ESTIMATION OF POINT-TO-POINT TRAFFIC DEMAND IN DYNAMIC ROUTING NETWORKS

Martin T.

Bellcore, P.O. Box 7050
100 Schultz Drive, Red Bank, New Jersey 07701-7050 USA

An accurate measurement or estimate of point-to-point traffic demand is essential to a cost-effective design of any communications network, especially for dynamic routing networks. In this paper, we present an estimation algorithm that addresses this need. This algorithm can be applied to networks with a sequential, dynamic routing scheme where the routing decisions are updated periodically with period $\Delta T$. The underlying network can be hierarchical or nonhierarchical, and $\Delta T$ can range from months to seconds. It uses primarily the trunk group measurements, but other network measurements can be included as well.

1. INTRODUCTION

An accurate measurement or estimate of point-to-point (P-P) traffic demand is essential to developing a P-P traffic demand forecast that can be used to design, engineer, and service communications networks. This is especially, but not uniquely, true for dynamic routing networks. In fact, it is anticipated that P-P data will be required for designing and maintaining any of the more dynamic networks of the future, including those based on ISDN, software-defined networks, and broadband technologies.

However, at present, few switches in the United States local exchange networks can directly measure such P-P traffic demand. The most common network measurements made by the existing switches are trunk group based measurements. Therefore, the estimation of P-P traffic demand based on the trunk group measurements has been a pressing issue for network designers for many years. As the dynamic routing technology is developed and implemented, this issue becomes even more important since the P-P traffic demand is a vital input to the dynamic routing control algorithms.

In this paper, we present an estimation algorithm that addresses this need. This algorithm can be applied to networks with a sequential, dynamic routing scheme where the routing decisions are updated periodically with period $\Delta T$. The underlying network can be hierarchical or nonhierarchical, and $\Delta T$ can range from months to seconds. Inputs include the trunk group measurements and the routing sequence in each $\Delta T$ interval. It turns out that the estimation of the P-P traffic demand is equivalent to the solving of a set of linear equations: The P-P traffic demand is the set of unknowns, and each trunk group usage measurement establishes an equation for these unknowns. As a consequence, the P-P traffic demand can be more accurately estimated for fully connected networks. For networks not fully connected, office totals can also be used as supplementary measurements.

In the remainder of this section, we briefly describe the routing scheme to which this estimation algorithm can be applied, and formulate the problem statement. In Section 2, we describe the notations and assumptions on which our analysis will be based. Section 3 contains traffic analysis on load and blocking. Two scenarios are considered: with and without the crankback capability at the via nodes. The estimation algorithm is presented in Section 4. In Section 5, we discuss the performance of the algorithm. In Section 6, examples are provided for illustration. Finally, our concluding remarks are presented in Section 7.

1.1 The Routing Scheme

The routing scheme that we consider here is of a
general class. DR5 [1] and DNHR [2], for example, belong to this class. In fact, any sequential, dynamic routing scheme that can be described in the following way belongs to this class.

The routing paths for each point-pair are updated periodically, with period $\Delta T$. At each update,

(a) The paths for each point-pair are ranked in a preference order. A minimum of one and a maximum of $M$ paths are selected.

(b) Until the next update, the routing rule consists in attempting the corresponding path-sequence selected in (a).

1.2 Problem Statement

Consider a network on which the routing scheme defined above is implemented. The following information is available periodically at each $\Delta T$ time interval:

- the trunk group measurements, which include the carried load (usage) and blocking (overflow to peg count ratio) on each trunk group, and
- the (fixed) routing sequence during the $\Delta T$ period.

Our objective is to estimate the P-P offered load and P-P blocking for each point-pair during the $\Delta T$ period.

2. NOTATIONS AND ASSUMPTIONS

We adopt the following notations for the network:

(a) $N$ = the number of nodes
$L$ = the number of links
$P$ = the number of point-pairs

(b) $x_i$ = the offered load of point-pair $i$
$B_i$ = the P-P blocking of point-pair $i$
$\alpha_i$ = the carried load of point-pair $i$

\[ \alpha_i = x_i(1-B_i) \] (1)

(c) For each point-pair, we allow at most $M$ sequential routes/paths. Each route/path may consist of one or two links.

(d) Path $ij$ ($i=1,2,\ldots,P$ and $j=1,2,\ldots,M$) denote the jth route for point-pair $i$.

(e) $OL_{ij}$ = the point-pair i load offered to path $ij$;
$PB_{ij}$ = the path blocking of path $ij$;
$CL_{ij}$ = the point-pair i load carried on path $ij$;

\[ CL_{ij} = OL_{ij}(1-PB_{ij}) \] (2)

(f) Path link $ijk$ denotes the kth link of path $ij$. Notice that

- $k=1$ or 2 since we consider only 1-link and 2-link paths.
- Each path link $ijk$ corresponds to a unique network link $l$ ($l=1,2,\ldots,L$), but each network link $l$ may correspond to a number of path links $ijk$. We denote this relationship by a mapping $R$, i.e.,

\[ R(ijk) = l, \quad l=0,1,2,\ldots,L \] (3)

$l$ is allowed to be zero so that

- If $R(ij1)=0$, it means that point-pair $i$ has at most $j-1$ routes.
- If $R(ij2)=0$, it means that the jth route for point-pair $i$ is a 1-link path.

(g) $LB_{ijk}$ = the link blocking of link $ijk$.

- $LB_{ij1} = 1$ if $R(ij1)=0$, and
- $LB_{ij2} = 0$ if $R(ij2)=0$.

(h) $\gamma_l$ = the total load carried on link $l$

The network is assumed to be a "lost calls cleared" system. In addition, we make two key assumptions in our analysis:

- On each link, calls from different point-pairs see the same blocking, which is the measured trunk group blocking.
- If a call is routed on a 2-link path, then we assume the event that this call will be blocked on one link is independent of the event that it will be blocked on the other link.

3. LOAD AND BLOCKING ANALYSIS

Because of our last assumption on the independence of call blocking on each link of a path, the path blocking is a simple function of its link blocking, i.e.,

\[ PB_{ij} = LB_{ij1} + LB_{ij2} - LB_{ij1}LB_{ij2} \] (4)

The P-P blocking and P-P offered load can be estimated by analyzing the load "offered to", and carried by each path $ij$, for each point-pair $i$. However, the load "offered to" each path will be different in two scenarios: when there is the "crankback" [3] capability at the via nodes and when there is not.
3.1 When There Is Crankback

When there is the crankback capability at the via nodes, then, if a point-pair i call is blocked on either link of its jth path (path iD), it will be routed to the next path in the routing sequence, when it exists. Let us derive $O_{ij}$ for each point-pair i and its jth path. For path i1,

$$O_{i1} = x_i$$  \hspace{1cm} (5)

since it is the first choice route for point-pair i. For path i2,

$$O_{i2} = x_iP_{i1}$$  \hspace{1cm} (6)

since it is the load rejected by path i1. In general,

$$O_{ij} = x_i \prod_{t=1}^{j-1} P_{it}$$  \hspace{1cm} (7)

From Eqs. (2) and (7), we obtain

$$C_{ij} = S_{ij}x_i$$  \hspace{1cm} (8)

where

$$S_{ij} \triangleq \left( \prod_{t=1}^{j-1} P_{it} \right) (1 - P_{ij})$$  \hspace{1cm} (9)

which is the proportion of point-pair i load carried on its jth route.

Now, based on the above analysis, we can synthesize two summation terms which will lead to the determination of the P-P blocking and the P-P offered load.

The first one is the sum of $C_{ij}$'s over j, which is the total point-pair i load carried on all its paths. Hence,

$$\alpha_i = \sum_{j=1}^{M} C_{ij} = \left( \sum_{j=1}^{M} S_{ij} \right) x_i$$  \hspace{1cm} (10)

From Eq. (1), we obtain

$$B_i = 1 - \sum_{j=1}^{M} S_{ij}$$  \hspace{1cm} (11)

Next, we need to calculate the total load carried on each link i from all point-pairs. This can be achieved by adding all point-pair i load carried on links ijk such that $R(ijk) = i$. Notice that the point-pair i load carried on link ijk is nothing but $C_{ij}$. Therefore, the second sum of $C_{ij}$'s is

$$y_i = \sum_{R(ijk) = i} C_{ij} = \sum_{R(ijk) = i} S_{ij} x_i$$  \hspace{1cm} (12)

Since there are L links, we have L such equations, one for each link. The P-P offered load $x_i$ can thus be obtained by solving these equations.

3.2 When There Is No Crankback

When there is no crankback capability at the via nodes, if a point-pair i call is blocked on the jth route (path iD), then this call will be routed to the next route in the routing sequence only if the call is blocked on the first link of path iD; this call will be abandoned if it is blocked on the second link.

In this situation, Eq. (5) remains valid for path i1. For path i2

$$O_{i2} = x_iL_{i11}$$  \hspace{1cm} (6')

since it is the load rejected by link i11, the first link of path i1. In general,

$$O_{ij} \triangleq x_i \prod_{t=1}^{j-1} L_{rt1}$$  \hspace{1cm} (7')

Eqs. (8), (10), (11), and (12) remain unaffected if $S_{ij}$ is redefined as follows.

$$S_{ij} \triangleq \left( \prod_{t=1}^{j-1} L_{rt1} \right) (1 - P_{ij})$$  \hspace{1cm} (9')

4. THE ESTIMATION ALGORITHM

The estimation of P-P blocking is straightforward. Given the blocking probability on each link ($L_{ijk}$), the path blocking ($P_{ij}$) can be derived from Eq. (4). The P-P blocking ($B_i$) can then be estimated using Eq. (11).

The estimation of P-P offered load ($x_i$) is not straightforward, although its relationship with the carried load on each link ($y_i$) is shown by Eq. (12). Eq. (12) for the L links can be written systematically as follows.

$$y = Hx$$  \hspace{1cm} (13)

where

$$y \triangleq \begin{bmatrix} y_1 & y_2 & \cdots & y_L \end{bmatrix}^T$$

$$x \triangleq \begin{bmatrix} x_1 & x_2 & \cdots & x_P \end{bmatrix}^T$$

$$H = [h_{ij}]$$

$$h_{ij} \triangleq \begin{cases} \frac{S_{ij}}{S_{ijr}} & \text{if link i is a link of the } r\text{th path of point pair } j \\ 0 & \text{otherwise} \end{cases}$$

Pseudo-inverse and various iterative methods can be applied to solve Eq. (13).
4.1 Pseudo-Inverse Method

The estimated P-P offered load, denoted by \( \hat{x} \), is given by

\[
\hat{x} = H^+ y
\]  

(14)

where \( H^+ \) is the pseudo-inverse \(^4\) of \( H \). To explain what \( H^+ \) is, some discussion on the \( L \times P \) matrix \( H \) is in order.

(a) For underdetermined systems \((L < P)\)

It is usually the case that the \( L \) rows are linearly independent. This is because each \( Y_I \), the carried load on trunk group \( I \), is a unique linear combination of the offered load from all the point-pairs. In this case, \( H^+ \) is defined by

\[
H^+ \triangleq H^+(HH^T)^{-1}
\]

(15)

(b) For square systems \((L = P)\)

\[
H^+ = H^{-1}
\]

(16)

(c) For overdetermined systems \((L > P)\)

This can happen, for example, when other network measurements, such as the office totals, are added to Eq. (13) (see Section 5). When \( L > P \), it is usually the case that the \( P \) columns are linearly independent. This is because each column \( i \) represents the unique routing paths selected for point-pair \( i \). In this case, \( H^+ \) is defined by

\[
H^+ \triangleq (H^T H)^{-1} H^T
\]

(17)

In any case, \( \hat{x} \) is the optimal estimate of \( x \), in the least-squares sense, based on the available measurements.

The memory needed for storing the \( H \) matrix is tremendous for large networks. The same is true for the computation and storage of \( H^+ \). To avoid such difficulties, iterative methods can be used instead to solve Eq. (13).

4.2 Iterative Methods

The basic idea of iterative methods is to start with an \( x^{(0)} \), an initial estimate of \( x \). Then, from the current estimate \( x^{(k)} \), a new estimate \( x^{(k+1)} \) is generated by considering the difference between \( y \) and \( Hx^{(k)} \). This process terminates when the difference between \( y \) and \( Hx^{(k)} \) is within a preset limit. In this iterative process, the \( x^{(k)} \)'s are generated one by one for each \( i \). Therefore, no matrix inversion is needed.

The iterative method that we used in testing the algorithm is a modified version of the Gauss-Seidel iteration \([5]\):

For \( i = 1, \ldots, P \), do the following:

Choose \( j \) such that \( h_{ij} \geq h_{mi}, 1 \leq m \leq L \)

\[
x^{(k+1)}_i = \frac{1}{h_{ij}} \left( y_i - \sum_{l=1}^{i-1} h_{il} x^{(k+1)}_l - \sum_{l=i+1}^{P} h_{il} x^{(k)}_l \right)
\]

(18)

Set \( x^{(k+1)}_i = 0 \) if the value obtained from Eq. (18) is negative.

Notice that \( h_{ij} \) is the largest element in column \( i \) of \( H \). This choice is to insure and speed the convergence of the process.

In coding the method into computer program, we made the following arrangements which are critical for the implementation of the algorithm on mini computers.

(a) \( y \), the \( L \)-dimensional vector of trunk group usage measurements, is stored in a matrix \( Y \). \( Y(i,j) \) denotes the usage measurement of the trunk group that interconnects Nodes \( i \) and \( j \).

(b) \( x \), the \( P \)-dimensional vector of the P-P offered load, is stored in a matrix \( X \). \( X(i,j) \) denotes the load between nodes \( i \) and \( j \). The initial estimate \( X^{(0)}(i,j) \) is assigned to be equal to \( Y(i,j) \).

(c) The matrix \( H \) is not stored. Its elements \( h_{ij} \) are calculated on line as needed in the iterative process. This reduces the memory requirements to the order of \( N^2 \) bytes.

5. PERFORMANCE ANALYSIS AND TEST

5.1 Performance Analysis

Three kinds of estimation error can result from this algorithm: the modeling error, the measurement error, and the error due to network disconnectivity.

(a) Modeling Error

The assumptions we made in the analysis, that ignore the effect of traffic peakedness on network call blocking, is expected to cause certain modeling error.

(b) Measurement Error

The accuracy of network measurements has been quantified before (see, e.g., \([6]\)). It was shown that, for large volume traffic, P-P load estimates based on direct trunk group measurements are of 5% root-mean-square error.
Due to Network Disconnectivity

This follows our discussion from Section 4.1. Recall that there are P unknowns and L equations in Eq. (13). For fully connected networks, L = P, and Eq. (13) has a unique solution. For non-fully connected networks, L < P, and Eq. (13) does not have a unique solution. This is where such estimation error comes in. This error will increase as the number L decreases. Our analyses show that the average percentage estimation error due to network disconnectivity is approximately 1 - UP, where UP is the network connectivity. UP = 1 for fully connected networks.

Notice that, for networks not fully connected, the "office totals" from each node can also be used as supplementary measurements. Let On denote the total (successful) traffic originating from Node n, and Tn the total (successful) traffic terminating at Node n. Then,

\[ O_n = \sum x_i(1 - B_i) \] (19)

where node n is the originating node of point-pair i.

\[ T_n = \sum x_i(1 - B_i) \] (20)

where node n is the terminating node of point-pair i. Eqs. (19) and (20) can be incorporated in Eq. (13) to better estimate x.

5.2 Test Results

We have tested this algorithm on 5-, 10-, and 52-node networks based on simulated network measurements. The connectivity of these networks ranges from 95 to 100%.

(a) P-P Blocking

Usually, it tends to overestimate the point-pairs that have very low blocking and underestimate those that have relative high blocking. As an example, when the realized P-P blocking in a simulated network is about 5%, an estimate from this algorithm can be 3% or less. The smaller the numbers is, the less accurate the estimate will be.

(b) P-P Offered Load

The load estimates are much more accurate than the blocking estimates. For networks with moderate (1%) or higher blocking, the P-P load estimates can be further improved by incorporating office total measurements. For very low blocking, fully-connected networks, such improvement is insignificant. Our results show that the overall average estimation error (EE), measured the following way, is around 7%.

\[ \text{EE} \triangleq \frac{\sum_{i=1}^{P} |x_i - \hat{x}_i|}{P} \] (21)

For the 52-node network, this algorithm takes 2 seconds CPU time on IBM RISC System/6000 model 320.

6. EXAMPLES

In this section, two examples are provided for illustration. The first example illustrates how we calculate P-P blocking and the second one illustrates how we construct the system equations to solve for the P-P loads.

Example 1. Consider a fully-connected 4-node network. The routing sequence for all the point-pairs is the following: 1) use the direct link as the first choice path. 2) use the other two nodes, with fixed order, as via nodes for the second and third choice path. What, then, is the P-P blocking for each point-pair?

Let us assume that all trunk groups have the same link blocking LB, and consider two cases: LB = .1 and LB = .2. The path blockings are:

<table>
<thead>
<tr>
<th>Path Type</th>
<th>LB = .1</th>
<th>LB = .2</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Crankback</td>
<td>.004</td>
<td>.026</td>
</tr>
<tr>
<td>Without Crankback</td>
<td>.011</td>
<td>.046</td>
</tr>
</tbody>
</table>

Example 2. Consider a 3-node network. The Nodes are denoted by A, B, and C. Link 1 interconnects Nodes A and B. Link 2 interconnects Nodes B and C. Link 3 interconnects Nodes A and C. The point-pairs and their routing sequences are given in the following table.
Based on the routing table, the mapping \( R(ijk) = 1 \) can be expressed as follows:

<table>
<thead>
<tr>
<th>ijk</th>
<th>1</th>
<th>ijk</th>
<th>1</th>
<th>ijk</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>1</td>
<td>311</td>
<td>3</td>
<td>511</td>
<td>2</td>
</tr>
<tr>
<td>112</td>
<td>0</td>
<td>312</td>
<td>0</td>
<td>512</td>
<td>0</td>
</tr>
<tr>
<td>121</td>
<td>3</td>
<td>321</td>
<td>2</td>
<td>521</td>
<td>3</td>
</tr>
<tr>
<td>122</td>
<td>2</td>
<td>322</td>
<td>1</td>
<td>522</td>
<td>1</td>
</tr>
<tr>
<td>211</td>
<td>2</td>
<td>411</td>
<td>1</td>
<td>611</td>
<td>3</td>
</tr>
<tr>
<td>212</td>
<td>0</td>
<td>412</td>
<td>0</td>
<td>612</td>
<td>0</td>
</tr>
<tr>
<td>221</td>
<td>1</td>
<td>421</td>
<td>2</td>
<td>621</td>
<td>1</td>
</tr>
<tr>
<td>222</td>
<td>3</td>
<td>422</td>
<td>3</td>
<td>622</td>
<td>2</td>
</tr>
</tbody>
</table>

The \( H \) matrix is

\[
H = \begin{bmatrix}
S_{11} & S_{22} & S_{32} & S_{41} & S_{52} & S_{62} \\
S_{12} & S_{21} & S_{32} & S_{42} & S_{51} & S_{62} \\
S_{12} & S_{22} & S_{31} & S_{42} & S_{52} & S_{61}
\end{bmatrix}
\]

Let us assume that all the links have the same blocking value of .1. Then, we obtain the values of \( s_{ij} \) from Eq. (9): \( s_{11} = .9 \) and \( s_{22} = .081 \). Thus we have,

\[
\begin{bmatrix}
y_1 \\
y_2 \\
y_3
\end{bmatrix} = \begin{bmatrix}
.9 & .081 & .081 & x_1 + x_4 \\
.081 & .9 & .081 & x_2 + x_5 \\
.081 & .081 & .9 & x_3 + x_6
\end{bmatrix}
\]

Assuming \( y_1 = 5 \text{ erl} \), \( y_2 = 7 \text{ erl} \), and \( y_3 = 10 \text{ erl} \), we obtain

\[
\begin{bmatrix}
x_1 + x_4 \\
x_2 + x_5 \\
x_3 + x_6
\end{bmatrix} = \begin{bmatrix}
4.06 \text{ erl} \\
6.05 \text{ erl} \\
10.16 \text{ erl}
\end{bmatrix}
\]

That is, the (two way) offered load between point-pair \((A,B)\) is 4.06 erl, between point-pair \((B,C)\) 6.05 erl, and between point-pair \((A,C)\) 10.16 erl.

7. CONCLUSION

In this paper, an algorithm for estimating the P-P traffic demand of a dynamic routing network is presented. Assuming no modeling error, the P-P traffic demand can be accurately estimated for fully connected networks. However, based on the available measurements, an optimal estimate, in the least-squares sense, can be obtained no matter how the network is connected. Studies show that estimates generated by this algorithm are of 7% estimation error in average. This estimation error can be further reduced by considering traffic peakedness factor in the load-blocking analysis.

ACKNOWLEDGEMENTS

The author would like to thank K. R. Krishnan and N. Koussoulas for their consultations and much effort in the simulation study.

REFERENCE


