "Multicast-Based Inference of Network-Internal Loss Characteristics," *IEEE Transactions on Information Theory*, vol. 45, pp. 2462–2480, 1999.

[4] K. Lai and M. Baker, "Nettimer: a tool for measuring bottleneck link, bandwidth," in *Proc. of USENIX Symposium on Internet Technologies and Systems*, 2001, pp. 11–11.

[5] V. Jacobson, "Pathchar: A tool to infer characteristics of internet paths." 1997. [Online]. Available: ftp://ftp.ee.lbl.gov/pathchar/

[6] X. Liu, K. Ravindran, and D. Loguinov, "Multi-Hop Probing Asymptotics in Available Bandwidth Estimation: Stochastic Analysis," in *Proc. of ACM/USENIX Internet Measurement Conference*, October 2005.

[7] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson, "Netalyzr: illuminating the edge network," in *Proc of IMC*, 2010.

[8] "Samknows." [Online]. Available: http://www.samknows.com/broadband/index.php

[9] F. Baccelli, S. Machiraju, D. Veitch, and J. Bolot, "The Role of PASTA in Network Measurement," *Proc. of ACM SIGCOMM*, vol. 36, no. 4, pp. 231–242, 11-15 Sep 2006. [Online]. Available: http://www.cubinlab.ee.unimelb.edu.au/~darryl/Publications/antipasta_camera.pdf

[10] M. Roughan, "A Comparison of Poisson and Uniform Sampling for Active Measurements," *IEEE J. Selected Areas in Communication*, vol. 24, no. 12, pp. 2299–2312, Dec 2006.

[11] E. Lawrence, G. Michailidis, and V. N.Nair, "Statistical inverse problems in active network tomography," in *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond, IMS Lecture Notes-Monograph Series*. IMS, 2007, vol. 54, pp. 24–44.

[12] F. Pin, D. Veitch, and B. Kauffmann, "Statistical estimation of delays in a multicast tree using accelerated em," *Queueing Systems*, vol. 66, no. 4, pp. 369–412, 2010.

[13] B. Kauffmann, "Inverse problems in networks," Ph.D. dissertation, Universit Pierre et Marie Curie, 2011.

[14] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks:shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

[15] L. Massoulie and J. Roberts, "Bandwidth sharing: objectives and algorithms," in *Proc. IEEE Infocom*, vol. 3, mar 1999.

[16] J. W. Roberts and L. Massoulié, "Bandwidth sharing and admission control for elastic traffic," in *Telecommunication Systems*, 1998, pp. 185–201.

[17] P. Key, L. Massoulié, A. Bain, and F. Kelly, "Fair internet traffic integration: network flow models and analysis," *Annals of Telecommunications*, vol. 59, 2004. [Online]. Available: http://dx.doi.org/10.1007/BF03179724

[18] F. Baccelli, B. Kauffmann, and D. Veitch, "Inverse Problems in Queueing Theory and Internet Probing," *Queueing Systems*, vol. 63, no. 1–4, pp. 59–107, 2009.