Estimating One-Way Delays From Cyclic-Path Delay Measurements

Omer Gurewitz* and Moshe Sidi
Electrical Engineering Department
Technion, Haifa 32000, Israel
gurewitz@tx.technion.ac.il moshe@ee.technion.ac.il

Abstract—**In this paper we present a novel approach for the estimation of one-way delays from cyclic-path delay measurements that does not require any kind of synchronization among the nodes of the network. Furthermore, this approach is taking into account the asymmetric nature of the network, and the fact that traffic flows are not necessarily the same in both directions. Our approach is based on cyclic-path delay measurements, each of which is extracted using a single (source) clock and therefore is accurate. The basic idea of the approach is to express the cyclic-path delays in terms of one-way delay variables. If there were enough independent cyclic-path delay measurements, then one could solve explicitly for the one-way delays. We show that the maximal number of independent measurements that can be taken is smaller, hence a procedure for estimating the one-way delay is proposed.**

Keywords— **Cyclic-path delay; Round-trip delay; One-way delay; Estimation; Measurements**

I. Introduction

PROPER analysis of network-observed data and measurements is essential for robust network performance and management. Such real-data analysis plays a pivotal role in the design of the network and in the control of its dynamic behavior. One of the most important network performance quantities is the delay. Commonly, the round-trip delay is observed and measured. Yet, in recent years it became apparent that one-way delay measurements are very important. The measurement of one-way delay instead of round-trip delay is motivated by several factors [1]. In many networks the path from a source to a destination may be different than the path from the destination back to the source ("asymmetric paths"). Therefore round-trip measurements actually measure the performance of two distinct paths together. Measuring each path independently highlights the performance difference between the two paths which may traverse radically different types of networks. Even when the two paths are symmetric, they may have radically different performance characteristics due to asymmetric queueing. Performance of an application may depend mostly on the performance in one direction. For example, a file transfer using TCP may depend more on the performance in the direction that data flows, rather than the direction in which acknowledgements travel. Finally, in quality-of-service (QoS) enabled networks, provisioning in one direction may be radically different than provisioning in the reverse direction, and thus the QoS guarantees differ. Measuring the paths independently allows the verification of both guarantees.

Cyclic-path delay measurements in networks are relatively simple. They can be done by keeping a time-stamp for a packet both upon its transmission and upon its return. Since the time-stamps are taken at the same location, the same clock is used for both. Hence, the difference between these time-stamps yields an accurate measure for the cyclic-path delay (clock skews or drifts are negligible during this interval of measurements). Note that round-trip delay measurements are simple cyclic-path delay measurements. One-way delay measurements, on the other hand, are quite complex to measure as they require a perfect synchronization among the clocks at the source and the destination of the packet. Global Positioning Systems (GPS) afford one way to achieve such synchronization, but GPS are still very scarce in network environment. Ordinary applications of Network Time Protocol (NTP) are designed to allow synchronization, but this synchronization depends on the stability and symmetry of delay properties among the NTP agents at the source and the destination, and this is exactly the delay that should be measured.

A common approach for estimating one-way delays is to measure round-trip delays and halving them [6]. As explained above, such an estimation is reasonable only when the network is symmetric and the traffic load in both directions is the same. Several recent papers presented novel methods for delay evaluations that are based on end-to-end measurements ([2], [3], [4], [5]). These methods are very attractive for round-trip delay evaluations, but their validity for one-way delay estimations depends on perfect synchronization among all clocks in the network that are involved in the end-to-end measurements - a task hard to achieve.

In this paper we present a novel approach for the estimation of one-way delays from cyclic-path delay measurements that does not require any kind of synchronization among the nodes of the network. Furthermore, this approach is taking into account the asymmetric nature of the network, and the fact that traffic flows are not necessarily the same in both directions. Our approach is based on cyclic-path delay measurements, each of which is extracted using a single (source) clock and therefore is accurate. The basic idea of the approach is to express the cyclic-path delays in terms of one-way delay variables. If there were enough independent cyclic-path delay measurements, then one could solve explicitly for the one-way delays. We show that the maximal number of independent measurements that can be taken is smaller, hence a procedure for estimating the one-way delay is proposed.

The paper is organized as follows. In Section II we present the underlying model used throughout the paper and introduce the estimation problem. Section III contains the analysis of the procedure. In particular, we compute the maximal number of

* Omer Gurewitz is currently with Mellanox Technologies, Israel. E-mail: omer@mellanox.co.il

independent cyclic-path delay measurements that can be taken. We then propose in Section IV a distributed algorithm that yields the necessary number of measurements. Finally, several examples of the approach are give in Section V.

II. The model and preliminaries

The goal of this paper is to introduce and analyze a novel approach for estimating one-way delays between the nodes of a network, based on cyclic-path delay measurements among these nodes. We begin by introducing the network model that is used.

The network is composed of a set of nodes connected by some links. The nodes of the network that are relevant to our study are those nodes that are participating in the cyclic-path delay measurements, i.e., the nodes that serve as origins (and hence final destinations) for such measurements. Let $\mathcal{N}$ denote this set of nodes and let $N = |\mathcal{N}|$. In general, a path between any two nodes in $\mathcal{N}$ is a mix of links and nodes, some nodes are in $\mathcal{N}$ and some are not in $\mathcal{N}$. We define a directed *logical link* between node i and node j in $\mathcal{N}$ as a directed path between these nodes that does not contain any other node in $\mathcal{N}$. Let $\mathcal{E}$ denote the set of all logical links and $E = |\mathcal{E}|$.

From now on we will concentrate on the underlying network $(\mathcal{N}, \mathcal{E})$ that is composed of the nodes in $\mathcal{N}$ and whose links are the members of the set $\mathcal{E}$. The link connecting node i with node j in the direction from i to j is denoted by $i \rightarrow j$. We assume that if link $i \rightarrow j$ exists, so does link $j \rightarrow i$. Network delays are usually dynamic. Yet, when looking at short intervals of time, we can assume that they vary very slightly. Hence throughout this paper we assume that the delay on each link is constant. Note however that the delay of the link in the direction $i \rightarrow j$ is not necessarily identical to the delay of the link in the direction $j \rightarrow i$. The reason is that the links of each direction may traverse different network equipment (routers) in the actual network, or the load in each direction may be different.

Our goal is to provide an estimation for the one-way delay for each of the directed links. The estimation is based on cyclic-path delay measurements. These measurements are done in a straightforward manner. A source node is sending a probe packet that is forwarded along several (non-repeated) nodes until the packet returns to the source node (completes a cycle). To control the sequence of nodes that the packet traverses, source routing can be used for these probe packets. A better way for sending the probe packets is to use the distributed algorithm that is proposed in Section IV. Time is recorded by the source node both when the message is sent and also when the message returns. The difference between these two times is the cyclic-path delay along the path that the probe packet traversed. We assume that the time the packet spends at the intermediate nodes is either part of the one-way delay, or can be computed and be subtracted from the total cyclic-path time. Note that the cyclic-path delay of each such path can be determined in a single measurement due to our assumption of constant delay. Future work will extend our estimation procedure to determine delay distributions based on several measurements.

For each source node, several cyclic-path delay measurements can be taken, by sending probe packets along different such paths. One of our goals in this paper is to propose the algorithm for the source nodes to determine which probe packets to send. After obtaining all the cyclic-path delays through various paths we are ready to estimate the one-way delay between any two nodes based on these measured times. In principle, the estimation will depend on the criteria used. We choose a least square error criteria as explained below.

For each link $i \rightarrow j$ in $\mathcal{E}$, let $x_{i,j}$ be the one-way delay from i to j on that link (these are the quantities we are after). Let $\hat{x}_{i,j}$ be the estimate of $x_{i,j}$. We formulate the estimation problem as a constrained optimization problem. The variables are $\vec{x} = \{x_{i,j}\}$ (similarly, the estimates are $\vec{\hat{x}} = \{\hat{x}_{i,j}\}$) and the constraints are the cyclic-path delays measured and the non-negativity of the variables $\vec{x}$. To formally define these constraints, assume that $\mathcal{L}$ measurements are taken. Recall that each measurement is cyclic-path. Let $a_{l,\{i,j\}} = 1$ if link $\{i, j\}$ appears along the path of the l-th measurement and $a_{l,\{i,j\}} = 0$ otherwise. Let α_l be the measured cyclic-path delay in the l-th measurement. Then the measurement constraints are given by $\mathbf{A} \cdot \vec{\mathbf{x}} = \vec{\alpha}$ where $\mathbf{A}$ is a $\mathcal{L}$x$\mathcal{E}$ matrix whose elements are $\{a_{l,\{i,j\}}\}$ and $\vec{\alpha}$ is a vector whose elements are $\{\alpha_l\}$.

Let the set Ω define all the values of $\vec{x}$ with $x_{i,j} > 0$ that comply with these constraints. Clearly, this set is convex. Our goal is to determine $\vec{\hat{x}}$ that yields the least square error, or

$$\min \left\{ \int_{\Omega} |\vec{x} - \vec{\hat{x}}|^2 d\,\vec{x} \right\}$$

under the constraints:

$$\Omega = \{ \vec{x} \mid x_{i,j} > 0 \; ; \; \mathbf{A} \cdot \vec{\mathbf{x}} = \vec{\alpha} \}$$

Note that if any further information is available upon the $x_{i,j}$, it can be incorporated as additional constraints in the definition of Ω.

It is important to note that our estimation is based only on cyclic-path delay measurements. Hence both the "send" time and the "receive" time are measured at the same clock. Therefore clock synchronization of any kind between any two nodes in the system is not required.

III. Analysis

Our estimation of the one-way delays is based on measuring cyclic-path delays. How many such measurements can and should be taken? At first glance it appears that the more measurements are taken, the better. Yet, trying to measure all possible cyclic-path delays is rather tedious, and definitely not scalable. For example, in an N-node fully connected network (a network where every two nodes are connected via a bi-directional link), the number of cyclic-paths that start at a specific node and pass through only one node is $N - 1$. The number of such paths that start at the same node and pass through two intermediate nodes is $(N - 1) \cdot (N - 2)$, and so on. The total number of cyclic-paths that start at a specific node and pass through each intermediate node once is:

$$\sum_{i=2}^{N-1} \frac{(N-1)!}{(N-i)!}$$

and the total number of cyclic-paths starting at any node is therefore:

$$N \cdot \sum_{i=2}^{N-1} \frac{(N-1)!}{(N-i)!} > N \cdot (N-1)! \qquad \forall N > 2$$

Consequently, measuring all the possible cyclic-paths will result in much more than $N \cdot (N-1)$ constraints (or equations). However, the number of variables (delays on each link) is $N \cdot (N-1)$ in a fully connected network. There are many more equations than there are variables (number of cyclic-paths $>$ number of links), which means that there is a redundancy in the equation set. The question is how many independent cyclic-paths exist that yield independent constraints. If there were $N \cdot (N-1)$ independent cyclic-paths then the set of constraints (equations) could have been solved and the "actual" one-way delays of each link could have been obtained.

In theorem 1 we prove that in an N-node connected network, the maximal number of independent equations obtained by measuring cyclic-path delays is smaller than the number of links by $(N-1)$.

Theorem 1: The maximal number of independent equations obtained by measuring cyclic-path delays in an N-node connected network falls behind the number of variables (links) by $(N-1)$, i.e., the maximal number of independent equations is $E-(N-1)$.

In order to prove theorem 1 we need the following:

Lemma 2: Adding a node to an $(N-1)$-node connected network, the number of links that are added exceeds by one the maximal number of independent cyclic-paths traversing the additional node that can be constructed.

Proof: Assume that the N-th node that is added is connected to m nodes in the network ($1 \le m \le N-1$). Let the added links be denoted by $N{\to}1, 1{\to}N, \ldots, N{\to}m, m{\to}N$ corresponding to the paths from node N to node 1, node 1 to node N, etc., respectively (note that there are $2m$ additional links). To prove the Lemma we need to show that there are $(2m-1)$ new independent cyclic-paths traversing through node N.

As the $2m-1$ equations, we choose the delays of the m cyclic-paths which start at node N traversing over a single link to one of N's neighbors and return to node N (all the paths of the form $N{\to}k{\to}N \quad k = 1, 2, \ldots, m$ where k is one of the nodes connected to N via a link). Since N is connected to the network via m such links there are m such paths. In addition we choose $m-1$ paths, each of which starts at node N, passes through a specific neighboring node, let us call it node 1 without any loss of generality, via link $N{\to}1$, continues to an arbitrary node k ($k = 2, \ldots, m$) via one or more links (zero or more intermediate nodes), and returns to N via the link $k{\to}N$. Since the network of $N-1$ nodes (prior to adding the N-th node) is connected, there must be at least one path between the nodes 1 and k. If there is more than one path between such nodes we choose only one.

The rest of the proof of the Lemma consists of two parts. In the first part we show that knowing the delays of the $2m-1$ chosen cyclic-paths is sufficient to compute the delays of all the cyclic paths containing node N. Hence, the maximal number of independent paths traversing the additional node is $2m-1$. In the second part we show that the $2m-1$ paths are independent (we cannot compute the delay of neither one of them as a combination of the others).

We start proving the first part by showing that by using the $2m-1$ path delays we chose, we can compute the delay of any cyclic-path that passes through node N. Let us look at the delay of the cyclic-path: $\{N{\to}i\ \Theta_{i\to j}\ j{\to}N\}$ where $\Theta_{i\to j}$ represents a path that starts at node i and ends at node j (at least one path $\Theta_{i\to j}$ exists which does not pass through node N, since the network was connected before adding node N).

$$\begin{aligned} Delay\{N{\to}i\ \Theta_{i\to j}\ j{\to}N\} = \\ \underbrace{Delay\{N{\to}i\ \Theta_{i\to 1}\ 1{\to}N\}}_{i} \\ + \underbrace{Delay\{N{\to}1\ \Theta_{1\to j}\ j{\to}N\}}_{ii} \\ + \underbrace{Delay\{\Theta_{1\to i}\ \Theta_{i\to j}\ \Theta_{j\to 1}\}}_{iii} \\ - \underbrace{Delay\{\Theta_{1\to i}\ \Theta_{i\to 1}\}}_{iv} \\ - \underbrace{Delay\{\Theta_{1\to j}\ \Theta_{j\to 1}\}}_{v} \\ - \underbrace{Delay\{N{\to}1{\to}N\}}_{vi} \end{aligned}$$

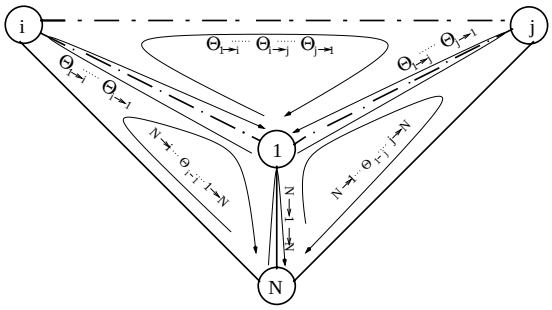


Fig. 1. Illustration of the proof

The various paths are depicted in Figure 1. Parts (ii) and (vi) are two out of the $2m-1$ chosen cyclic-paths. Parts (iii), (iv) and (v) are cyclic-paths which do not include node N, hence their delays are known. Part (i) is known since if we know the delay of a path in one direction and the round-trip delays on each link comprising the path, we can compute the delay of the other path direction simply by subtracting the round-trip delays on each link from the delay of the path:

$$\begin{aligned} Delay\{k{\to}j{\to}\cdots{\to}2{\to}1\} = \\ Delay\{1{\to}2{\to}\cdots{\to}j{\to}k\} \\ -Delay\{1{\to}2{\to}1\} - \cdots - Delay\{j{\to}k{\to}j\} \end{aligned}$$

We showed that we can compute the delay of any cyclic-path that starts (and terminates) at node N. It is obvious that the delay of any cyclic-path that starts at some node z and passes

through node N is the same as the matching cyclic-path starting at N, since in both cases the same links are traversed in the same direction exactly once. This completes the first part of the lemma proof.

Next we show that the $2m-1$ paths are independent. The proof follows by contradiction. Suppose that one cyclic-path can be eliminated and yet the delay of the eliminated path can be computed as a combination of the rest of the set of chosen paths. If the eliminated path is of the form $N{\rightarrow}k{\rightarrow}N$, it means that none of the remaining cyclic-paths pass through the link $N{\rightarrow}k$ (for $k=1$ there is no path passing through the link $1{\rightarrow}N$). Hence the variable $x_{N,k} = Delay\{N{\rightarrow}k\}$ does not appear in any of the equations and hence is impossible to compute. If the eliminated path is of the form $N{\rightarrow}1{\rightarrow}\Theta_{1\rightarrow j}{\rightarrow}k{\rightarrow}N$ we will have that the two variables $x_{N,k} = Delay\{N{\rightarrow}k\}$ and $x_{k,N} = Delay\{k{\rightarrow}N\}$ appear only together and hence cannot be separated which makes it impossible to compute any cyclic-path including only one of them.

This concludes the proof. The set of $(2m-1)$ cyclic-paths is independent, and there is no redundant path.

We now turn to prove Theorem 1.

Proof: The proof of the theorem follows by induction. We start by constructing a network of two nodes connected via a bi-directional link, meaning that we start with two links and one round-trip path (one degree of freedom). In each step we add one node that is connected to the previous step sub-network by at least one bi-directional link. We add the node and all the links connected to the previous step sub-network. According to lemma 2, in each step we add one degree of freedom. So if in step i we added node i and the e_i links which connect it to the step $(i-1)$ sub-network consisting of nodes $1 \ldots i-1$, we added only $|e_i|-1$ independent paths. The construction of the network $(\mathcal{N},\mathcal{E})$ will last $N-1$ steps, hence the degree of freedom is $N-1$.

Theorem 1 implies directly that in an N-node *connected* network $(\mathcal{N},\mathcal{E})$, using a correct choice of $(N-1)$ links and $E-(N-1)$ cyclic-paths whose delays are measured, we can represent the one-way delays of all the E links. We also showed one particular set of $(N-1)$ links and $E-(N-1)$ cyclic-paths that describe the rest of the links.

The next theorem states that instead of minimizing the function $\sum_{all\ links}(x_i - \hat{x_i})^2$ over all links, we can minimize a similar function only upon specifically chosen $(N-1)$ links of an N-node connected network. To state the theorem, denote by $w_1, w_2, \ldots, w_{N-1}$ the variables that correspond to the one-way delays of the $N-1$ chosen links. Similarly, let $\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_{N-1}$ be their respective estimates.

Theorem 3: The target function

$$\sum_{all\ links}(x_{ij} - \hat{x_{ij}})^2$$

can be presented as a function of the $(N-1)$ one-way delays of the chosen links as follows:

$$\sum_{k=1}^{N-1}\sum_{l=1}^{N-1} D_{k,l}\cdot(w_k-\hat{w}_k)(w_l-\hat{w}_l) \qquad (1)$$

where $D_{k,l}$ are constants.

Proof: According to theorem 1, in the N-node connected network $(\mathcal{N},\mathcal{E})$, there are E links (their delays are our variables), and only $E-(N-1)$ independent cyclic-paths (which are the set of independent equations). Clearly, each one-way link delay $x_{i,j}$ can be presented as a linear combination of the chosen $(N-1)$ independent one-way link delays:

$$x_{i,j} = \gamma_{i,j} + \sum_{k=1}^{N-1} b_{i,j}^{(k)} w_k$$

where $b_{i,j}^{(k)}$ are integers and $\gamma_{i,j}$ are constants. We take our estimates $\hat{x}_{i,j}$ to have the same form, i.e.,

$$\hat{x}_{i,j} = \gamma_{i,j} + \sum_{k=1}^{N-1} b_{i,j}^{(k)} \hat{w}_k$$

Now let us concentrate on one element in the sum of the target function, namely let us develop $(x_{i,j}-\hat{x}_{i,j})^2$:

$$\begin{aligned}(x_{i,j}-\hat{x}_{i,j})^2 &= \left(\gamma_{i,j} + \sum_{k=1}^{N-1} a_{i,j}^{(k)} w_k \right. \\ &\quad \left. - \ \gamma_{i,j} - \sum_{k=1}^{N-1} a_{i,j}^{(k)} \hat{w}_k\right)^2 \\ &= \left(\sum_{k=1}^{N-1} a_{i,j}^{(k)} (w_k - \hat{w}_k)\right)^2 \\ &= \sum_{k=1}^{N-1}\sum_{l=1}^{N-1} a_{i,j}^{(k)} a_{i,j}^{(l)} (w_k-\hat{w}_k)(w_l-\hat{w}_l)\end{aligned}$$

Using the last result and summing over all links will result in Eq. (1). From this derivation it is also clear that $D_{k,l} = D_{l,k}$ for all l, k.

An example of theorem 3 is the target function of the fully connected network for which $D_{k,k} = 2(N-1)\ \ \forall k$ and $D_{k,l} = 2\ \ \forall k \neq l$.

In order to complete the analysis, we still have to find

$$\min\left\{\int_\Omega |\vec{x}-\vec{\hat{x}}|^2 d\,\vec{x}\right\}$$

which according to theorem 3 is of the form:

$$\min\left\{\int_\Omega \sum_{k=1}^{N-1}\sum_{l=1}^{N-1} D_{k,l}\cdot(w_k-\hat{w}_k)(w_l-\hat{w}_l) d\vec{w}\right\}$$

Let us partially differentiate the above with respect to each variable $\hat{w}_p$, and equate it to zero.

$$\frac{\partial}{\partial \hat{w}_p}\left\{\int_\Omega \left(\sum_{k=1}^{N-1}\sum_{l=1}^{N-1} D_{k,l}\cdot(w_k-\hat{w}_k)(w_l-\hat{w}_l)\right) d\vec{w}\right\} = 0$$

Using the fact that $D_{k,l} = D_{l,k}$ we obtain:

$$\int_\Omega \left(D_{p,l}\cdot\sum_{l=1}^{N-1}(w_l-\hat{w}_l)\right) d\vec{w} = 0$$

or

$$\hat{w}_l = \int_\Omega \frac{1}{\int_\Omega d\vec{w}} w_l d\vec{w} \quad 1 \le l \le N-1 \qquad (2)$$

Note that this solution is unique due to the convexity of Ω. It is interesting to see that the best estimate $\hat{w}_l$ is some kind of averaging of w_l over Ω.

From the derivation in this section it is clear that one can apply the estimation procedure even if the number of cyclic-path delay measurements that are taken is smaller than the maximal number $(E-(N-1))$. The solution would be exactly as in Eq. (2), with the modified set Ω. In the extreme case where only round-trip delay measurements among neighboring nodes will be taken, the result of the estimation above for the one-way delay would be halving the round-trip delay.

IV. A Distributed Algorithm

In this section we present a distributed network algorithm that conducts the measurement of the cyclic-path delays in $E-(N-1)$ independent paths that yield the constraints necessary for the estimation procedure. Designing a centralized algorithm that performs the same measurements is trivial, since we suggested in theorem 1 a particular set of $E-(N-1)$ independent cyclic-paths.

We consider a connected network $(\mathcal{N}, \mathcal{E})$ and assume that each node has a unique identity. Also, each node knows its adjacent links, and the identity of its immediate neighbors.
The algorithm uses the following messages:
$MSG_l^S(\Lambda, d)$- message initiated at node S, received from neighbor l and Λ is the list of nodes that the message visited; d serves as a flag which indicates whether MSG^S is going downstream or upstream.
msg^k -message initiated at node k. Relates to the part of the algorithm which measures round-trip delay on a single link.
The algorithm uses the following variables:
G_i - set of neighbors of node i
m_i^S - indicates whether i has already entered the algorithm (values 0,1)
p_i^S - neighbor from which the first MSG^S is received
λ_i - sublist of Λ including all nodes that appear after node i in the list.
$T_i(start)$- time first MSG^S is received
$T_i(stop)$- current time

The algorithm

Initialization

if i receives a MSG_l^S, then
- just before receiving the first $MSG^S(\Lambda)$, $m_i^S = 0$; $p_i^S = nil$
- after receiving the first MSG^S, node i can receive only MSG^S's sent in the present protocol

Algorithm for node $i \neq$ S

For $MSG_l^S(\Lambda, d)$
  if $m_i^S = 0$, then:
    $T_i(start) \leftarrow$Time
    $m_i^S \leftarrow 1$; $p_i^S \leftarrow l$; add ID(i) to Λ;
    send $MSG^S(\Lambda)$ to all neighbors except p_i^S
    send msg^i to all neighbors with $ID(l) > ID(i)$ $\ l \in G_i$
  else if $((ID(l) < ID(i))\ or\ (d == 1))$
   if $i \in \Lambda$
    $T_i(stop) \leftarrow$Time
    $Delay\{i \to \Theta_{\lambda_i} \to i\} = T_i(start) - T_i(stop)$
   else if $i \notin \Lambda$
    add ID(i) to end of list $\Lambda = \{\Lambda, ID(i)\}$;
    send $MSG^S(\Lambda, d = 1)$ to p_i^S

For msg^l
  if $ID(l) > ID(i)$ (Response message to a message initiated by i), then:
    $T_i(stop) \leftarrow$Time
    $Delay\{i \to l \to i\} = T_i(start) - T_i(stop)$
  else $ID(l) < ID(i)$
    send msg^i to neighbor l

Algorithm for node S

For START
    $m^S \leftarrow 1$;
    $T_S(start) \leftarrow$Time;
    send $MSG^S(\Lambda = \{S\};\ d = 0)$ to one neighbor.
    send msg^S to all neighbors with $ID(l) > ID(i)$ $\ l \in G_S$

For $MSG_l^S(\Lambda, d)$
    $T_S(stop) \leftarrow$Time
    $Delay\{S \to \Theta_\Lambda \to S\} = T_S(start) - T_S(stop)$

For msg^l
  if $ID(l) > ID(S)$ then:
    $T_S(stop) \leftarrow$Time
    $Delay\{S \to l \to S\} = T_S(start) - T_S(stop)$
  else
    send msg^S to neighbor l

V. Numerical Solution Method

The solution of Eq. (2) might become complicated, mainly due to the constraints. In this section we develop a numerical procedure that is based on approximating the integral by a sum. We have to sum over all vectors which are in the region Ω. In other words we have to sum over all the vectors $\vec{x}$ such that $x_{i,j} \ge 0$, and solve the equation $\mathbf{A} \cdot \vec{\mathbf{x}} = \vec{\alpha}$. The resolution can be as fine as desired trading off running time. In order to find the vectors we add to the equation set $N-1$ independent equations which relate to the $N-1$ independent variables. Let us label them as $(w_1, w_2, \ldots, w_{N-1})$ (note that the w's are actually part of our variables $x_{i,j}$). The $N-1$ equations we add are: $w_1 = \beta_1, w_2 = \beta_2, \ldots, w_{N-1} = \beta_{N-1}$.

We have now a set of E equations with E variables. In matrix form we can write it as $\mathbf{B} \cdot \vec{\mathbf{x}} = \vec{\beta}$ where the ExE matrix $\mathbf{B}$ and the vector $\vec{\beta}$ are:

$$\mathbf{B} = \begin{pmatrix} & & & \mathbf{A} & & & & \\ 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ & & & & \vdots & & & \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}$$

$$\vec{\eta} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{N-1} \\ \beta_1 \\ \vdots \\ \beta_{N-1} \end{pmatrix}$$

Since $\mathbf{B}$ is a non-singular matrix (all equations are independent), we can find the inverse matrix $\mathbf{B^{-1}}$, and solve the equation $\vec{x} = \mathbf{B^{-1}} \cdot \vec{\eta}$. Now all we have to do is choose $(\beta_1, \beta_2, \cdots, \beta_{N-1})$, construct the vector $\vec{\eta}$, multiply it from the left with $\mathbf{B^{-1}}$, and check if all the elements of the derived $\vec{x}$ are non-negative $(x_{i,j} \geq 0,\ \forall x_{i,j})$ then $\vec{x} \in \Omega$. Of course we do not have to check over all $\vec{\beta} = (\beta_1, \beta_2, \cdots, \beta_{N-1})$. Since $x_{i,j} \geq 0$ we should check only $\vec{\beta}$ where all elements are non-negative. In addition, the w's are part of the $x_{i,j}$ hence each of them obey an equation of the form $w_i + Delay\{\cdot\} = \alpha_l$ for some l. Therefore, we can restrict $\vec{\beta}$ only to $\vec{\beta}$ where $\beta_i \leq \alpha_l$. We can go further and narrow the region of possible $\vec{\beta}$ (for example β_i should be no bigger than the delay of any path passing through the link w_i). Notice that the matrix $\mathbf{B^{-1}}$ should be computed only once.

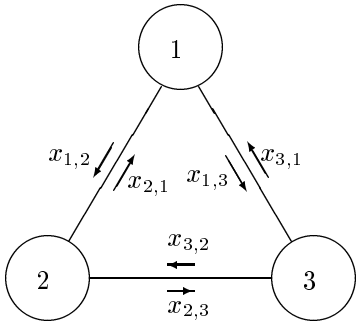
Now after we know all the points $\vec{x} \in \Omega$, we can compute $\hat{\vec{x}}$ according to Eq.(2):

$$\hat{\vec{x}} = \frac{1}{m} \sum_{\vec{x}_r \in \Omega} \vec{x}_r$$

where m is the number of points in Ω.

Example 1

Consider the following simple example of a fully connected 3-node network.



The following cyclic-path delays were measured:

$$\begin{aligned} x_{1,2} + x_{2,1} &= 50 \\ x_{2,3} + x_{3,2} &= 230 \\ x_{3,1} + x_{1,3} &= 50 \\ x_{1,2} + x_{2,3} + x_{3,1} &= 30 \end{aligned}$$

We have a set of $E - (N-1) = 6 - 2 = 4$ equations with $N \cdot (N-1) = 6$ variables. The matrix $\mathbf{A}$ and the vector $\vec{\alpha}$ are respectively:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

$$\vec{\alpha} = \begin{pmatrix} 50 \\ 230 \\ 50 \\ 30 \end{pmatrix}$$

We add the following $N - 1 = 2$ equations:

$$\begin{aligned} w_1 &= x_{1,2} = \beta_1 \\ w_2 &= x_{2,3} = \beta_2 \end{aligned}$$

$\mathbf{B}$ and $\vec{\eta}$ are:

$$\mathbf{B} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\vec{\eta} = \begin{pmatrix} 50 \\ 230 \\ 50 \\ 30 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

$\mathbf{B}$ is non singular, hence $\mathbf{B^{-1}}$ can be found:

$$\mathbf{B^{-1}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{pmatrix}$$

Now we can solve the equation $\vec{x} = \mathbf{B^{-1}}\vec{\eta}$ for each β_1 and β_2. Search for all $\vec{x}$ such that all elements are non-negative $(x_{i,j} \geq 0)$ to find all $\vec{x} \in \Omega$. Since $x_{1,2} + x_{2,3} + x_{3,1} = 30 \Rightarrow \beta_1 + \beta_2 + x_{3,1} = 30 \Rightarrow 0 \leq \beta_1 \leq 30$, $0 \leq \beta_2 \leq 30$, we can limit the search only to β_i in the range above. After we find all $\vec{x} \in \Omega$ according to the resolution on which we decided to check β_1 and β_2, we can compute

$$\hat{\vec{x}} = \frac{1}{m} \sum_{\vec{x}_r \in \Omega} \vec{x}_r$$

which results in:

$$\hat{x}_{1,2} = 10\ ;\ \hat{x}_{2,1} = 40\ ;\ \hat{x}_{2,3} = 10$$

$$\hat{x}_{3,2} = 220\ ;\ \hat{x}_{3,1} = 10\ ;\ \hat{x}_{1,3} = 40$$

Note that based only on single link round-trip delays, and halving the measured round-trip delays on each link, will result in:

$$\hat{x}_{1,2} = 25\ ;\ \hat{x}_{2,1} = 25\ ;\ \hat{x}_{2,3} = 115$$

$$\hat{x}_{3,2} = 115\ ;\ \hat{x}_{3,1} = 25\ ;\ \hat{x}_{1,3} = 25$$

which is definitely not the correct one-way delays.

Example 2

Consider the following example of a 10-node $(N = 10)$ connected network with $E = 24$ links.

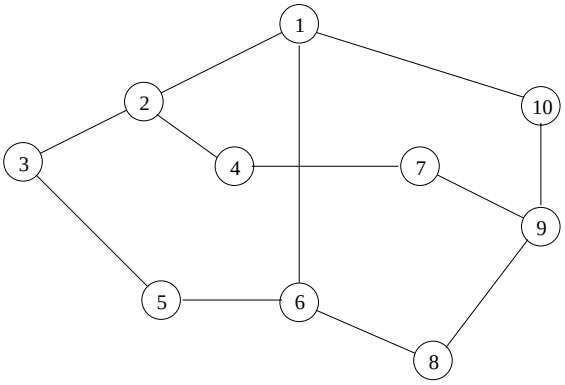


Fig. 2. Network for example 2

The following cyclic-path delays were measured:

$$\begin{array}{ll} x_{1,2} + x_{2,1} = 5 & x_{2,3} + x_{3,2} = 5 \\ x_{3,5} + x_{5,3} = 5 & x_{5,6} + x_{6,5} = 5 \\ x_{2,4} + x_{4,2} = 5 & x_{4,7} + x_{7,4} = 5 \\ x_{6,8} + x_{8,6} = 5 & x_{8,9} + x_{9,8} = 5 \\ x_{7,9} + x_{9,7} = 5 & x_{9,10} + x_{10,9} = 5 \\ x_{6,1} + x_{1,6} = 5 & x_{10,1} + x_{1,10} = 5 \end{array}$$

$$\begin{aligned} x_{1,2} + x_{2,4} + x_{4,7} + x_{7,9} + x_{9,10} + x_{10,1} &= 24 \\ x_{1,2} + x_{2,3} + x_{3,5} + x_{5,6} + x_{6,1} &= 18 \\ x_{1,6} + x_{6,8} + x_{8,9} + x_{9,10} + x_{10,1} &= 19 \end{aligned}$$

Note that the maximal independent cyclic-path delay measurements that can be taken is $E - (N - 1) = 15$ in this example.

Applying our estimation procedure for this example we obtain:

$$\begin{array}{llll} \hat{x}_{1,2} = 4.35 & \hat{x}_{2,1} = 0.65 & \hat{x}_{2,3} = 3.74 & \hat{x}_{3,2} = 1.26 \\ \hat{x}_{3,5} = 3.74 & \hat{x}_{5,3} = 1.26 & \hat{x}_{5,6} = 3.74 & \hat{x}_{6,5} = 1.26 \\ \hat{x}_{2,4} = 3.63 & \hat{x}_{4,2} = 1.37 & \hat{x}_{4,7} = 3.63 & \hat{x}_{7,4} = 1.37 \\ \hat{x}_{6,8} = 3.83 & \hat{x}_{8,6} = 1.17 & \hat{x}_{8,9} = 3.83 & \hat{x}_{9,8} = 1.17 \\ \hat{x}_{7,9} = 3.63 & \hat{x}_{9,7} = 1.37 & \hat{x}_{9,10} = 4.38 & \hat{x}_{10,9} = 0.62 \\ \hat{x}_{6,1} = 2.43 & \hat{x}_{1,6} = 2.57 & \hat{x}_{10,1} = 4.38 & \hat{x}_{1,10} = 0.62 \end{array}$$

As with the previous example we observe that halving round-trip measurements would result in a one-way delay of 2.5, which is incorrect.

REFERENCES

[1] G. Almes, S. Kalidindi and M. Zekauskas, "A One-way Delay Metric for IPPM," RFC 2679, September 1999.

[2] R. Caceres, N.G. Duffield, J. Horowitz, D. Towsley, Multicast-based Inference of Network-Internal Loss Characteristics, *IEEE Transactions on Information Theory*, November 1999.

[3] R. Caceres, N. Duffield, J. Horowitz, D. Towsley, T. Bu, "Multicast-Based Inference of Network-Internal Characteristics: Accuracy of Packet Loss Estimation," *IEEE Infocom'99*, New York, March 1999.

[4] N. Duffield, F. Lo Presti, "Multicast Inference of Packet Delay Variance at Interior Network Links," *IEEE Infocom 2000*, Tel-Aviv, March 2000.

[5] F. LoPresti, N.G. Duffield, J. Horowitz, D. Towsley, "Multicast -Based Inference of Network-Internal Delay Distribution," UMass CMPSCI 99-55, 1999.

[6] V. Ozdemir, S. Muthukrishnan, I. Rhee, "Scalable, Low-Overhead Network Delay Estimation," *IEEE Infocom 2000*, Tel-Aviv, March 2000.