

Performance Evaluation of ATM Shortcut Connections in Overlaid IP/ATM Networks *

Victor Firoiu [†]
vfiroiu@nortelnetworks.com

Bay Architecture Lab
Nortel Networks
3 Federal St., Billerica, MA 01821 USA

Jim Kurose
kurose@cs.umass.edu

Don Towsley
towsley@cs.umass.edu

Department of Computer Science
University of Massachusetts
Amherst, MA 01003 USA

Abstract - In this paper we present methods to evaluate the benefit of using direct ATM connections (shortcuts) between IP nodes in IP over ATM networks, and we identify the combinations of IP and ATM network topologies where ATM shortcut benefits are likely to be high. We model an IP/ATM network with and without ATM shortcuts as two loss networks. We propose a metric for network performance comparison, the Network Load Ratio, that gives the ratio of the number of flows accepted by two networks at the same network blocking probability. We derive an estimator of this metric, the Asymptotic Load Ratio, that has low computational complexity. This estimator forms the basis of a methodology for network performance comparison. We use this method in simulation experiments using random networks. These experiments indicate that in many cases the utilization of an IP/ATM network increases proportionally to the decrease in the average path length when ATM shortcuts are used. We have also found that there is almost no correlation between the increase in network utilization (when using ATM shortcuts) and the IP to ATM node ratio.

1 Introduction

The exponential growth of interconnected data networks such as the Internet is probably the most important development in the telecommunication sector in the current decade. The improvement in the capacity of telecommunication networks through fiber optic and ATM technology has been an essential contribution to Internet's transition to

*This material is based upon work supported in part by the National Science Foundation under Grant NCR-95-08274 and CDA-95-02639. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[†]This work was done in part while Victor Firoiu was with University of Massachusetts, Amherst.

a mass scale world network. Today, at most of its structural levels (Backbone, WAN, regional Internet Service Providers), the Internet is carried over telecommunication networks that are increasingly deploying ATM technology (e.g., NSF's very High Speed Backbone Network Service, vBNS, [8]). In such cases, IP networks are configured as virtual networks over ATM network infrastructures. The results, overlaid IP/ATM networks, feature a virtual separation between the IP and ATM networks. This is exemplified in Figure 1, where IP traffic between *A* and *C* is forwarded over the IP default route constructed from the fixed ATM VCs *A - B* and *B - C*.

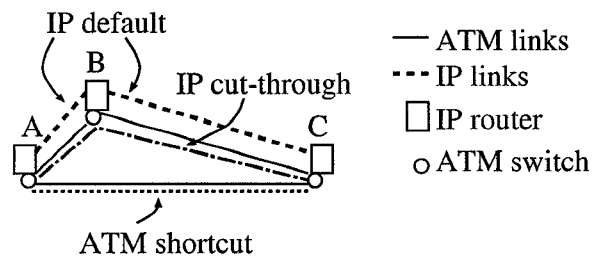


Figure 1: The three IP/ATM types of service for IP flows

The recent development and deployment of network layer ATM protocols such as ATM signaling [18] and PNNI [16] makes it possible for IP to interact with ATM over the same physical network. Recently, several proposals have been put forward to improve the efficiency of IP/ATM networks through an interaction between IP and ATM network layers. One proposal [15] is to *cut through* the IP processing of IP packets and transmit the data of an IP flow through a separate ATM virtual circuit (VC). The VC connects two IP routers at the edge of the ATM network, and the route of the ATM VC follows the IP route (in

Figure 1 the IP cut through is realized through the switched VC $A - B - C$).

Another set of proposals goes one step further and permits the ATM VC, called *ATM shortcut*, connecting two IP routers, to be routed by the ATM network. ATM shortcuts are proposed by the Next Hop Resolution Protocol (NHRP) [14] for unicast IP flows and by Multicast Address Resolution Protocol (MARS) [1] for multicast IP flows. Alternatively, in [2] we proposed a combination of RSVP and ATM signaling for resource reservation of IP flows with Quality of Service (QoS) requirements in IP/ATM networks. For example, in Figure 1 the ATM shortcut between A and C is the switched VC $A - C$.

IP cut-through has proven to be beneficial in experimental systems and commercial products by increasing the average throughput of IP routers embedded in ATM networks [12, 4]. Additional performance improvements are possible in the ATM shortcut operation mode since routing IP flows in the ATM network provides the opportunity for more efficient network resource utilization than the IP cut through mode. However, there has not yet been any theoretical or practical confirmation of the benefit of ATM shortcuts. In this paper we address the following question:

Given an overlaid IP/ATM network, what is the performance benefit of ATM shortcuts?

This question is important in IP/ATM network design, where the benefits of ATM shortcuts can be weighed against the simplicity of a separate model where IP and ATM are maintained at separate layers. The purpose of this paper is to develop a methodology for evaluating ATM shortcuts, given an IP/ATM network, and to assess the benefits of ATM shortcuts in a variety of scenarios.

To answer the above question, we first model an IP/ATM network, with and without ATM shortcuts, as two distinct network graphs. We then propose a metric to compare the performance of the two networks: the Network Load Ratio. For a given IP/ATM configuration, the Network Load Ratio represents the additional load the IP/ATM network can accept on average when using ATM shortcuts (compared with the “IP cut through” case), for the same network blocking probability. We develop a method to compute the Network Load Ratio for a given network blocking probability based on the Fixed Point Method [17]. We also develop the notion of Asymptotic Load Ratio which is an approximation to the Network Load Ratio when the networks operate in underload conditions. We find experimentally through simulation that this provides a good approximation (the 90-percentile of the relative error is less than 0.12) to the Network Load Ratio for a range of network blocking probabilities (less than 0.02), thus giving a general indication of the relative performance of the two networks. This second method is very attractive due to its

low computational complexity. We use this method in simulation experiments to identify network topological conditions where ATM shortcuts bring high benefits. We find that in many cases the utilization of an IP/ATM network is improved significantly when the average route length decreases upon using ATM shortcuts. We also find that there is almost no correlation between the increase in network utilization (when using ATM shortcuts) and the IP to ATM node ratio.

The paper is organized as follows. In Section 2 we propose a metric for network performance comparison, the Network Load Ratio and its limiting value for underloaded conditions, the Asymptotic Load Ratio, and two methods for their approximate computation. In Section 3 we evaluate the accuracy with which the Asymptotic Load Ratio estimates the Network Load Ratio. In Section 4 we use the Asymptotic Load Ratio in simulation experiments to determine network topological conditions where ATM shortcuts bring high benefits. Section 5 concludes the paper.

2 A Metric and Two Methods for Network Performance Comparison

Our approach in quantitatively evaluating the benefits of ATM shortcuts consists of comparing the flow loads offered to two networks having the same blocking probability. We will compare a given IP/ATM network where IP flows are routed by IP routing, with the same IP/ATM network where IP flows are routed by ATM routing. In this section we propose a metric to compare the performance of two loss networks, and we develop two methods for estimating this metric. The technical details rely on the theory of loss networks; see [20, 9, 10, 17] and references therein.

We assume that both IP and ATM routing are fixed. Alternate routing, possible in ATM and also under study for IP (see for example [3]), may bring additional benefits in increased network utilization. Nevertheless, computing network loss probability is much more complex when using alternate routing (see for example [6]). Moreover, alternate routing in ATM, as well as QoS-based routing in IP, is work in progress, and therefore, not completely specified. Therefore, here we consider fixed routing as a first step in evaluating the performance of ATM shortcuts. An extension to our study that would include alternate routing constitutes an interesting future work.

We consider a network consisting of a set of nodes \mathcal{V} and a set of links \mathcal{L} . A link $l \in \mathcal{L}$ has capacity C_l . The flows that arrive to the network have origin and destination in a set $\mathcal{A} \subseteq \mathcal{V}$, called *access nodes*. (In the next sections, \mathcal{V} will be the set of ATM nodes, and \mathcal{A} , the set of IP nodes.) These flows are classified into a set of classes \mathcal{K} according to their end-point nodes, i.e., there is a separate class $k \in \mathcal{K}$ for each pair of nodes (A, B) , $A, B \in \mathcal{A}$.

Since we assume that routing is fixed, it follows that all

flows of class $k \in \mathcal{K}$ are routed on route r_k , where r_k is the sequence of links traversed by flows of class k . In this study, we assume that all flows have the same bandwidth requirements (henceforth taken as a unit of bandwidth), irrespective of class. The more general case, where flows can have different bandwidth requirements, although tractable, is more complex, and we leave it for future work. A class k flow is admitted iff one unit of capacity can be reserved at each link $l \in r_k$. We define $\mathcal{R} = \{r_k | k \in \mathcal{K}\}$ to be the set of routes in the given network.

Flows arrive to the network according to a Poisson process with rate λ and have exponential holding times with mean $1/\mu$. The intensity of offered load is $\rho = \lambda/\mu$. With probability ν_k a flow becomes a class k flow. A flow is blocked at a link if the flow's bandwidth requirement is larger than the link's available bandwidth. The probability that a flow of a class is blocked at link j is denoted by L_j . A flow is blocked (its establishment is denied) if it is blocked at any link on its path (route). The probability that a class k flow is blocked (not admitted) is denoted by B_k . The probability that a flow is blocked, called the network blocking probability, is denoted by P and is given by

$$P = \sum_{k \in \mathcal{K}} \nu_k B_k = P(\rho) \quad (1)$$

where P depends on ρ through B_k . Let us define ϕ_k to be the intensity of offered load of class k

$$\phi_k = \nu_k \rho \quad \forall k \in \mathcal{K} \quad (2)$$

We also introduce η_i , the probability that a flow traverses link i ,

$$\eta_i = \sum_{k \in \mathcal{K}, i \in r_k} \nu_k \quad \forall i \in \mathcal{L} \quad (3)$$

and ψ_i , the intensity of offered load on link i ,

$$\psi_i = \sum_{k \in \mathcal{K}, i \in r_k} \phi_k = \eta_i \rho \quad \forall i \in \mathcal{L} \quad (4)$$

Let us now consider two networks N_1 and N_2 with link sets \mathcal{L}_1 and \mathcal{L}_2 and route sets \mathcal{R}_1 and \mathcal{R}_2 , that have the same set of access nodes, $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{A}$ (i.e., the same set of flow end points), the same set of flow classes, $\mathcal{K}_1 = \mathcal{K}_2 = \mathcal{K}$ and the same offered load probabilities $(\nu_k)_{k \in \mathcal{K}}$. We define the following metric for comparing the performance of the two networks.

Definition 1 The Network Load Ratio of networks N_1 and N_2 at blocking probability p , $R(N_1, N_2, p)$, is defined to be the value of R such that

$$P_1(R\rho) = P_2(\rho) = p \quad (5)$$

Network N_1 is said to perform R times better than N_2 at blocking probability p .

In other words, if N_1 and N_2 have a Network Load Ratio of R , N_1 can handle a load that is R times greater than N_2 for a network blocking probability p . Observe that the Network Load Ratio R depends on the network blocking probability p and on the characteristics of the two networks: $R = R(N_1, N_2, p)$. Henceforth, when no confusion is possible, we drop the parameters N_1 and N_2 from the expression of R .

The Network Load Ratio for a given network blocking probability p is given by

$$R(N_1, N_2, p) = \frac{P_1^{-1}(p)}{P_2^{-1}(p)} \quad (6)$$

where $P_i^{-1}(\cdot)$ is the inverse of function $P_i(\cdot)$, $i = 1, 2$. In general, the network blocking probability cannot be computed exactly due to its computational complexity [13]. However, a good approximation is given by the Fixed Point method, also known as the Reduced Load approximation [10]. Following the Fixed Point method, an approximation of the blocking probability B_k of flows in class k , B_k^* , is given by

$$B_k^* = 1 - \prod_{j \in r_k} (1 - L_j^*) \quad (7)$$

where L_j^* , an approximation of the blocking probability L_j at link j , is the solution of the following system of equations,

$$L_j^* = Er\left(\sum_{k \in \mathcal{K}, j \in r_k} \phi_k \prod_{i \in r_k - \{j\}} (1 - L_i^*), C_j\right) \quad j \in \mathcal{L} \quad (8)$$

Here $Er(\cdot)$ is the Erlang loss formula,

$$Er(\rho, C) = \frac{\rho^C / C!}{\sum_{n=0}^C \rho^n / n!} \quad (9)$$

It follows that an approximation for the network blocking probability is,

$$P^* = \sum_{k \in \mathcal{K}} \nu_k B_k^* \quad (10)$$

The system of equations (8) can be solved by repeated substitutions, and is known to converge in most practical cases, giving a good approximation to B_k . The inverse of the network blocking probability function, $\rho = P^{-1}(p)$, can be computed using any numerical method for approximate root computation (e.g., the secant method) in conjunction with the Fixed Point method. The procedure based on equations (2), (6)-(10) allows us to compute the Network Load Ratio of two networks that operate at a given network blocking probability. It constitutes the first method that we propose for network performance comparison.

A problem with this method is its rather high computational complexity. The computation of Network Load Ratio based on equations (6-10) has complexity $O(L^2CKFS)$, where L is the number of links, C is the maximum of link capacities, K is the number of flow classes, F is the number of iterations in the Fixed Point method, and S is the number of iterations of the secant method. This complexity arises from the fact that the computation in (8) is done for all L links, with the product in (8) being $O(L)$, the sum in (8) being $O(K)$, and the computation of the Erlang function being performed in $O(C)$. From our experiments in Section 3, for an approximation error of 0.01, F typically falls between 3 and 100, and S typically falls between 10 and 50.

Another problem with this network comparison method is that it yields the Network Load Ratio $R(p)$ of two networks for a given blocking probability p . The point value $R(p)$ may not provide much insight into the relative behavior of the two networks for a range of network blocking probability values p .

After a series of empirical experiments with various networks of different topologies and loads, we have come to the conclusion that the Network Load Ratio $R(p)$ generally exhibits little variation in a range of values for p that are of interest to us. For example in Figure 3 we plot the network blocking probability of four networks with 17 access nodes and four different topologies: the complete graph, tandem, star and a model (that we labeled “N17”) of the NSF Backbone (see Figure 2) as it was in 1995. In each network, for each pair of nodes (and thus for each flow class k), we take $\nu_k = \nu, \forall k \in \mathcal{K}$. We observe that the Network Load Ratio is almost constant for blocking probabilities in the range $p \in [10^{-5}, 0.01]$. In the following we provide a formal motivation for this empirical observation and derive a second method for comparing two networks.

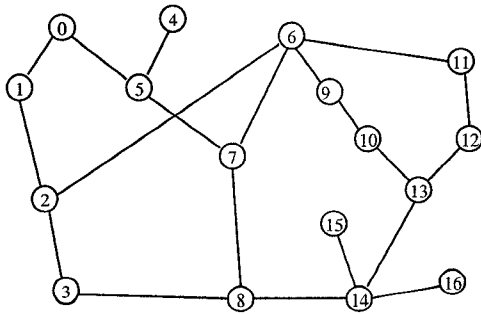


Figure 2: Topology of N17 (a model of NSF Backbone)

We observe (as illustrated in Figure 3) that a “small” value of network blocking probability (i.e., $p \in [10^{-5}, 10^{-1}]$) does not necessarily correspond to a small

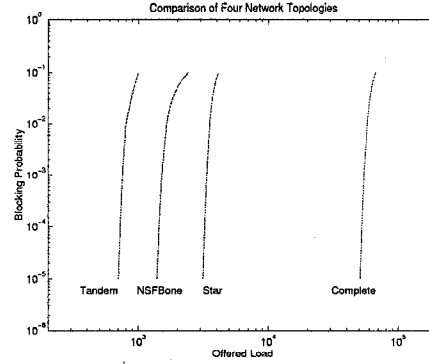


Figure 3: Comparison of four network topologies

value of ρ (in Figure 3, $\rho \in [10^2, 10^5]$). Thus, we consider the limit of the network load ratio as $p \rightarrow 0$, and not as $\rho \rightarrow 0$ (ρ approaches 0 more slowly).

Let us consider two networks N_1 and N_2 having all parameters defined earlier.

Definition 2 The limit of the Network Load Ratio $R(N_1, N_2, p)$ as $p \rightarrow 0$,

$$A_0(N_1, N_2) = \lim_{p \rightarrow 0} R(N_1, N_2, p) = \lim_{p \rightarrow 0} \frac{P_1^{-1}(p)}{P_2^{-1}(p)}$$

is the Low-Blocking Load Ratio of networks N_1 and N_2 .

Observe that A_0 is a function of link capacities C_i . In [5] we establish the following result.

Proposition 1 If all link capacities are the same in both networks, $C_i = C, \forall i \in \mathcal{L}_q, q = 1, 2$, then the limit to the Low-Blocking Load Ratio as $C \rightarrow \infty$, named Asymptotic Load Ratio, is:

$$A(N_1, N_2) = \lim_{C \rightarrow \infty} A_0(N_1, N_2) = M_2/M_1 \quad (11)$$

where

$$M_q = \max_{i \in \mathcal{L}_q} \eta_i \quad q = 1, 2 \quad (12)$$

Intuitively, Proposition 1 says that the relative performance of two networks in underload conditions is given by the ratio of loads on their bottleneck links. This result confirms the heuristic of “minimizing the maximum congestion” used in several alternate routing algorithms in circuit switching networks (see [17], Section 7.1 and references therein).

Proposition 1 yields a simple approximation of the Asymptotic Load Ratio when link capacities are large ($C > 100$, as is the case with most current networks).

For the general case, where the values of C_i are not restricted, we make the following conjecture:

Conjecture 1 *The Low-Blocking Load Ratio is well approximated by the Asymptotic Load Ratio*

$$A(N_1, N_2) = M_2/M_1 \quad (13)$$

where

$$M_q = \max_{i \in \mathcal{L}_q} \eta_i / C_i \quad q = 1, 2 \quad (14)$$

In Section 3 we present a set of experiments where we find the Asymptotic Load Ratio to be a good approximation for the Network Load Ratio.

The computation of the Asymptotic Load Ratio given in equations (11), (12), (13) and (14), constitutes our second proposed method for network comparison. Note that the complexity of the computation is $O(LK)$ where L is the number of links in the two networks and K is the number of flow classes. This makes the computation of the Asymptotic Load Ratio much simpler than that of Network Load Ratio.

3 Accuracy of Asymptotic Load Ratio

In this section we evaluate the accuracy of the Asymptotic Load Ratio. First, we verify that the Asymptotic Load Ratio, computed as in (11), (12) and (13), (14), is the limit of Network Load Ratio as the network blocking probability approaches zero. Second, we seek to determine the range of network blocking probabilities for which the Asymptotic Load Ratio is a good approximation to the Network Load Ratio. Third, we measure and compare the computation times incurred while computing the Network Load Ratio and Asymptotic Load Ratio.

We perform the evaluation through a set of simulation experiments, where the two methods are applied to pairs of randomly generated network topologies. We use networks generated at Georgia Tech using the methods proposed in [19], here named RAND, and in [21], named Transit-Stub (TS).¹ The RAND networks have unstructured topologies, whereas the TS networks exhibit hierarchical structures and smaller diameters, which are claimed [21] to more accurately resemble real networks. All networks contain 100 nodes. All links in the RAND networks have OC3 (155 Mb/s) capacity, whereas only the backbone links in the TS networks have OC3 capacity and the rest, T3 (45 Mb/s) capacity. The bandwidth of a flow is 100 Kb/s. The shortest path routing policy is used to generate the routes.

In the first experiment, 10 networks of each network type were considered, and compared to each other using the two comparison methods. Specifically, the Network Load Ratio $R(N_i, N_j, p)$ is computed for each pair of networks N_i, N_j and for the following values of network

blocking probability

$$p \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.02, 0.03, 0.04, 0.05, 0.07, 0.1\}$$

For each network pair N_i, N_j , the Asymptotic Load Ratio $A(N_i, N_j)$ is also computed. We then compute the relative error of Asymptotic Load Ratio compared to Network Load Ratio for each network blocking probability value and each pair of networks

$$e_{N_i, N_j}(p) = \frac{A(N_i, N_j) - R(N_i, N_j, p)}{R(N_i, N_j, p)}$$

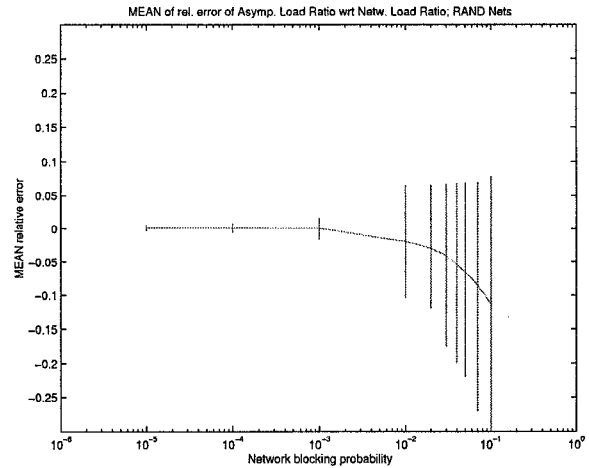


Figure 4: Mean of relative error of Asymptotic Load Ratio with respect to Network Load Ratio; RAND networks

In Figure 4 we plot the results for RAND networks. For each value of network blocking probability p , we plot the average of the relative error values $e_{N_i, N_j}(p)$, over the set of all pairs (N_i, N_j) of networks. We also plot a vertical bar indicating the 5-percentile and the 95-percentile of the set $(e_{N_i, N_j}(p))_{i, j}$. We observe that the relative error converges to zero as $p \rightarrow 0$, which confirms Proposition 1 that the Asymptotic Load Ratio is the limit of Network Load Ratio. We also note that the 90-percentile of the relative error is less than 0.12 for a network blocking probability less than 0.02, indicating that the Asymptotic Load Ratio is a good estimate for the Network Load Ratio in this range. In [5] we show similar results for TS networks, thus confirming Conjecture 1.

In [5] we also show results from a second experiment where we treat each of the networks considered in the first experiment as a base network. Starting from each base network we create 10 “supranets” by adding a random number of links (between 1 and 10). We then compute the Network Load Ratio and Asymptotic Load Ratio for each base network and supranet combination. The relative error statistics are processed as in the first experiment, and we obtain

¹The networks and the generating code can be found at <http://www.cc.gatech.edu/fac/Ellen.Zegura/graphs.html>

similar results: the 90-percentile of the Asymptotic Load Ratio is within 10% of Network Load Ratio for $p < 0.02$.

For the above experiments we have also recorded the time required to compute the Network Load Ratio and Asymptotic Load Ratio for each pair of networks compared. The average computation time has been measured with the Unix code profiler *prof* on an 100MHz SGI workstation. Table 1 displays the average computation times for RAND and TS networks respectively. The difference in computation time in favor of Asymptotic Load Ratio confirms the simplicity predicted at the end of Section 2.

Table 1: Comparison of computation times for Network Load Ratio and Asymptotic Load Ratio

	RAND	TS
Network Load Ratio	103.927 s	129.676 s
Asymptotic Load Ratio	0.008 s	0.012 s

In conclusion, we consider the Asymptotic Load Ratio to be preferable to the Network Load Ratio for comparing network performance for two reasons:

- The Asymptotic Load Ratio is empirically found to give a good approximation (the 90-percentile of the relative error is less than 0.12) for the relative performance of two networks (the Network Load Ratio) for a range of network blocking probability values ($p \in (0, 0.02)$).
- The Asymptotic Load Ratio is very simple computationally (four orders of magnitude faster) compared to the Network Load Ratio method.

4 Evaluating the Benefit of ATM Shortcuts on Random Networks

In this section we study the sensitivity of the Asymptotic Loss Ratio (ALR) with respect to several topological characteristics of the IP and ATM networks such as average route (path) length or average node degree. For this purpose we have performed simulation experiments where we generate IP and ATM network topologies using the Georgia Tech network generator [21]. We computed the networks' topological characteristics and ALR, and derived statistical dependencies of ALR on the networks' characteristics.

The objective of these experiments is to generate pairs of (IP and ATM) networks with realistic topologies. The current overlaid IP/ATM networks exist mostly in the Internet backbones (such as vBNS [8]). ATM backbone topologies are expected to carry both telephone and data traffic. Hence, we expect ATM networks to be designed

similar to telephone networks. Telephone networks backbones are "nearly fully connected" ([11], page 10): the average node degree is high and the average path length is small. By contrast, IP networks (including backbones) are usually "less connected". Also, backbones usually have a flat (non-hierarchical) topology. In our experiments we generate networks with these observations in mind.

We generate random networks where a link is generated between two nodes with probability α . Thus, α is a parameter that influences the "connectedness" of the network. We generate ATM networks with $N_A \in \{16, 32\}$ nodes and with connectedness parameter $\alpha_A \in \{0.2, 0.4, 0.6, 0.8\}$. We generate 10 ATM networks for each combination of (N_A, α_A) values. All links have capacity OC3 (155Mb/s). The IP networks are generated with $N_I = q * N_A$, where $q \in \{0.25, 0.5, 0.75, 1.0\}$, and using a connectedness parameter $\alpha_I \in \{0.25, 0.5, 0.75, 1\} * \alpha_A$. This produces 16 sets of (N_I, α_I) values for each (N_A, α_A) pair, and we generate 10 IP networks for each set of (N_I, α_I) values.

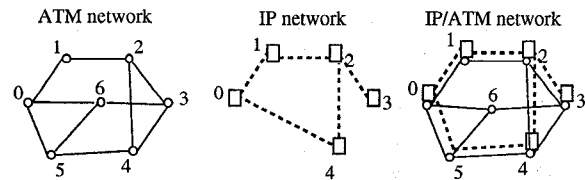


Figure 5: Overlying an IP network over an ATM network

For each pair of ATM and IP networks, we overlay IP over ATM by mapping IP nodes to ATM nodes and the IP links to the shortest paths (hop count) in the ATM network (see example in Figure 5). For each overlaid IP/ATM network, the load consists of IP flows requesting reservations of 100Kb/s between any pair of IP nodes with the same intensity $\phi_k = \phi, \forall k \in \mathcal{K}$ (ϕ_k as defined in (2)). The IP flows can be routed in two ways: one is over the shortest path in the IP network following the path of each IP link in the ATM network; the other is over the shortest path in ATM network (i.e., using ATM shortcuts). For example, in Figure 5 an IP flow between nodes 0 and 3 has route 0 - 1 - 2 - 3 if routed in IP, and route 0 - 6 - 3 if routed in ATM (i.e., 0 - 6 - 3 is an ATM shortcut).

We evaluate the benefit of ATM shortcuts for each IP/ATM overlaid network by computing the ALR for each of the two ways of routing of IP flows. Also, for each IP/ATM overlaid network, we compute the following topological characteristics for each of the two ways of routing of IP flows:

- N_A = the number of ATM nodes, N_I = the number of IP nodes (these are also ATM nodes);

- B_A, B_I , the average degree = average number of links connected to each node;
- C_A, C_I , the normalized average degree, $C_A = B_A/N_A, C_I = B_I/N_I$;
- D_A, D_I , the diameter = length of longest route;
- E_A, E_I , the average depth = average route length;
- F_A, F_I , the number of biconnected components. Here a biconnected component is a maximal set of edges such that any two edges in the set are on a common simple cycle.

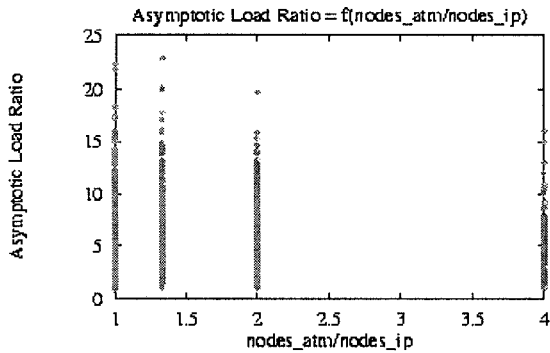


Figure 6: Asymptotic Load Ratio versus Node Ratio

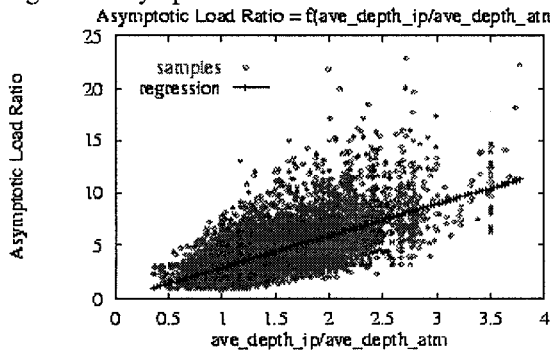


Figure 7: Asymptotic Load Ratio versus Average Depth Ratio

In Figures 6 and 7, we display scatter plots of ALR as a function of ratios of number of nodes and average depth respectively, where each point represents an IP/ATM network pair (we have 12800 points in each figure). We seek to establish if ALR depends on any of these topological ratios. First, we observe that the node ratio N_A/N_I has little influence on ALR (see Figure 6), whereas ALR is clearly correlated with the ratio of average depth (see Figure 7). In order to quantify the correlation, we present, in Table 2, the coefficients of correlation between ALR and each of the topological ratios. These values indicate a weak correlation of ALR with respect to node ratio and average degree

Table 2: Coefficient of correlation between Asymptotic Load Ratio and various topological ratios

	Coeff. of correlation w.r.t. ALR
(N_A/N_I)	0.232
(B_A/B_I)	0.164
(C_A/C_I)	0.469
(D_I/D_A)	0.610
(E_I/E_A)	0.672
(F_I/F_A)	0.622

Table 3: Coefficient of correlation between various topological ratios

	(C_A/C_I)	(D_I/D_A)	(E_I/E_A)	(F_I/F_A)
(C_A/C_I)	1.000	0.791	0.746	0.364
(D_I/D_A)	0.791	1.000	0.930	0.604
(E_I/E_A)	0.746	0.930	1.000	0.686
(F_I/F_A)	0.364	0.604	0.686	1.000

ratio, and a significant correlation with respect to diameter ratio, average depth ratio and number of biconnected ratio. The linear regression computed using the least square error method is depicted by a line in Figure 7.

A further question to ask is what is the cumulative dependence of ALR on diameter, average depth, normalized average degree and number of biconnected ratios. If the parameters exhibited independent effects on ALR, we could perform a multiple linear regression [7]. In Table 3 we present the coefficient of correlation between the four topological ratios of interest. We observe a strong correlation between the diameter ratio and the average depth ratio (it is somewhat natural for them to be related, since the diameter and average depth measure the maximum and average of the same set: the path lengths). Given this, we can eliminate the diameter ratio from the multiple linear regression. Consequently, we have computed the following multiple linear regression:

$$ALR = 0.5 + 1.92 * (E_I/E_A) + 0.26 * (F_I/F_A) + 0.08 * (C_A/C_I)$$

We observe in the above regression that ALR has only a weak dependence on the Normalized Average Degree Ratio. By also observing in Table 3 a significant correlation between Normalized Average Degree Ratio and Average Depth Ratio, we conclude that we may drop it from the multiple linear regression, which then becomes:

$$ALR = 0.5 + 2.0 * (E_I/E_A) + 0.25 * (F_I/F_A)$$

In conclusion, we observe the following trends regarding the benefits of ATM shortcuts in IP/ATM networks:

- The network using ATM shortcuts is likely to accept more IP flows than the network not using them provided the network not using shortcuts has a high average depth or large number of bicomponents.
- The ATM shortcut benefits are generally not influenced by the number of nodes or the average degree of the IP/ATM network, with or without ATM shortcuts.

5 Conclusion

In this paper we have considered the problem of evaluation of benefit of ATM shortcuts in IP/ATM networks, i.e., the benefit of ATM routing of IP flows in ATM networks. We proposed to measure the benefit of ATM shortcuts with the Network Load Ratio, that expresses the increase in the number of flows accepted by an IP/ATM network for the same network blocking probability. We developed a low complexity computation for Asymptotic Load Ratio, which estimates the Network Load Ratio relatively well (the 90-percentile of the relative error is less than 0.12) for underload network conditions (network blocking probability less than 0.02). An important result of this paper is a methodology for comparing network performance, which can be used to evaluate the benefit and tradeoff of ATM shortcuts, and in the more general context of network design. We use this method in simulation experiments using random networks. These experiments indicate that in many cases the utilization of an IP/ATM network increases proportionally to the decrease in the average path length when ATM shortcuts are used. We have also found that the decrease in the number of biconnected components of the IP/ATM network when using ATM shortcuts has also an influence on the network utilization, although not as strong as the average path length. Moreover, we have found that there is almost no correlation between the increase in network utilization (when using ATM shortcuts) and the IP to ATM node ratio.

This work can be extended in several directions. First, the conjecture that extends the Asymptotic Load Ratio to networks with heterogeneous capacities is an open problem that needs to be proved. Second, the results and methods may also be extended to networks with alternate routing, flows with heterogeneous bandwidth requirements, and multicast flows.

References

- [1] G. Armitage. Support for Multicast over UNI 3.0/3.1 based ATM Networks. Internet RFC 2022, November 1996.
- [2] A. Birman, V. Firoiu, R. Guérin, and D. Kandlur. Support for RSVP-based Services over ATM Networks. In *IEEE Global Internet*, pages 10–15, 1996.
- [3] E. Crawley, R. Nair, B. Rajagopalan, and H. Sandick. A Framework for QoS-based Routing in the Internet. Internet draft, work in progress, draft-ietf-qosr-framework-05.txt, May 1998.
- [4] A. Feldmann, J. Rexford, and R. Caceres. Reducing Overhead in Flow-Switched Networks: An Empirical Study of Web Traffic. In *IEEE INFOCOM*, 1998.
- [5] V. Firoiu, J. Kurose, and D. Towsley. Performance Evaluation of ATM Shortcut Connections in Overlaid IP/ATM Networks. Technical Report UM-CS-1997-040, University of Massachusetts, Amherst, revised 1998. <ftp://ftp.cs.umass.edu/pub/techrept/techreport/1997-/UM-CS-1997-040.ps>.
- [6] A. Greenberg and R. Srikant. Computational Techniques for Accurate Performance Evaluation of Multirate, Multihop Communication Networks. *IEEE/ACM Transactions on Networking*, 5(2):266–290, April 1997.
- [7] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley & Sons, 1991.
- [8] J. Jamison and R. Wilder. vBNS: The Internet Fast Lane for Research and Education. *IEEE Communications Magazine*, 35(1), Jan 1997.
- [9] F. Kelly. Blocking Probabilities in Large Circuit-Switched Networks. *Adv. Appl. Prob.*, 18, 1986.
- [10] F. Kelly. Loss Networks. *The Annals of Applied Probability*, 1(3), 1991.
- [11] S. Keshav. *An Engineering Approach to Computer Networking*. Addison-Wesley, 1997.
- [12] S. Lin and N. McKeown. A Simulation Study of IP Switching. In *ACM SIGCOMM*, 1997.
- [13] G. Louth. *Stochastic Networks: Complexity, Dependence and Routing*. PhD thesis, University of Cambridge, 1990.
- [14] J. Luciani, D. Katz, D. Piscitello, and B. Cole. NBMA Next Hop Resolution Protocol (NHRP). Internet draft-ietf-rolc-nhrp-11.txt, 1997.
- [15] P. Newman, T. Lyon, and G. Minshall. Flow Labelled IP: a Connectionless Approach to ATM. In *IEEE INFOCOM*, 1996.
- [16] R. Cherukuri and D. Dykeman (eds.) and M. Goguen (chair). ATM PNNI Draft Specification. ATM Forum 94-0471, November 1995.
- [17] K. Ross. *Multiservice Loss Models for Broadband Telecommunication Networks*. Springer, 1995.
- [18] P. Samudra. Draft of UNI Signaling 4.0. ATM Forum 94-1018 R4, July 1995.
- [19] B. M. Waxman. Routing of Multipoint Connections. *IEEE Journal on Selected Areas in Communications*, 6(9):1617–1622, December 1988.
- [20] W. Whitt. Blocking when Service is Required from Several Facilities Simultaneously. *AT&T Technical Journal*, 64(8), October 1985.
- [21] E. Zegura, K. Calvert, and S. Bhattacharjee. How to Model an Internetwork. In *IEEE INFOCOM*, 1996.