Traffic Matrix Inference in IP Networks

N. Benameur, J. W. Roberts
{nabil.benameur, james.roberts}@francetelecom.com
France Telecom R&D
38-40 rue du Général Leclerc
92794 Issy Les Moulineaux, France
Tel : +33 1 45 29 51 93

Abstract

While the traffic matrix is used as basic data for many network planning tasks, it must be recognized that its inference in IP networks is particularly difficult and error prone. This paper discusses the issue of defining representative traffic demands and surveys proposed techniques for directly measuring elements of the traffic matrix or inferring their value from link measurements.
Key words: Traffic matrix, inference, traffic measurement, statistical estimation, origin-destination matrix.

1 Introduction

The traffic matrix quantifies the demand between all pairs of origin and destination nodes in a network. Consider the network represented by the 12-node, 19-edge graph in Figure 1 where each node is a potential traffic source or sink and each edge represents two one-way links. The traffic matrix in this example would specify demand on the 132 routes linking distinct nodes. The way demand is expressed depends on the type of network in question and its traffic characteristics. For the Internet, as discussed in the next section, demand can generally be summarized in terms of the overall bit rate generated by all user applications, averaged over a busy period typically of length one hour.

The traffic matrix is necessary for many network planning functions. It is clearly essential to know the volume of expected demand in order to size the network adequately to handle that demand with satisfactory quality (low delay and loss of transmitted data). Demand on any link of the network depends on the way traffic is routed. In current IP networks, the path of a given origin-destination route is the shortest path where the "length" of a link is an administratively assigned weight. Routing for a given origin-destination demand is thus fixed and it is consequently very important to carefully plan route assignments to ensure that the demand on any link is within capacity. To plan these assignments means appropriately choosing the administrative weights (Fortz & Thorup 2000). Since link sizes are upgraded in large discrete steps (e.g., link capacity is typically increased by a factor of 4 in moving from one level of the optical transmission hierarchy to the next), it is not possible to dissociate sizing and routing optimization.

Unfortunately, it proves particularly difficult to derive the traffic matrix for meshed IP networks like that of Figure 1. A significant source of difficulty is the lack of correspondence between IP addresses and geographical locations. To deduce the route to which a packet belongs it is necessary to apply routing protocols (so-called internal and external gateway protocols) to the configuration data in the routers at the time the packet was observed. The amount of traffic observed on links is not sufficient to deduce the amount of traffic on the routes. For instance, in Fig. 1, it would be necessary to deduce 132 route traffics from only 38 link measurements. Finally, it is well known that IP traffic is highly volatile both in time and in space. The traffic matrix can change significantly due both to rapid growth and to changing traffic source locality. Frequently changing inter-domain routing policies (BGP advertisements), often beyond the control of those planning the network in question, also cause the relation between route and link traffic to be particularly unstable.

There are some similarities between the problem of traffic matrix inference in IP networks and the derivation of origin-destination trip matrices for transportation planning purposes. In both cases link measurements are much more readily available than direct estimations of end to end flows. The common objective is to anticipate and avoid potential points of congestion although the potential for re-sizing and route optimization is clearly much greater in IP networking. Nevertheless, the techniques developed over the years in the transportation context constitute an interesting source of inspiration for the communication network planner.

Our objective in the present paper is to more clearly identify the issues faced in inferring the traffic matrix of an IP network. We first discuss the nature of IP traffic and how it can be adequately represented for network planning purposes. We then briefly survey the inference techniques proposed in the literature, distinguishing direct observation methods and approaches relying on inference from link measurements. Since all methods seem to have an inevitably high degree of imprecision, we briefly discuss in a concluding section how the network might be better designed to deal with demand uncertainty.

2 Characteristics of IP traffic

Traffic in an IP network results from a very wide variety of user applications producing data flows of many different forms. Traffic characteristics change continually as new applications gain popularity with two recent remarkable examples being the Web and peer to peer applications like Napster. However, it remains true that this traffic can be divided into two main categories: *streaming* flows, produced by audio and video applications, and *elastic* flows, produced by all applications involving the transfer of some form of digital document. From this perspective, evolutions in usage simply change the volumes and proportions of the two types of flow. Elastic traffic (which uses TCP) currently contributes around 90 to 95 percent of transported bytes.

The amount of data traffic observed at a given measurement point in successive intervals of time is notoriously variable making it difficult to characterize and measure (Leland *et al.* 1994). However, relative variability decreases as smaller streams are aggregated on network links (as predicted by the central limit theorem). Practical experience suggests that it is possible to define a mean "busy hour" traffic on high capacity links which is roughly stable on successive working days. This traffic is expressed as a bit rate equal to the volume of data observed divided by the length of the observation interval. Figure 2 shows the evolution over one week and one day of incoming and outgoing traffic on a high capacity IP network link. The curves plot traffic averaged over successive 5 minute intervals.

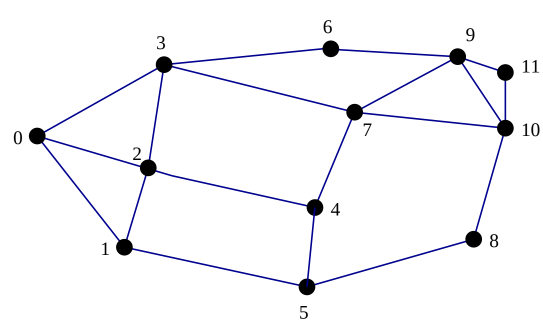

Figure 1: Network topology

The link capacity is 10 Gbit/s and is quite uncongested, like most links in present IP backbones. The measured traffic is thus a true representation of expressed demand.

It is clear that the busy period is well defined and that traffic in this period is roughly the same on successive working days. It is meaningful in this case to speak of a representative traffic matrix. If the network is sized and routing is optimized to handle this traffic without congestion, it may be assumed that the network will provide adequate quality at all times. In some regions like North America, the fact that a network spans several time zones may be exploited by planning the network to accommodate a number of representative traffic matrices. This is current practice for the telephone network but not for the Internet.

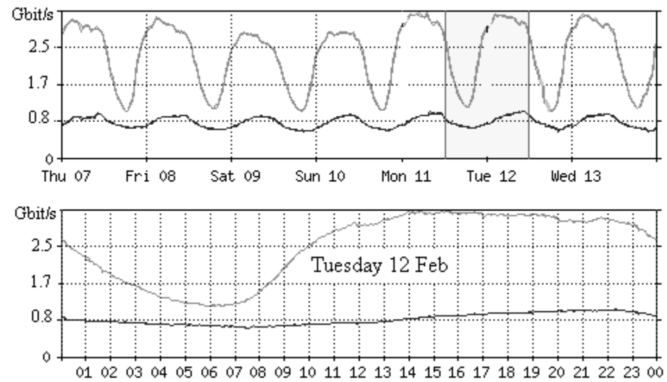

Figure 2: Weekly and daily demand profiles on an OC192 link, February 2002 (in & out traffic)

Recent results on modelling the performance of elastic traffic demonstrate that knowing just the average traffic offered is sufficient to determine the most significant quality of service measures such as the expected response time of a document transfer (Ben Fredj *et al.* 2001). Should streaming traffic grow significantly in relative volume it may be necessary to characterize demand with additional parameters describing the traffic mix (the different rates of audio and video flows, for instance). However, for present purposes, we assume that the only data required for each route is a representative expected offered traffic in bits/s.

There are currently no well-established rules for defining the representative traffic to be used for IP network planning. One current practice, mainly for billing purposes, is to measure traffic in successive 5 minute intervals and to select as representative value the 95-th percentile of these values. ITU recommended practice for traffic measurements in the telephone network precisely defines daily, weekly and monthly representative values to be used for planning purposes (ITU Rec. E.500). This practice is not obviously adaptable to the context of the Internet, however, due notably to the current extremely high growth rate.

The traffic matrix can be defined using these representative values. However, frequent changes in routing protocols due to changes in peering agreements or shifts in the location of preferred traffic sources suggest that the traffic matrix can vary frequently. Figure 3 shows an example of such variations on a backbone link in a period of one month where traffic in one direction (the darker line) changes significantly between weeks 47 and 49. This volatility suggests the need for simple traffic matrix inference methods and for robust engineering approaches.

3 Direct measurement of the traffic matrix

A direct method for evaluating the traffic matrix consists in observing packets at their origin on an ingress router and deducing their egress from the destination IP address. The latter deduction

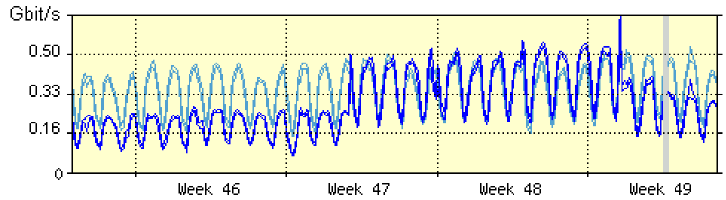

Figure 3: Demand variation over one month on an OC48 link, December 2001.

is not straightforward, however, since there is no strict correspondence between IP addresses and geographical locations. It is necessary to deduce the location from knowledge of the internal and external gateway routing protocols of the considered domain. The routing information necessary to fix the destination is much more complete and complex than that necessary for packet forwarding which simply specifies the next hop. A two-step process is necessary:

- traffic data is collected during the network busy period;
- an off-line destination analysis is applied using the routing tables effective when the measurements were made.

This direct approach is advocated by Feldmann *et al.* (2000) who collect incoming and outgoing traffic statistics on just the peering links of the considered AT&T IP backbone. From this data they can deduce the traffic matrix for routes having origin or destination outside their backbone. They rely on the Netflow tool to gather packets into flows having the same origin and destination (IP addresses and transport level port numbers) (Netflow 2002). The collected flow data is exported at 15 minute intervals to a central storage point to be analysed off-line using additional data collected from router configuration files and forwarding tables. This approach has significant disadvantages. The amount of data to be transported to the central storage device is very high and the activation of Netflow (or similar software) in routers is known to consume a non-negligible amount of CPU time and can compromise router performance.

An alternative method of gathering traffic data is used by Sridharan *et al.* (2001) in a study of routing strategies for the Sprint backbone. This is not advocated as an operational procedure but is still worthy of mention. The authors have performed extremely detailed analyses of traffic entering and leaving the network via one PoP (point of presence). Specialized equipment records the salient details of all packets observed on all ingress and egress interfaces. Post-processing allows the reconstitution of flow data at any required granularity. By associating known BGP routing tables it is possible to deduce the proportions of traffic between that PoP and all the other PoPs of the Sprint network. This allows them to constitute one row and one column of a PoP-to-PoP traffic matrix. To complete the matrix, they then apply an *ad hoc* extrapolation method. To derive an accurate traffic matrix it would be necessary to perform similar measurements at every PoP. While this method thus appears clearly too complex for routine operational purposes, it is certainly useful in identifying some interesting particularities of IP traffic such as the very high variability in the size of point to point flows, also noted by Feldmann *et al.* (2000).

The need to know the routing tables which are valid at the time the traffic measurements are made can be a significant constraint. This requirement is avoided with a technique proposed by Duffield and Grossglauser (2000). They point out that the trajectories of flows through a network can be observed directly by means of a particular sampling technique. A small proportion of packets is sampled on all network links by means of a hash function of the invariant packet content yielding 1 (packet sampled) or 0 (packet ignored). Data for all sampled packets are collected and analysed off line to reconstitute a "photograph" of the flow of traffic through the network. If a packet is sampled at one router, it is necessarily also sampled at all other routers and it is possible to identify its path. The number of sampled packets must be chosen as a trade-off between accuracy and generated overhead. A significant drawback with this approach is the requirement to instrument all observed router interfaces.

4 Inference from link traffic measurements

In contrast to the difficulty involved in obtaining direct measurements on end-to-end flows, it is routine in IP networks to derive traffic counts on all router interfaces via SNMP (the *simple network management protocol*). These data can be used to infer elements of the traffic matrix. Recent studies in the present area of IP networks have been inspired by earlier work on traffic inference from link counts in road transportation.

In road transportation, it is difficult to obtain Origin-Destination matrices by measurements, interviews or surveys. Various approaches to estimating the traffic matrix using traffic counts on links have been developed. The objective is to obtain the most likely matrix causing the observed link counts. Some models solve this problem by using deterministic techniques to find the most likely matrix given a general model of trip distribution or a prior estimated traffic matrix while other models adopt statistical inference techniques. The purpose of this section is to provide a brief survey of some of these approaches. A recent publication provides a comparative evaluation (Medina *et al.* 2002).

4.1 Deterministic inference

We assume that traffic on network links is known precisely and that we also know the paths followed by all end-to-end routes. Let the number of links be L and the number of routes R. Let y_l denote the traffic on link l for $1 \leq l \leq L$, and x_r the traffic on end-to-end route r for $1 \leq r \leq R$. Note that for convenience we represent the traffic matrix here as a vector $X = \{x_r\}$. The volumes are defined by the vector $Y = \{y_l\}$.

Network paths are defined by a matrix $A = [a_{rl}]$ where a_{rl} represents the proportion of route r traffic that uses link l. In networks with fixed routing we would have $a_{rl} \in \{0, 1\}$. Current IP networks allow either fixed routing or load sharing over a set of equal cost paths in which case the a_{rl} would be equal for the links composing those paths.

Each route r corresponds to an end-to-end demand of volume x_r such that:

$$\sum_{r=1}^{R} a_{rl} x_r = y_l, \text{ for } 1 \leq l \leq L. \tag{1}$$

Or equivalently:

$$Y = AX. \tag{2}$$

Since $L \ll R$, the above equations for the unknowns x_r are largely under-specified (i.e., a large range of traffic matrices produces the same set of link traffics). To determine the "best" solution, it is usual to make use of additional information in the form of an initial traffic matrix $\{\widetilde{x}_r\}$. This may, for example, be derived from historical data or from additional knowledge about the user population. The objective is to find the solution to (1) which minimizes some metric measuring the distance between $\{x_r\}$ and $\{\widetilde{x}_r\}$ under the constraints expressed in equation (1).

The above problem has been considered in some detail in the context of road traffic, (e.g., Bell 1983, Bierlaire & Toint 1994, Van Zuylan & Willumsen 1980[1]). Methods differ depending on the metric used. A popular approach is to preserve the "information content" of the matrix $\{\widetilde{x}_r\}$ by maximizing entropy (Van Zuylan & Wiliamson 1980). The objective is to find the traffic matrix $\{x_r\}$

[1] We have not attempted to provide a complete survey but simply cite some significant contributions which usefully illustrate the different techniques.

that adds as little information as possible to the information in the initial traffic matrix $\{\widetilde{x}_r\}$. This leads to a distance metric of the form $\sum x_r(\log x_r/\widetilde{x}_r - 1)$. An alternative distance metric is the sum of squares $\sum(x_r - \widetilde{x}_r)^2$ yielding a traffic matrix $\{x_r\}$ that is as close as possible to the initial matrix $\{\widetilde{x}_r\}$.

A deterministic inference technique well-known in telephone network planning is the method of Kruithof (Krupp 1979). It is assumed here that we dispose of measurements of the aggregate traffic entering and leaving all terminal nodes together with an initial estimate $\{\widetilde{x}_r\}$. The elements of $\{\widetilde{x}_r\}$ are successively adjusted first to match the outgoing traffic measurements (row sums) and then the incoming traffic measurements (column sums). The process is repeated until convergence is obtained. This method is not well suited to the topology of an IP network, however, since routes are typically much longer than in the telephone network, and fails to make use of readily available link traffic data.

4.1.1 Evaluation

To illustrate the usefulness of deterministic inference we consider a simple example using the sum of squares distance metric. The problem to solve is the following:

$$\text{Minimize } \sum_{r=1}^{R}(x_r - \widetilde{x}_r)^2 \tag{3}$$

subject to :

$$\sum_{r=1}^{R} a_{rl}x_r = y_l, \text{ for } 1 \le l \le L \text{ and } x_r \ge 0. \tag{4}$$

We first define a traffic matrix for the network illustrated in Figure 1 by randomly selecting traffic volumes distributed between 0 and 200 Mbit/s for all 132 node pairs. This traffic is routed over fixed routes and the link loads are calculated. To represent the initial estimate of the traffic matrix $\{\widetilde{x}_r\}$, we perform a random perturbation of the true matrix. The perturbation consists in multiplying each traffic volume x_r by a random factor uniformly distributed between $-a$ and a. The parameter a gives the level of perturbation. Non-uniformly distributed factors have also been tested.

We then calculate the matrix $\{x_r\}$ satisfying equations (4) and minimizing the sum of squares (3).

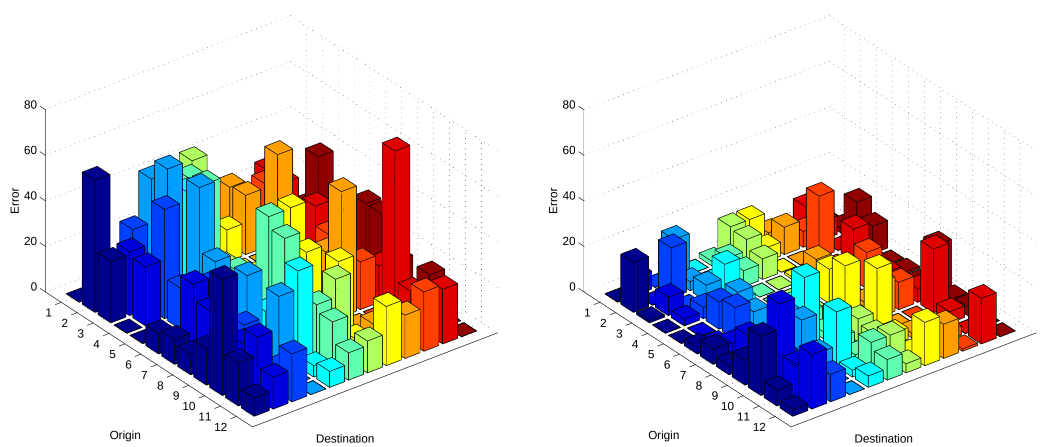

Figure 4: Differences between the true and the estimated matrix before and after correction.

Results for a particular experiment are illustrated in Figure 4. The graph on the left shows the initial perturbation with error terms of up to 60 Mbit/s. The result of the optimization on the right shows that the errors are significantly reduced, all being less than 30 Mbit/s in this case. These results are typical of many similar experiments using different initial data and different perturbations, as shown in Table 1.

	Initial matrix		Corrected matrix	
	Average error	Max error (Mb/s)	Average error	Max error (Mb/s)
Uniform perturbation	31,49	151	20,73	92
	11,72	57	7,02	32
	3,86	19	2,39	10
Non-Uniform perturbation	44,64	157	27,09	143
	16,90	59	6,79	33
	5,84	20	2,11	9

Table 1: Differences between the true and the estimated matrix before and after correction.

4.2 Statistical inference

Some recent work on traffic matrix inference for IP networks makes use of statistical techniques, again pioneered in the field of road transportation planning, (e.g., Spiess 1987, Tebaldi & West 1998, Willumsen 1984). The data available are sets of traffic measurements for each link obtained in successive time periods. Such measurements might, for example, be derived by polling the router management information data base at 5 minute intervals using SNMP.

Given an initial assumption about the form of the distribution of the size of the end-to-end traffic flows, the objective is to determine the most appropriate values of its parameters, given the set of traffic measurement data.

The most popular approach is the maximum likelihood method (Spiess 1987). The prior traffic matrix and the traffic counts are regarded as observations resulting from the true traffic matrix. The method consists in maximizing the likelihood of observing $\widetilde{X} = \{\widetilde{x}_r\}$ and $Y = \{y_r\}$ conditional on $X = \{x_r\}$. Assuming that $\{\widetilde{x}_r\}$ and $\{y_r\}$ are independent, this can be expressed as :

$$L(\widetilde{X}, Y|X) = L(\widetilde{X}|X).L(Y|X) \tag{5}$$

In (Spiess 1987), the likelihood is calculated under the assumptions that the matrix $\{\widetilde{x}_r\}$ follows either a multinomial distribution when the sampling size is assumed small or a Poisson distribution when the sampling size is large. A Poisson distribution is also assumed for the link counts. The traffic matrix estimation consists in maximizing the obtained likelihood over all possible traffic volumes $\{x_r\}$ given by the constraints in (1).

Another approach is the Generalised Least Squares method where $\widetilde{X}$ is supposed to be obtained from X with a probabilistic error term ϵ with zero mean and finite covariance (Cascetta 1984, Bell 1991). This leads to the minimization of a distance between $\widetilde{X}$ and X involving the covariance matrix.

In an alternative method, Vardi supposes the end-to-end traffic flow volumes $X = \{x_r\}$ have a Poisson distribution ($X \sim Poisson(\lambda)$) and calculates the respective mean rates by means of a maximum likelihood estimation (Vardi 1996). The goal here is to estimate the vector $\lambda = \{\lambda_r\}$ based on the observed link counts. Assuming K repeated observations of the link volumes denoted by $Y^k = \{y_r^k\}$, the maximum likelihood estimation yields the following equation:

$$\lambda = \frac{1}{K} \sum_{k=1}^{K} E_\lambda[X^k|Y^k = AX^k]. \tag{6}$$

Expectation-Maximisation can be used to estimate the parameter λ by applying the following iteration:

$$\lambda^{n+1} = \frac{1}{K}\sum_{k=1}^{K} E[X^k|Y^k,\lambda^n]. \tag{7}$$

However, this requires finding all solutions of $AX = Y$. In order to solve this problem, one possibility consists in using an approximation for Y based on the central limit theorem:

$$\bar{Y} = \frac{1}{K}\sum_{k=1}^{K} Y^k \rightarrow N(A\lambda, \frac{1}{K}A\Lambda A') \text{ with } \Lambda \equiv diag(\lambda). \tag{8}$$

The method consists then in maximizing the likelihood expressed as follows:

$$l(\lambda) = -log|A\Lambda A'| - K(\bar{Y} - A\lambda)'(A\Lambda A')^{-1}(\bar{Y} - A\lambda) \tag{9}$$

This approach has been refined by Cao *et al.* (1999, 2000). In this work the authors suppose a normal distribution for the x_r, accounting for the observed fact that the distribution of traffic volumes is generally more variable than Poisson. The authors also propose methods to account for intensities which vary over the sampling period. The methods proposed in (Cao *et al.* 1999) are somewhat complex and suitable for small networks only. A decomposition approach is proposed in (Cao *et al.* 2000) allowing an extension to large networks.

Still more general assumptions about the form of the traffic distribution are possible using the Bayesian approach proposed by Tebaldi and West (1998). These authors notably identify a problem of bias with classical maximum likelihood methods which tend to significantly overestimate the traffic on low intensity routes. The authors show that a choice of informed priors $p(\lambda)$ (deduced from an initial estimate of the traffic matrix, as in Section 4.1) considerably improves the accuracy of estimations when the end-to-end traffic flow volumes $X = \{x_r\}$ follow a Poisson distribution ($X \sim Poisson(\lambda)$):

$$p(X,\lambda) = p(\lambda)\prod_{i=1}^{R} \lambda_i^{x_i} exp(-\lambda_i)/x_i! \tag{10}$$

The goal is to compute the likelihood of observing $X = \{x_r\}$ conditional on $Y = \{y_r\}$. This distribution is related to the two posterior distributions $p(\lambda|X,Y)$ and $p(X|\lambda,Y)$ through:

$$p(X|Y) = p(X|\lambda,Y)p(\lambda)/p(\lambda|X,Y) \tag{11}$$

The computation of these two posterior distributions involves iterative simulation methods such as Markov chain Monte Carlo (MCMC) sampling for generating λ and X.

The evaluations performed by Medina *et al.* (2002) show that the expectation maximization approach of Cao *et al.* (2000) outperforms the Bayesian methods of Tebaldi and West (1998). However all inference methods appear to suffer from significant inaccuracy. An important determining factor is the precision of the initial estimation of the traffic matrix $\{\tilde{x}_r\}$.

4.3 The initial estimates

The above inference methods clearly rely on the accuracy of the initial estimate of the traffic matrix $\{\tilde{x}_r\}$. This estimate can be derived from a variety of sources including historical data on traffic distributions or from sampling measurements performed using techniques like those in (Sridharan *et al.* 2001) or (Duffield & Grossglauser 2000). A promising approach is to use a so-called gravity model for the traffic distribution in the network (Kowalski & Warfield 1995, West 1994) . The classical gravity model relates the traffic on route r linking origin i and destination j to the user populations M_i and M_j and the distance d_{ij} between i and j, as follows:

$$\widetilde{x}_r = \alpha \frac{M_i M_j}{d_{ij}^{\beta}} \tag{12}$$

with appropriate choice of model parameters α and β.

This particular formulation can be applied to the estimation of telephony traffic (Kowalski & Warfield 1995) or road traffic (West 1994) but is less valid in the case of IP networks where distance has a smaller impact on traffic.

However, the gravity model can be adapted using alternative factors such as total incoming and outgoing traffic (as in Kruithof's method) or factors reflecting structured features such as the location of peering points or customers for an access backbone. The so-called choice models proposed by Medina *et al.* (2002) produce encouraging results.

5 Conclusions: dealing with uncertainty

The problem of traffic matrix inference in IP networks is difficult and yet extremely important for network planning. The direct observation method presented by Feldmann *et al.* (2000) appears as the most accurate solution but implementation is complex and currently relies on proprietary software. Inference from more readily available link measurements is simpler and can readily be used to improve imprecise preliminary estimates. The accuracy of these inference methods depends significantly on the precision of the initial estimated traffic matrix.

However, it does appear that the derived traffic matrix can never constitute a very accurate long term measure of demand. This is because of the inherent volatility of IP traffic, due notably to the instability of routing protocols and to frequent changes in traffic patterns as the location of preferred data sources shifts. This volatility is largely beyond the control of the network planner and must be taken into account in developing more robust engineering methods.

The network architecture might also be revised to facilitate traffic fluidity. Current routing protocols are designed more to ensure logical connectivity than to guarantee the efficiency and quality of user transactions. Use of MPLS would allow greater control over traffic routing and also provide scope for direct route traffic measurement when the route coincides with a label switched path (Rosen *et al.* 2001). At longer term, one might envisage the use of traffic aware adaptive routing, as in the telephone network, to ensure that flows are routed over paths with sufficient capacity to handle them. If adaptive routing were employed, precise knowledge of the traffic matrix is less critical since traffic flows automatically find the paths with sufficient capacity.

The art of estimating IP network traffic matrices is still in its infancy. Lessons learned in the field of transportation, as well as in that of traditional telecommunications networks, are extremely useful but do not provide a complete solution. The specific constraints arising from the nature of IP traffic and the way it finds its way through the network via the routing protocols makes this a particularly important and challenging research issue.

References

Bell, M. G. H., 1983 *"The Estimation of an Origin Destination Matrix from Traffic Counts"* Transportation Science 17, pp 198-217.

Bell, M. G. H., 1991 *"The Estimation of Origin Destination Matrices by Constrained Generalised Least Squares"* Transportation Research 25B, pp 13-22.

Ben Fredj, S., Bonald, T., Proutière, A., Regnié, G., Roberts, J., 2001 *"Statistical bandwidth sharing: a study of congestion at flow level"*, Proceedings of ACM Sigcomm 2001, Computer Communication Review, Vol 31, No 4.

Bierlaire, M., Toint, Ph. L., 1994 *"MEUSE : An Origin Destination Matrix Estimator That Exploits Structure"* Transportation Science 31(4), pp 363-371.

Cao, J., Davis, D., Vander Wiel, S., Yu, B., 2000 *"Time-Varying Network Tomography : Router Link Data"* Journal of the American Statistical Association, Vol 95, No 452, pp 1063-1075.

Cao, J., Vander Wiel, S., Yu, B., Zhu, Z., 2000 *"A scalable Method for Estimating network traffic matrices"* Bell Labs technical report, , available at http://cm.bell-labs.com/~cao.

Cascetta, E., 1984 *"Estimation of Trip Matrices from Traffic Counts and Survey Data, a Generalized Least Squares Estimator"* Transportation Research 18B, pp 288-299.

Duffield, N. G., Grossglauser, M., 2000 *"Trajectory Sampling for Direct Traffic Observation"* Proceedings of ACM Sigcomm 2000, Computer Communication Review, Vol 30, No 4.

Feldmann, A., Greenberg, A., Lund, C., Reingold, N., Rexford, J., True, F., 2000 *"Deriving Traffic Demand for Operational IP Networks : Methodology and Experience"* Proceedings of ACM Sigcomm 2000, Computer Communication Review, Vol 30, No 4.

Fortz, B., Thorup, M., 2000 *"Internet Traffic Engineering by Optimizing OSPF Weights"* Proceedings of Infocom 2000.

ITU Recommendation E.500, , 1998 *"Traffic intensity measurement principles"*, Geneva.

Kowalski, J.P., Warfield, B., 1995 *"Modelling Traffic Demand Between Nodes in a Telecommunication Network"*, ATNAC'95.

Krupp, R.S., 1979 *"Properties of Kruithof's projection method"*, The Bell System Technical Journal, 58, pp517-538.

Leland, W., Taqqu, M., Willinger, W., Wilson, D., 1994 *"On the self-similar nature of Ethernet traffic (extended version)"*, IEEE/ACM Trans. Networking, pp1-15.

Medina, A., Taft, N., Salamatian, K., Bhattacharyya, S., Diot, C., 2002 *"Traffic matrix estimation: Existing techniques and new directions"*, Proceedings of ACM SIGCOMM 2002.

Netflow 2002 *"Netflow Aggregation"* Cisco IOS release 12.0(5)T.

Rosen, E., Viswanathan, A., Callon, R., 2001 *"Multiprotocol Label Switching Architecture"*, IETF RFC 3031.

Spiess, H., 1987 *"A maximum-likelihood model for estimating origin-destination matrices"* Transportation Research 28B, pp 395-412.

Sridharan, A., Bhattacharyya, S., Diot, C., Guerin, R., Jetcheva, J., Taft, N., 2001 emph"On the Impact of Aggregation on the Performance of Traffic Aware Routing". Proceedings of ITC 17, Elsevier, pp 111-123.

Tebaldi, C., West, M., 1998 *"Bayesian Inference on Network Traffic Using Link Count Data "* Journal of the American Statistical Association.

Van Zuylen, H.G., Willumsen, L.G., 1980 emph"The most likely trip matrix estimated from traffic counts" Transportation Research 143, pp 281-293.

Vardi, Y.,1996 emph"Network Tomography : estimating source-destination traffic intensities from link data" Journal of the American Statistical Association, Vol 91, n433, Theory and Methods.

West, M.,1994 *"Statistical Inference for Gravity Models in Transportation Flow Forecasting"* Discussion Paper 94-40, IS- DS, Duke University.

Willumsen L.G., 1984 *"Estimating time dependent trip matrices from traffic counts"* Proceeding of the ninth international symposium on traffic theory.