Do We Need Resource Reservation?

Ping Pan
Juniper Networks
1194 N. Mathilda Avenue,
Sunnyvale, CA 94089
pingpan@juniper.net

Henning Schulzrinne
Computer Science Department
Columbia University
New York, NY 10027
schulzrinne@cs.columbia.edu

April 25, 2001

Abstract

Reservation reservation has been a very controversial issue among academic scholars and network operators. This paper examines the question of whether or not the Internet is better off with resource reservation from several aspects. An analytical model shows that reservation-capable networks provide better performance for real-time traffic than that in best-effort-only networks. A study on today's Internet traffic volume, bandwidth pricing, and network topology indicates that network heterogeneity could become a significant problem that prevents ISP's from delivering predictable services to end users. Finally, we look into some of the potential network applications, and conclude that supporting resource reservation is a key to the deployment of those applications.

We don't claim that we have given a definite proof on the necessity of resource reservation, and that the data we have collected represents today's traffic condition in all ISP networks. However, we believe our results provide strong evidences that network over-provisioning has its limitation, and reservation protocols can be implemented and deployed with manageable impact on network's scalability and on router's overall performance. We hope that such evidence can foster more research on Internet reservation protocols.

1 Background

In 1993, Zhang *et al.* [1] described *resource reservation* as an important component for providing service guarantees in the Internet. Resource reservation is to set aside network resources (*e.g.,* link bandwidth and router's packet buffers) for specific user packet streams, or *flows*. To obtain a reservation, users signal the network to request the amount and the quality of service; the network in turn decides to whether or not it can satisfy the request. RSVP had been proposed as the signaling protocol that is responsible for setting up reservations in the network. Despite it's standardized several years ago, there appears to be little commercial deployment of RSVP.

Over the years, many people [2, 3] have been debating the issue of whether or not there is a need for resource reservation in the Internet in the first place. The arguments against resource reservations include:

Over-provisioning is good enough: Because bandwidth is becoming a cheap commodity, the networks can afford to be over-provisioned to eliminate link congestion and to reduce packet transmission latency. This popular argument seems to be confirmed by the fact that several large backbone providers, UUnet and Verio, have been deploying high-speed links (up to OC-192 in 2000), and used over-provisioning to overcome link jitter and congestion. It is important to note here that this argument is based on the expectation that, as more advanced technology (such as optical networking and gigabit Ethernet) becomes available, the bandwidth price will be driven down rapidly and consistently in the near future.

Applications can adapt: End users can use different adaptation mechanism, such as buffering, signal reconstruction, and adaptive compression ratio, at application layer to adapt to network delay and packet losses. For example, NeVot [4] has built-in buffering mechanism to adjust for play-out delay, and vat [5] can reconstruct lost signal.

Real-time traffic volume is too small: Most of the network traffic is *elastic*, and has loose time constraints. Most of the multimedia data (video and audio) transmission should not be considered as real-time either, since users can always down-load or cache data in advance. Therefore, engineering a network to accommodate small amount of inter-active real-time traffic seems to be overkill.

Adding processing cost to routers: Processing control messages, maintaining reservations states and running admission control contribute to heavy processing and memory overhead at routers. This argument has gained a lot of support during the design and implementation of RSVP [6] and IntServ [7, 8, 9, 10].

Adding complexity to service providers: Reservations categorize user traffic into multiple “flows”, each must have a different accounting and billing procedure, and needs to be handled separated by the ISP’s. Given that the sheer number of “flows” can be very large, a scalable management and billing system would be complicated and costly to develop.

Do we need to use reservations in the Internet? The answer to this question depends on many factors, such as, bandwidth cost, Internet growth rate, service deployment process, and the emergence of “killer” applications in the future that can significantly change traffic pattern and traffic dynamic. In this paper, we will argue the necessity of resource reservation from several aspects, and evaluate some of the flaws in the above argument.

2 Performance Evaluation for Real-Time Applications

Shenker and Breslau have studied [11, 12] the performance of real-time applications in the Internet through the use of a utility function formulation and several simple models. They concluded that depending on traffic pattern, reservation-capable networks retain significant advantages over best-effort-only networks, no matter how cheap bandwidth is.

Two types of real-time application are defined here: *rigid* and *adaptive*. The rigid applications are extremely sensitive to delay and have hard real-time requirements. These applications need their data to arrive within a given delay bound. An example of such an applications is the traditional telephone service.

Adaptive applications, on the other hand, can be audio and video applications that are implemented to tolerate occasional delay-bound violations and dropped packets.

The utility functions for both rigid and adaptive applications are illustrated in Figure 1. The utility function for a rigid application flow is

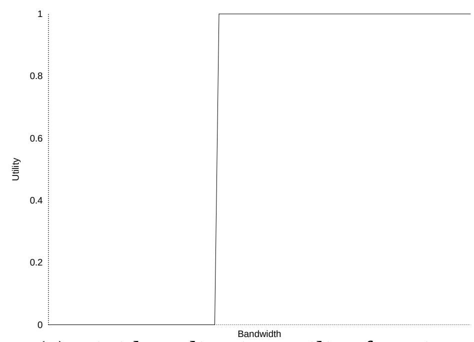
$$\pi(b) = \begin{cases} 0 & \text{for } b < \overline{\mathrm{b}} \\ 1 & \text{for } b \geq \overline{\mathrm{b}} \end{cases}$$

where $\overline{\mathrm{b}}$ is the bandwidth needed for real-time applications.

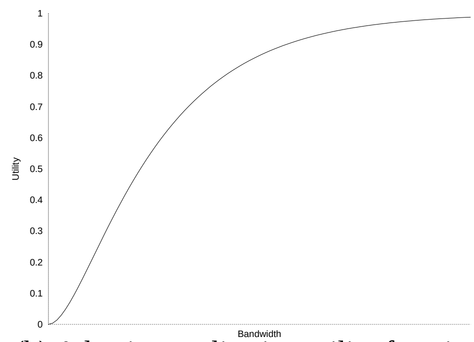
For applications that can adapt to both changes in rate and delay, the utility function is defined as:

$$\pi(b) = 1 - e^{\frac{-b^2}{\kappa + b}}$$

where κ is a constant. Similar to other elastic applications (file transfer and email), at high bandwidth, there is only a marginal performance gain with additional bandwidth. Nevertheless, the adaptive applications always have an intrinsic bandwidth requirement, and this intrinsic generation rate is independent of the network congestion. The performance degrades badly as soon as the bandwidth share becomes smaller than the intrinsic rate. Thus the network can become overloaded, though the applications can adapt to network congestion.



(a) Rigid application utility function



(b) Adaptive application utility function

Figure 1: The utility function for rigid and adaptive applications.

The fundamental difference between a best-effort-only network and a reservation-capable one is that, in the former, user flows are never denied access to the network, whereas a reservation-capable network can deny reservation requests.

Let $P(k)$ be the possibility that there are k flows requesting service, and C be the link capacity. The average number of the requesting flows in the network is $\bar{k} = \sum_{k=0}^{\infty} P(k)k$.

Hence, the total utility of a best-effort-only network is given by:

$$V_B(C) = \sum_{k=1}^{\infty} P(k)k\pi(\frac{C}{k})$$

For a reservation-capable network that can admit at most $k_{max}(C)$ flows, the total utility becomes

$$V_R(C) = \sum_{k=1}^{k_{max}(C)} P(k)k\pi(\frac{C}{k}) + \sum_{k=k_{max}(C)+1}^{\infty} P(k)k_{max}(C)\pi(\frac{C}{k_{max}(C)})$$

The normalized utility functions for best-effort-only and reservation-capable networks are respectively,

$$B(C) = \frac{V_B(C)}{\bar{k}}$$

and

$$R(C) = \frac{V_R(C)}{\bar{k}}$$

By modeling $P(k)$ as Poisson, exponential and algebraic load distributions[1], they observed the following important insights:

- **Performance and bandwidth benefits:** The best-effort-only networks may be simple to build, but they require additional bandwidth in order to match the performance of the reservation-capable networks. This additional bandwidth is defined as the *bandwidth gap*, $\Delta(C)$, such that, $R(C) = B(C + \Delta(C))$.

  When evaluating the utility functions and the bandwidth gaps for both rigid and adaptive applications, it was found that, not surprisingly, rigid applications perform much better in a reservation-capable network than that in a best-effort-only network. However, surprisingly, for Poisson and exponential load distribution, the performance results for adaptive applications are almost identical in both best-effort-only and reservation-capable networks. The bandwidth gap disappears at high link capacity region. This behavior demonstrates the advantage that adaptive applications can tolerate network overload conditions, and raised the issue whether reservations have any advantage. However, when using algebraic load distribution in the model, the bandwidth gap between two networks is a con-

[1] In the paper [12], the Poisson distribution is defined as $P(k) = \frac{\nu^k e^{-\nu}}{k!}$, and the exponential distribution is $P(k) = (1 - e^{-\beta})e^{\beta k}$. The algebraic distribution is defined as $P(k) = \frac{\nu}{\lambda + k^z}$. Note, when setting λ to be zero, it becomes a simple Pareto distribution, which has been used to model the *self-similarity* behavior of Internet traffic. For detailed mathematical analysis, refer to the paper directly. We only present the model and the results here for clarity.

stant[2] regardless link capacity. This implies that the best-effort-only networks always require more bandwidth than reservation networks to match their performance.

- **Cost and bandwidth tradeoff:** The reservation-capable networks impose the burden of additional complexity compared to best-effort-only networks. In an attempt to estimate the complexity, the authors present a simplified economic analysis, and convert the cost of extra complexity in building a reservation-capable network to the cost of additional bandwidth.

  They first define the welfare provided by the network as:

  $$W(p) = V(C(p)) - p\,C(p)$$

  where p is the cost per unit bandwidth. When setting the welfare of reservation-capable and best-effort-only networks to be equal, (*i.e.*, $W_R(\bar{p}) = W_B(p)$), $\bar{p}$ is the cost per unit bandwidth in reservation-capable networks, and the ratio $\gamma(p) = \frac{\bar{p}}{p}$ would provide the quantified information on added complexity in reservation-capable networks.

  It was shown that when using adaptive applications, for Poisson and exponential distributions, $\gamma(p) \to 1$ as $p \to 0$. This implies that as bandwidth becomes cheaper, $\bar{p} \to p$, *i.e.*, if adding reservation capability to the network introduces any additional per-unit bandwidth cost, it would make the best-effort-only network the preferable choice.

  However, in contrast, in the algebraic case, $\gamma(p)$ is always greater than one, and does not converge. Thus, when by introducing a small amount of additional per-unit bandwidth cost, a reservation network is always the preferable choice over a best-effort-only one, no matter how inexpensive the bandwidth becomes.

In conclusion, the work showed significant performance improvement and bandwidth gaps between best-effort-only and reservation-capable networks for rigid applications. For adaptive applications, with Poisson and exponential load distributions, there is no strong need for reservations, but with algebraic load distribution, reservations yielded significant benefits.

Hence, the need for reservation-capable networks seems to be depending on the traffic pattern in the future. Some of the recent studies [13, 14, 15, 16] have shown that the Internet traffic pattern exhibits *self-similarity* behavior. A close approximation of self-similarity is the Pareto distribution, which is a variation of the algebraic distribution we have defined here. This seems to lead us to believe that if network traffic behavior continues the current trend, reservation-capable networks would provide more benefits than that of best-effort-only networks, in terms of performance and cost.

3 Current Network Condition

Despite the claims by some providers [17, 18] that the networks have little jitter and congestion, end-user traffic suffers from significant delay and jitter.

[2] For $z \to 2^+$, the constant is e.

Paxson had formally analyzed the dynamics of Internet traffic from over 20,000 TCP connection traces among 35 widely spread sites from 1994 to 1997 [19, 20]. Specifically, he defined the notion of *available bandwidth* as the proportion of the total network resources that were consumed by an end-to-end connection itself. Values of 1 mean that bandwidth is available on all physical links that are used by the connection, whereas values of 0 imply that none of the link bandwidth is available. From his measured data, we found the following interesting results:

- Available bandwidth varies widely, from very little to almost 1.
- The end-to-end packet delay variations occurred primarily on time scale of 100-1,000 ms, but extended out quite frequently to much larger times.

The results imply the following: the network has both high-bandwidth links, where packets can sail through without any disturbance, and bandwidth *bottleneck* links. When packets go through the bottleneck links, the transmission can take very long time.

The range of the delay variation was particularly troublesome: to transmit one-way voice data, the maximum allowed end-to-end delay is approximately 200 ms [21]. Let's estimate that the total time for packetization and coding/decoding is approximately 100 ms according to the ITU recommendation [22]. This leaves approximately 100 ms, which is the lower bound in Paxson's measured delay variation, to be the total network latency allowed for voice packets. This implies that the end-user applications need to absorb as much as 900 ms of delay in order to support real-time voice traffic.

Hence, it is reasonable to claim that, *to support delay-sensitive end-user applications in the Internet, unless the network can always transmit user packets over the links that have plenty of bandwidth, the applications will suffer from jitter.*

This raised the following questions: where are the bandwidth bottleneck links? Can we simply improve the bandwidth on these bottleneck links? How real is the problem that end-user traffic actually travels through the bottleneck links? We will try to provide some insights to these questions in the following sections.

3.1 Bandwidth Bottlenecks

Network links cover a large spectrum of speeds. Typically, backbone networks have link speeds ranging from OC-3 (155 Mb/s) to as high as OC-192 (10 Gb/s), and likely to operate at higher speed in the future, as the next-generation optical equipment being deployed. Often, a backbone connection between two nodes is installed in such a way of having multiple physical links or wavelength in parallel.

Within private and local-area networks, each user normally has at least 10 Mb/s (Ethernet LAN's) of bandwidth to share, normally with a small number of other users. Most of the residential dial-up users run at 56 kp/s modem speed. With the emergence of high-speed LAN's (100 Mb/s Ethernet) and DSL/cable modem technologies, end users will eventually have more bandwidth to access the Internet. For wireless users, currently, the cellular (GSM) technology can only provide approximately 10 kb/s of bandwidth. However, in 3G, user data rates vary extensively depending on the technology (GPRS, 3G-1x, HDR, etc.). In general, wireless data rates can range from 64 kb/s all the way up to almost 2 Mb/s.

From recent measurement results [23], it was estimated that private, local-area and backbone networks are lightly loaded, as shown in Table-1. The particular reason for light traffic load within corporate networks is to maintain a low transaction latency within the network. This seems to be saying that so long user traffic stays in the same network, latency and jitter are not a problem. In other words, if there is any congestion in the Internet at all, it's most likely coming from network interconnecting places.

Network Type	Average Utilization
Backbones	10-15%
Private Networks	3-5%
LAN's	1%

Table 1: Network utilization estimation from [23].

We refer to the lines that connect private networks to the Internet, or interconnect provider networks at POP's, NAP's and private peering points, as *access links*. It is believed that the access links are the main points of congestion. The NAP's are particularly problematic. Unfortunately, most of the commercial ISP's do not publicly reveal the traffic statistics on their access links. We have to rely on traffic traces from several academic and research networks to piece together the link utilization conditions.

SWITCH [24] is a regional ISP that provides Internet connectivity to Swiss academic and research institutions. Presently, the backbone consists of E3 (34 Mb/s) and OC-3 trunks. Table 2 shows the traffic statistics on some its access links. The measurement was made by collecting traffic counters from the routers every 5 minutes using SNMP. From the collected data, we notice that although the average link utilization is low, the peak utilization at the NAP and the direct peering links are quite high. In fact, the peak utilization at CERN NAP had reached the maximum link capacity, since the NAP is operating on an Ethernet. Unfortunately, there is no data available for us to determine congestion duration.

Link Type	Average Utilization		Peak Utilization	
	In	out	In	out
Transatlantic OC-3 Link (Zurich - New York)	16.1%	5.6%	49.4%	32.1%
Trans-Europe OC-3 Link	10.5%	17.0%	23.9%	48.9%
CERN NAP, Geneva (100M Ethernet)	13.7%	25.1%	70.3%	89.6%
Direct Peering (to Swisscom) (100M Ethernet)	3.0%	12.5%	64.8%	67.0%

Table 2: Monthly link utilization in SWITCH access links on September, 2000.

We then investigated traffic volume and congestion condition in a larger backbone network, NORDUnet [25], which interconnects the Nordic national networks to the rest of the world. The backbone itself is a mix of OC-3 and OC-12 trunks. NORDUnet connects to several NAP's, including Chicago NAP, and is peering with several other large provider backbones, such as TeleGlobe, Telia and FUNET. Most of the access links are OC-3 trunks, except that the Chicago NAP link runs at DS-3 speed, and the link to DGIX NAP is a 100 Mb/s Ethernet connection. Similar to other providers, NORDUnet uses multiple links to peer with other providers, and traffic to and from other providers is evenly distributed across all the links. We selected the 12 busiest links out of a total of 36 access links in the backbone, and analyzed the peak and average link utilization shown in Figures 2, 3 and 4.

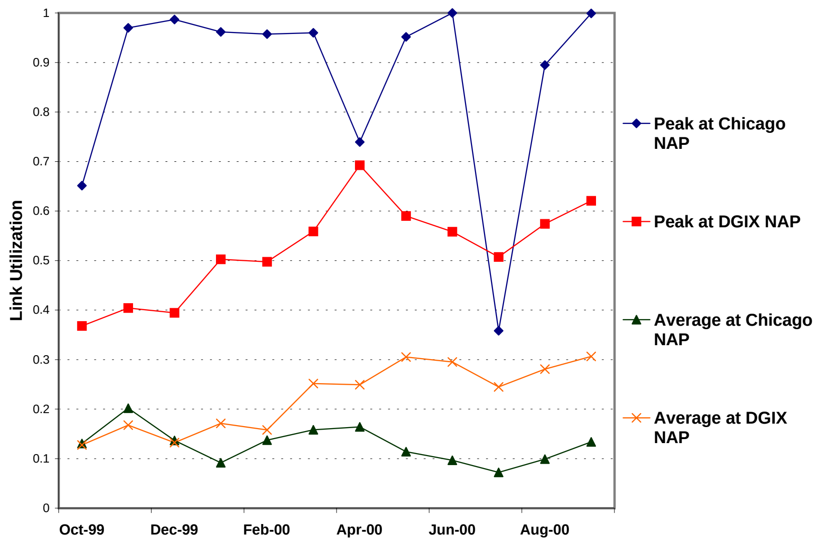
We observe the following interesting results:

- Both NAP (Figure 2) and private peering links (Figure 3) can get congested. This is somewhat surprising, since it has been perceived that unlike NAP's, private peering links are better managed and thus have less congestion.
- Trans-atlantic links are a part of the provider's backbone, and have had very high peak utilization (Figure 4). This simply implies that there are bottleneck links within provider networks, and user traffic that stays within the same provider's network can also get congested.
- Most importantly, the average link utilization is quite low, around 20% to 30%, however, the peak link utilization is high on all links. Some of the links had reached 100% link utilization. The congestion occurrence varies depending on links. From the statistics maintained at NORDUnet, the Chicago NAP link operated at peak only once in August, 2000, which lasted approximately 7 hours. The DGIX NAP link operated at peak (over 80% link utilization) five times in September, 2000, and only once in August, 2000. Each congestion in DGIX lasted between around 1 hour to as many as 10 hours. We have also observed frequent and long-lasting congestion on several peering links. During August and September, 2000, the TeleGlobe links had operated at peak (over 80% link utilization) almost every workday. Each peak lasted about 8 hours.

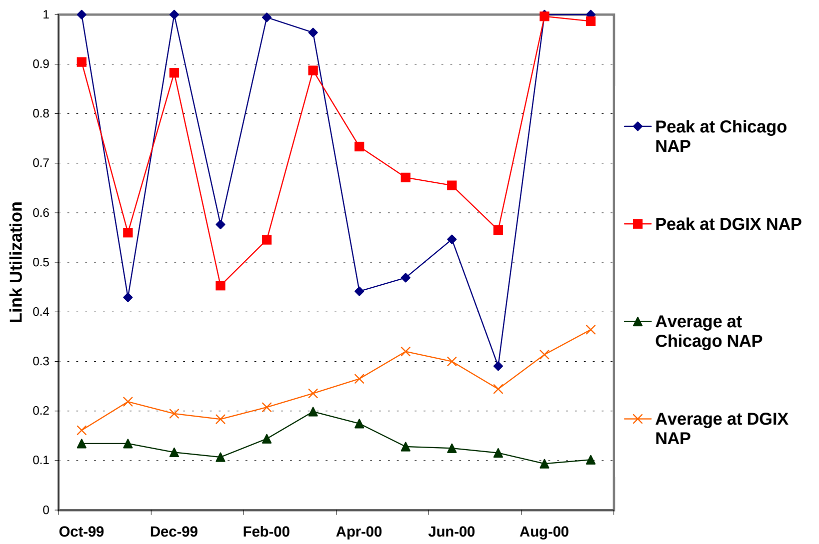
To understand the reasons for link congestion, we studied the link traffic history extensively, and exchanged e-mails with the technical support in NORDUnet [3]. We discovered that network links are being frequently re-configurated and upgraded. One of its reasons is that providers regularly re-arrange or shift user traffic between links to balance traffic load and optimize bandwidth usage. For example, the reason for the sharp traffic increase and subsequent link congestion on the TeleGlobe links (Figure 3-(a)) was because in an effort to reduce the traffic "flooding" from external users and to make bandwidth available for internal users, NORDUnet had shut off one of the three OC-3 links to the TeleGlobe backbone.

Since we do not have the access to many other ISP's traffic statistics, we cannot make the claim that the traffic condition in NORDUnet and SWITCH networks are typical for the rest of the Internet. However, from the statistics that have been released by several other ISP's, such as Above.net [26] and BBC [27], we discovered that they all have the similar traffic behavior. *The average link utilization is always reasonably low, and many links are lightly loaded at all time. However, every network always has busy links (particularly, access links at NAP's and peering points) that have long lasting high bandwidth utilization.*

[3] We acknowledge the kind and detailed response from Havard Eidnes at NORDUnet.

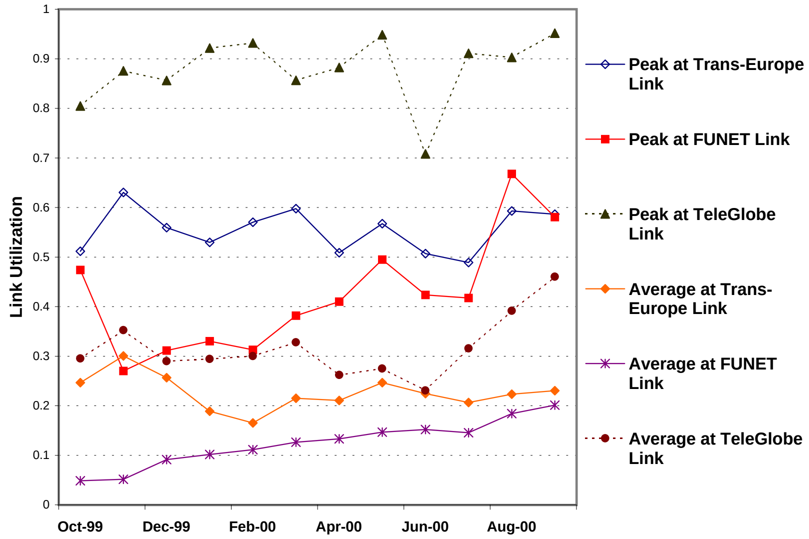


(a) Input traffic link utilization at the NAP's.

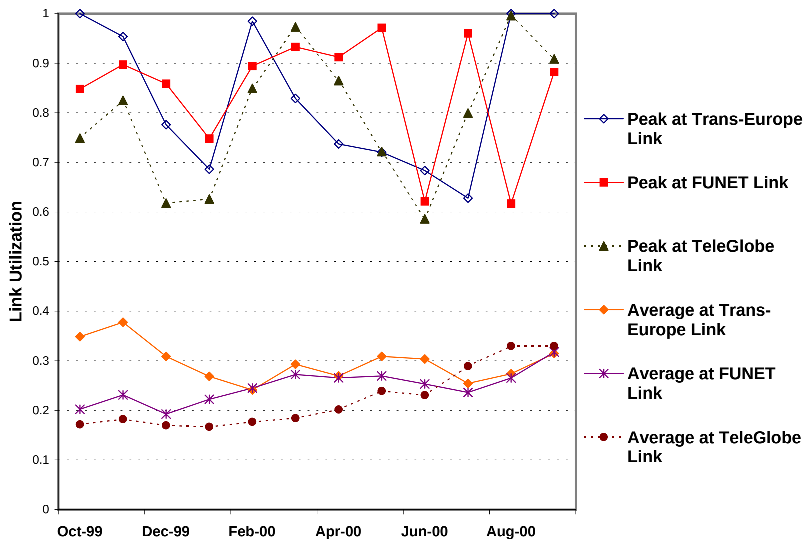


(b) Output traffic link utilization at the NAP's.

Figure 2: NORDUnet traffic average and peak link utilization at the NAP's from October 1999 to September 2000. Note, the DGIX NAP runs on a 100 Mbps Ethernet, that has a maximum link utilization of 70-80%.



(a) Input traffic link utilization with peers.



(b) Output traffic link utilization with peers.

Figure 3: NORDUnet traffic average and peak link utilization with three different peers from October 1999 to September 2000. NORDUnet has multiple peering links to FUNET and TeleGlobe. The data here is the average from those links.

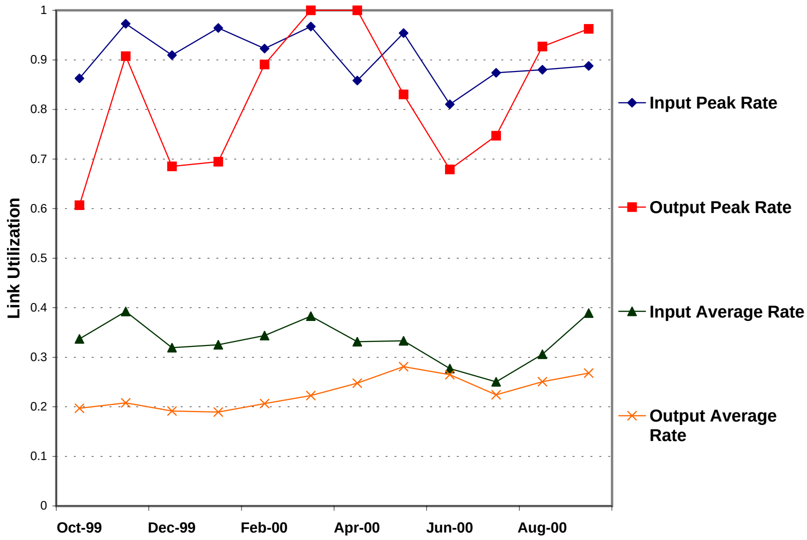


Figure 4: NORDUnet traffic average and peak link utilization on the transatlantic links from October 1999 to September 2000. NORDUnet has three OC-3 transatlantic links. The data here is the average from those links.

3.2 Getting Connected Is Costly

One obvious solution to reduce or eliminate the congestion and jitter problems is to add more bandwidth at congested links, such as the access links. Unfortunately, this is unlikely to happen any time soon in many networks, especially those operated by the small ISP's. This is because most of these links are leased from telephone companies, and are *very* expense.

Despite many people have predicted that bandwidth would become "dirt" cheap, the reality is that leased line prices have not decreased *consistently* and *rapidly*.

Andrew Odlyzko in [28] had estimated the tariffed price for a T1 link from New York City to Pittsburgh, a distance of about 300 miles, which is about the average of long distance private line links. In 1987, the link it was priced at $10,000 per month. After a consistent decrease for 5 years, the price was reduced to $4,000 in 1992. But from 1992 to 1998, the link price has climbed by over 50%, to $6,000 per month.

What is surprising about this finding is that the price increase took place during the same period for Internet commercialization, which has been one of the largest network infrastructure expansions in history: thousands of ISP's were founded, and the demands of leased lines were skyrocketing. For much of the 1990's, telecommunication equipment vendors had been continuously improving their products, and were able to manufacture routers and switches that support T1, T3 and OC-3 at very reasonable prices.

It is important to know that the bandwidth price is almost independent from the actual costs of cable installation. Generally, cable installation expense includes engineering and construction expense (such as digging), negotiating right-of-way (ROW) costs and fiber costs. It was estimated that for phone or cable companies to install cables, the average cost for installation is between $7 and $15 per meter [29]. Typically fiber installations are done in 36, 96, or 192 fiber strand bundles. The fiber cost is approximately 5 to 15 cents per strand per meter depending the strand size. However, network providers would have to pay almost 10 times as much the price [30] for fiber and maintenance.

In spite the bandwidth price increase in the 1990's, comparing with Europe, North America has much lower leased line prices [31]. At the end of 1998, for comparable distances (300 km) and bandwidth (2 Mb/s), US leased lines are four times less expensive than ones within the same Europe country, and 16 times less expensive than international links within Europe.

Not only are leased lines expensive, interconnecting ISP's can be costly as well. Table- 3 shows the connection fees at some of the NAP's.

NAP	E-3 (34 Mbps)	DS-3 (45 Mb/s)	OC-3 (155 Mb/s)	OC-12 (600 Mb/s)
Chicago (Ameritech)	-	3,900	4,700	-
Bay Area (PacBell)	-	3,750	5,000	13,500
France (Parix)	4,086	-	4,458	-

Table 3: Monthly Cost (in US Dollars) to connect to a NAP in a 3-year term. The NAP's listed here use ATM PVC to guarantee bandwidth for peer-to-peer traffic. The NAP's that operate with Ethernet and provide shared bandwidth among peers have a much lower price.

A network provider needs to pay neighboring providers for transferring its traffic. This transit traffic expense can be quite high. Recently, some North American providers [32] have claimed that the cost for a transit DS-3 link was $50,000 per month, and a OC-3 link costs up to $150,000 per month. To reduce costs, providers have begun to adapt the method of establishing *private peering* among each other, and to reduce the amount of traffic going through public NAP's. Typically, in a private peering arrangement, two connected providers share the cost of a single link, assuming both provider's traffic volume on the link are symmetrical. For example, Telia, a backbone provider in Sweden, had analyzed their transit costs and recognized that approximately 85% of their traffic at MAE-East, a large public NAP, was to their transit provider and the remaining 15% was through peering relationships. By focusing on establishing peering relationships with the top 25 destination AS's, they shifted the mix to 70% through private peering, with the remaining 30% of traffic going through the NAP. The result was increased traffic efficiency and a reduction in the cost of transit.

In summary, we believe that, like for many other commodity products, such as computer software and microprocessors, the rapid technological progress and the rising consumer demand *do not* necessarily translate to rapid price decrease in the market. It appears that due the monopoly by a small number of major players (phone companies, cable companies, and large Internet providers) [33], and strict and protective government regulations [34], the price for bandwidth will not decrease fast enough, despite the inexpensive fiber installation and maintenance costs.

3.3 The Vast and Flat Internet

So far, we have shown the existence of bandwidth bottleneck links in the network, many of which are located at network access places. But how much end-user traffic is actually going over these links?

The answer depends on the size and the geographic location of the provider networks. For example, UUnet has a very dense network installation in the United States. When two UUnet users communicate with each other in US, it is most likely that their traffic stays within the same network, and won't go over bandwidth bottleneck links. However, for European, Asian and Australian users, a large percentage of their Internet traffic is actually with US networks. According to an estimation several years ago [35], 60% of Telstra (the dominant Australian provider) Internet traffic was with US provider networks.

Thus, it's reasonable to assume that a large portion of end-to-end traffic traverse over multiple routers and provider networks. The average end-to-end router hop-count was 16 in a 1996 measurement [36]. As measured by Paxson [37], the longest end-to-end hop-count in the Internet was greater than 30 in 1997.

When user packets traverse multiple providers, they are likely to experience congestion at access links. A typical example is the following: suppose that users from regional network $\mathcal{A}$ need to communicate with users in another regional network $\mathcal{D}$. Backbone provider $\mathcal{B}$ provides transit service for $\mathcal{A}$; backbone $\mathcal{C}$ for $\mathcal{D}$. Backbone $\mathcal{B}$ and $\mathcal{C}$ exchange traffic at a NAP or through private peering. Thus, end-to-end traffic involves all four provider networks: $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$, and has a provider hop-count of 4. Of course, it is also possible that there are more than two transit networks ($\mathcal{B}$ and $\mathcal{C}$, as illustrated here) to interconnect sparsely located user populations.

Figure 5 shows the AS path information gathered at BGP border routers at University of Oregon [38] and BELNET [39] [4]. The x-axis is the provider hop-count (represented in AS length), and the y-axis is the percentage of the destination networks. The figure shows the number of providers that a user at the measuring network has to travel in order to reach to a destination network. Most of the destination networks are 3 to 5 providers away. This result has been further confirmed by a recent Bell Labs study [40] and data gathered by Telstra [41].

However, we need to take two things into consideration here. First, the AS information is announced by the destination networks, and distributed by BGP. Border routers can receive the same AS reachability data from multiple neighboring networks. It's up to the border routers to determine the most suitable routing path. Hence, the collected AS path information may not reflect the actual data forwarding path.

The BGP routing information is aggregated at border routers. The reasons for BGP aggregation are to reduce the total number of states maintained by the routers, and to localized the route flapping effects to a small portion of the network. It is possible that the measured AS path length has been compressed, and the actual AS length (or provider hop-count) is longer.

Obviously, the more hops and providers that user packets travel over, the more likely that they will go over bandwidth bottleneck links. Given the on-going rapid Internet deployment in Far East and Latin America, and continuing growth in US and Europe, we predict both router and provider hop-counts between large user populations will increase.

[4] The data was collected by Olivier Bonaventure of BELNET in December 1999. We acknowledge his timely response to our inquiry.

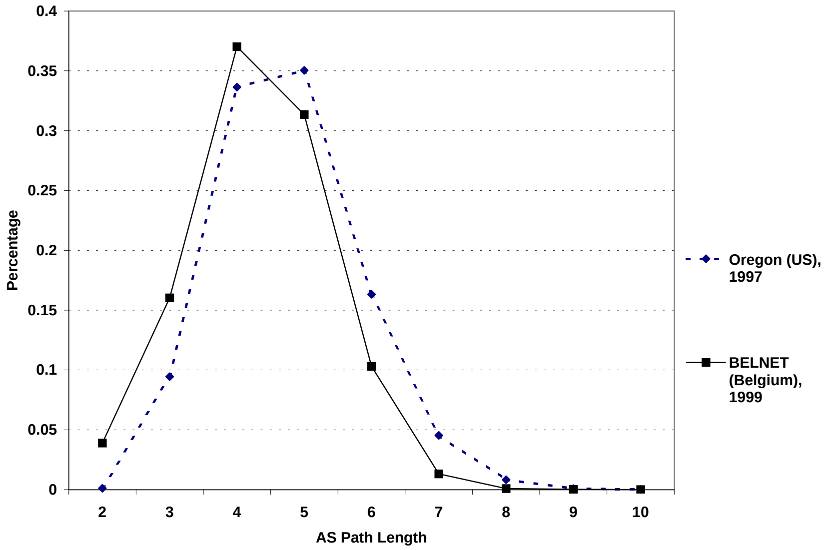


Figure 5: The AS path length information gathered from BGP AS-Path attribute.

4 Sustaining Internet Traffic Growth

The Internet is in a transitional period. According to an 1998 estimation by Coffman and Odlyzko [42], the Internet traffic has a 100% annual growth rate. We predict that, in the near future, some of the new applications may not only accelerate traffic growth, but change traffic characteristics as well. This will add more network deployment and management challenges to providers. Here, we list some of the emerging applications.

- **VPN traffic:** In studies conducted by Coffman and Odlyzko [42, 35], it was shown that in 1997, the traffic carried by private lines and the Internet were 3,000 - 5,000 TB/month, and 2,500 - 4,000 TB/month, respectively. In 1999, there were 5,000 - 8,500 TB/month for private lines and 10,000 - 16,000 TB/month for the Internet. In 1997, the revenue for retailed leased lines for private non-Internet carriers companies was 10 billion dollars in US. The numbers imply that private networking is a profitable and growing business.

  Currently, there are two trends in private networking: one is to interconnect private networks over the public Internet infrastructure, and the other is to replace existing private networks with the Internet. In both cases, it requires the Internet to be able to duplicate the same secure and predicable services as in the private networks. Specifically, the legacy services such as SNA require timely packet delivery and are extremely sensitive to packet loss. One way to achieve this goal is to apply resource reservation, or traffic engineering, in the Internet. In fact, creating customer-level VPN's in the form of MPLS [43, 44] has been one of the main driving forces behind the recent attempts on developing and deploying resource reservation protocols in the Internet.

- **Real-time data:** Though streaming data with real-time requirements represents a very small portion of the total Internet traffic today, this situation can be quickly altered if Internet is to carry telephony voice traffic. In 1999 [35], the voice traffic in US telephone systems was 40,000 TB/month, while the Internet traffic during the same period was 10,000 - 16,000 TB/month. That is, voice traffic volume is significantly larger than that of Internet traffic. Even a small portion of the voice communication going through Internet can dramatically change traffic pattern and traffic dynamics. Unfortunately, due to network delay and occasional link congestion, the current Internet is *not* suitable for end-to-end voice traffic delivery.

  Some have argued that there is no good economical reason for introducing voice service to the Internet, because the net cost of using the Internet and the existing phone networks is pretty much the same. However, this is not a proper measure to evaluate the emergence of IP telephony technology. The most significant advantage that IP telephony can offer is *not* as a telephone service replacement, rather it is meant for service integration, where voice, messaging and audio/video streaming applications can be accessed by end users from a single networking interface. It thus force the networks to add the capability to deliver user packets that have different criteria in terms of delay and throughput.

- **Disruptive innovations:** Network engineering and deployment take time. The past history tells us that user traffic triggered by some "killer" applications could potentially "flood" the network before the providers can respond and adjust their network capacity. For instance, due to Web browsing, the Internet traffic has an abnormal growth [35] that doubles every three or four months in 1995 and 1996, which represents a 100-fold traffic explosion over a two-year span. Another traffic explosion took place more recently. In 1999, several university networks had observed abnormal traffic growth [45] due to Napster [46]. Fear sever network congestion and communication degradation on other Internet applications, many universities were forced to ban or limit Napster usage.

5 The *Scaling* Issues in Resource Reservation

From the analysis in Section 2, it was proven that reservation-capable networks deliver better performance than best-effort-only networks for both rigid and adaptive traffic depending on the traffic assumption. However, reservation-capable networks require the routers to have admission control capability, which is believed to impose various scalability problems. A typical example for lack of scalability has been the RSVP-IntServ model.

In this section, we elaborate and evaluate the scalability problems in resource reservation.

5.1 Per-flow Queueing

In a network where there are many flows, the processing overhead associated with real-time scheduling and queuing becomes non-negligible [47]. The overhead is caused by maintaining queues for each "micro" (per-user) flow and assigning packets to each such queue.

To reduce per-flow queuing overhead, several alternative architecture have been proposed, including the IETF DiffServ model [48] and the *dynamic packet state* [49] architecture. In the DiffServ model, backbone routers simply implement a set of buffer management and priority-like queuing disciplines for each of a very small number of traffic "classes", providing them with coarse grained rate guarantees. At network edge, user flows are aggregated into these classes. It was shown in [50] that the edge routers were indeed capable of providing reasonably accurate rate guarantees and fair distribution of excess resources, with minimal impact on their raw performance.

Another important factor is that, in recent years, route vendors have changed the traditional single-processor router design, and adapted more advanced and mature technologies that use ASIC chips for packet classification, forwarding and queuing. As a result, today's routers can support thousands of packet filters for QoS usage [51].

In summary, it appears that we can overcome the per-flow (or micro-flow) limitation by adapting different QoS approaches at core and edge [52], and replace some of the old equipment with new and more advanced routers.

5.2 Router Memory Usage

When RSVP was first proposed in 1993, routers were not equipped with much memory due to the high cost of SRAM/DRAM at the time. Router's software was carefully designed and developed to make a good use of the limited memory to support rapid growing IP routes and other link-layer data (MAC addresses and Frame Relay DLCI's). For example, the last NSFnet backbone routers from IBM in 1993 managed to operate with 16 MB of fast memory. During the implementation of RSVP, the developers quickly realized that storing one reservation could take as many as 500 bytes [53, 54]. That means, to support 10,000 RSVP sessions, routers need to allocate 5 MB of memory. At the time, there were concerns among developers over RSVP's large memory requirement.

However, in the subsequent years, the cost of memory has gone down dramatically. Today's routers are equipped with more memory: Cisco 12416 routers [51] can support up to 256 MB memory, and Juniper M-160 routers [51] are equipped with 768 MB of DRAM by default. Nowadays, assuming good software engineering practice, memory usage is no longer a bottleneck for reservation protocols.

5.3 Bandwidth Usage

Frequent control message exchanges between routers may result in consuming significant amount of link bandwidth. This is particularly significent in soft-state based protocols such as RSVP.

Generally, on a network link, control message bandwidth B_{ctrl} is proportional to the total number of reservations, N, the total message size, S, that is required to setup and maintain a reservation, and the soft-state refresh period R:

$$B_{ctrl} = \frac{N \cdot S}{R}$$

We associate each reservation with an user flow rate r. Hence, the ratio between control message and reserved data traffic is determined by

$$\eta = \frac{S}{R \cdot r}.$$

For dial-up modem and cellular users, as a common sense, the number of reservation sessions has to be small - most likely one session at any given time. A typical RSVP PATH message is 200 bytes; a typical RESV (Controlled Load and Fixed Filter) message is 150 bytes. Thus, S = 350 bytes. On a 64 kb/s modem link, a reservation has $B_{ctrl} \sim$ 93 b/s, and $\eta \sim$ 0.146%.

On a backbone link that supports N = 10,000 RSVP sessions, we choose R to be 30 seconds, which is the default value recommended by the RSVP specification [6]. We have $B_{ctrl} \sim$ 0.93 Mbps. Assume r to be 64 kb/s, the typical voice transmission rate, $\eta \sim$ 0.146%, which is the same as for a single user.

When RSVP is used to make QoS (*e.g.*bandwidth) reservations, the maximum number of sessions on any given link is bound by the link capacity. Therefore, the upper bound for control message bandwidth is predictable and manageable. However, when reservation protocols, such as RSVP, are used to acquire and maintain MPLS "labels", in this case, reservation sessions are not necessarily associated with specific link bandwidth. B_{ctrl} is no longer bounded, and can become quite large.

Recently, several mechanisms have been proposed to reduce B_{ctrl}. In [55], we have proposed a scheme called *staged refresh timers* to reduce control message bandwidth by increasing refresh interval R, while improving the reliability in RSVP message delivery. Another mechanism, *summary refresh*, has been introduced in [56] to compress reservation message size S.

Another way to reduce B_{ctrl} is to limit the total number of reservations, N, in the network through *reservation aggregation*. Guerin *et al.* [57], Berson *et al.* [58], and Baker *et al.* [59] have proposed various mechanism to support reservation aggregation within the RSVP-IntServ framework.

5.4 Message Processing Cost

Message processing cost is directly related to protocol complexity. Here, we must separate the idea of having resource reservation from the complexity and the scalability problems that are involved in reservation protocol designs and implementations.

As being advocated in the previous sections, networks and users can benefit from using resource reservations. However, when we look into RSVP, there is little doubt that implementing this receiver-initiated, multicast-driven, two-pass reservation protocol is very difficult. In particular, supporting wildcard filter and blockade state efficiently can be very challenging. Nevertheless, the IBM RSVP router implementation [60] was able to setup a new RSVP session in 1.1 msec, and process a refresh message in 0.64 msec for a low-cost low-power 32 MHz Motorola processor.

By simplifying the reservation sequence on routers, we can drastically reduce processing overhead. For instance, we proposed a lightweight sender-initiated, one-pass reservation protocol, YESSIR. On the same IBM router platform, the YESSIR implementation improved the reservation setup time by 70% over RSVP. It was also shown that with careful implementation and by using some of basic hashing techniques to manage reservation states, a stand-alone YESSIR implementation can process up to 10,000 flow setups per second (or support up to 300,000 flows per box) on a commodity 700 MHz Pentium PC [61].

Hence, we have shown that routers have the capacity to support a large number of reservations. However, we need to understand protocol design techniques and simplify the reservation sequence at routers.

5.5 State Management Cost

Each individual reservation needs to be managed separately by the providers. Management includes tasks such as performing various actions according to service-level agreement (SLA), and monitoring and collecting user traffic for accounting and billing. The complexity and the overhead involved in reservation management increase linearly with the number of reservations. Therefore, there can be a scaling problem if there are too many reservations in the network.

To grasp the scope of this issue, we need to understand the maximum number of "states" that providers are capable or willing to handle. A close analogy is the routing policies managed by each ISP. Presently, network providers regulate incoming and outgoing transit traffic by setting up route policies at border routers. A typical route policy consists of a list of destination networks whose traffic is allowed or disallowed to receive [62]. Each policy is derived from the bilateral agreement among neighboring providers. If an ISP does not manage its routing policies properly, it can cause traffic "flooding" and looping in the Internet. The consequence is catastrophic and can result to neighboring providers blocking all traffic coming from and going to the offending ISP. The maximum number of route policy entries at a large provider's backbone is in the order of 90,000, which is the total number of routes in today's Internet, though in reality the actual number of routing policy entries is much smaller due to route aggregation. We believe that handling resource reservations is comparable to managing routing polices. For one thing, improper reservation management can cause service disruption and wasting network resources.

We now investigate the maximums number of reservations that a provider can possibly have. Figure 6 shows the total number of end users, networks and routing domains in the entire Internet over the years. Given that regional providers normally have a small user population (in thousands), it's feasible to have per-user reservation within regional networks. On the other hand, backbone networks are generally peering with each other, and carry user's traffic from other networks. Figure 6 indicates that there are around 100 million end users in the Internet today. Since it is not unreasonable to assume that a substantial percentage of the total user population goes through a large transit backbone, we hence believe that per-end-user reservation will not work in the backbone environment.

There have been several proposals [65, 66] for supporting resource reservation in the backbones.

The PASTE approach [66] is to use RSVP at the control layer and MPLS at the data forwarding layer to create a scalable traffic management architecture that supports service differentiation. However, using RSVP can run into the N-square problem, where N is the number of reservation participants in the network.

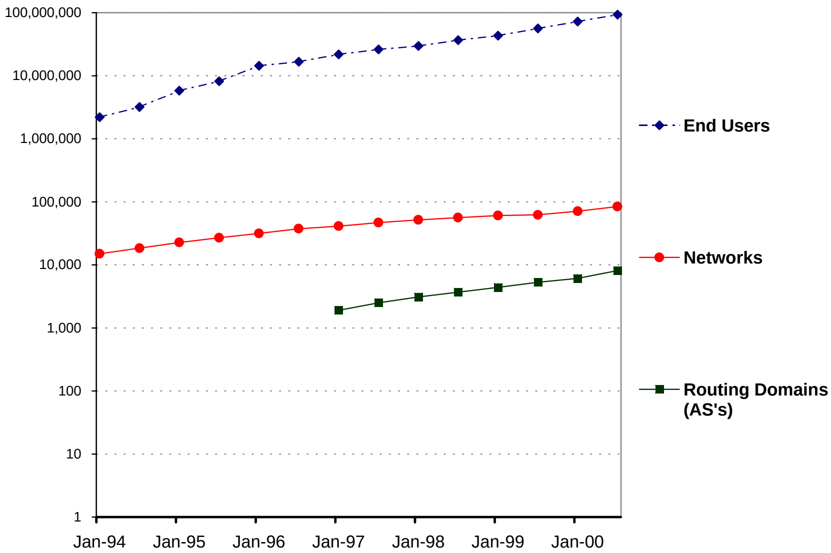


Figure 6: The growth of the Internet from 1994 to 2000 [63, 41, 64]. The number of networks was collected by Geoff Huston of Telstra at AS-1221 by monitoring BGP routes. The AS data was collected by Tony Bates from GX Network. Due to CIDR aggregation, both numbers can be slightly different at different location inside the network.

This is due to the design of the RSVP protocol itself, where each router has to maintain both traffic source and destination for each reservation flow. In a network with $\mathcal{N}$ users, the total number of states at each router thus becomes $O(\mathcal{N}^2)$. Figure 6 gives the possible reservation participant at different granularity. It clearly shows that due to the potential N-square problem, RSVP should not be used to setup reservations at per-user, per-network or per-AS levels.

The BGRP proposal [65] is to setup reservations at AS-level. To counter the N-square problem, BGRP is to create reservation sink-trees in the network, where each reservation destination domain is considered as a "sink". Reservations going to the same destination domain are aggregated at transit domains. Consequently, the total number of reservations that routers has to maintain is $O(\mathcal{N})$, where $\mathcal{N}$ is the total number of AS's in the network. The current number of AS's in the Inertent is is over 10,000.

In conclusion, we believe that there is a scaling problem in dealing with inter-provider reservations. The solution to this problem is two-folded. First, we need to introduce a hierarchical reservation model, where the regional networks setup reservations at user-level, and the backbone networks need to maintain reservations at network-level or AS-level. Second, we need to design a new reservation protocol that has good scaling property for the backbone networks.

6 Discussion

In recent years, networking infrastructures have attracted great amount of investment. Many backbone networks were built, and a bandwidth commodity market was developed. This created a pressure on margins of service providers. As a result, revenues of service providers are not growing fast enough in relationship to the growth of investment in ever-more-expensive infrastructures. This has been one of the reasons for recent telecommunication stock market downfall.

To generate revenues, the provders need to create services differentiated in quality. This is more than an issue that some applications need a better QoS than others. Fundamentally, it is needed due to economic reasons. Differentiated services mean different packets are treated differently in the network, and that in turn means adding "costs" to the network. So instead of trying to avoid "costs", the real issue becomes how to minimize them.

From our investigation, the current Internet has bandwidth bottleneck links that can cause long-lasting congestion and jitter. Meanwhile, leased line cost has not been reduced sufficiently in a timely manner for many ISP's to deploy high-speed links everywhere in their networks. In reality, the applicability of over-provisioning is questionable.

It is possible that over-provisioning is sufficient to reduce packet transaction delay and link congestion *within* some of the large well-funded backbones [17, 18], however, it is certainly not a cost-effective solution for all ISP's. In fact, to reduce the cost and the volume of transmitting data, many providers have been adapting techniques such as web caching and private peering. Keep in mind, if bandwidth is truly a cheap commodity, why would the ISP's bother with adjusting internal configuration to support private peering, and web cache servers?

Nobody can predict exactly what will happen in the future. In today's fast growing inter-networking environment, network deployment is not about meeting today's demand, instead, it is to anticipate and be ready for the new services and new types of traffic. It is a near-sighted argument to rule out the possibility of wide-deployment of inter-active real-time applications (such as voice over IP) in the future. It is almost impossible to sufficiently over-provision network resources for the future. To continue the expansion of the Internet, we have to seek for a better solution that can accommodate network growth at a reasonable cost. If we continue to allow some traffic sources to get free resource that they want, and provide no traffic isolation and protection to mission-critical data flows, we are bounded to face network congestion and unexpectable end-to-end delays.

Using resource reservation can bring many benefits to the network. It can provide traffic protection and service guarantees to user flows inside the network, so that the remotly located user populations can have acceptable and consistent communication over the Internet. Also, it can be used as a tool to prevent unregulated traffic from flooding the network to degrade other user's applications.

We need to separate the concept of resource reservation from the complexity and scalability issues in resource reservation protocols. Issues for the former include when and how to apply reservations, whereas message processing, state maintenance and state aggregation are the engineering issues that fall into the latter category.

We need to be careful with where to apply resource reservation. Just because reservations can protect user flows from temporary link congestion, it is *not* a tool to compensate for resource inadequacies caused by over-subscription, and sloppy network design. In fact, applying resource reservation in an over-subscribed network can cause service degradation to many users, and aggravating congestion conditions.

Resource reservation requires states to be installed and maintained at the routers. To support large number of reservation flows efficiently, reservation protocols must be scalable and relatively simple to process. From our studies on the reservation scalability problem, we argue that the scalability problem is not only an issue about the number of the flows that routers can process, but also an issue about the number of the flows that providers can manage for policing, accounting and billing. We believe that much future research is needed on understanding network resource manageability.

7 Acknowledgment

We appreciate and acknowledge many people for providing helpful information and insightful comments in the past several months. They are Bill St. Arnaud, Fred Baker, Olivier Bonaventure, Randy Bush, Havard Eidnes, Roch Guerin, Ellen Hahne, Andrew Odlyzko, Vern Paxson, Jennifer Rexford, Luca Salgarelli and Yechiam Yemini.

References

[1] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala, "RSVP: a new resource ReSerVation protocol," *IEEE Network*, vol. 7, pp. 8–18, Sept. 1993.

[2] L. Zhang, S. Shenker, D. Clark, C. Huitema, S. Deering, and D. Ferrari, "Reservations or no reservations," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (Boston, Massachusetts), Apr. 1995. panel-discussion slides.

[3] F. Baker, J. Crowcroft, R. Guerin, H. Schulzrinne, and L. Zhang, "Reservations about reservations," in *Proc. of IFIP Fifth International Workshop on Quality of Service (IWQOS '97)*, (New York, NY), May 1997.

[4] H. Schulzrinne, "Guide to nevot 3.34." http://www.cs.columbia.edu/ hgs/nevot/.

[5] UCB/LBNL, "vat – lbnl audio conferencing tool." http://www-nrg.ee.lbl.gov/vat/.

[6] R. Braden, Ed., L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource ReSerVation protocol (RSVP) – version 1 functional specification," Request for Comments 2205, Internet Engineering Task Force, Sept. 1997.

[7] R. Braden, D. Clark, and S. Shenker, "Integrated services in the internet architecture: an overview," Request for Comments 1633, Internet Engineering Task Force, June 1994.

[8] J. Wroclawski, "The use of RSVP with IETF integrated services," Request for Comments 2210, Internet Engineering Task Force, Sept. 1997.

[9] J. Wroclawski, "Specification of the controlled-load network element service," Request for Comments 2211, Internet Engineering Task Force, Sept. 1997.

[10] S. Shenker, C. Partridge, and R. Guerin, “Specification of guaranteed quality of service,” Request for Comments 2212, Internet Engineering Task Force, Sept. 1997.

[11] S. Shenker, “Fundamental design issues for the future internet,” *IEEE Journal on Selected Areas in Communications*, vol. 13, Sept. 1995.

[12] L. Breslau and S. Shenker, “Best-effort versus reservations: A simple comparative analysis,” *ACM Computer Communication Review*, vol. 28, pp. 3–16, Sept. 1998.

[13] W. E. Leland, M. S. Taqq, W. Willinger, and D. V. Wilson, “On the self-similar nature of Ethernet traffic,” in *SIGCOMM Symposium on Communications Architectures and Protocols* (D. P. Sidhu, ed.), (San Francisco, California), pp. 183–193, ACM, Sept. 1993. also in *Computer Communication Review* 23 (4), Oct. 1992.

[14] V. Paxson and S. Floyd, “Wide area traffic: The failure of poisson modeling,” *IEEE/ACM Transactions on Networking*, vol. 3, June 1995.

[15] M. E. Crovella and A. Bestavros, “Self-similarity in world wide web traffic: Evidence and possible causes,” *IEEE/ACM Transactions on Networking*, vol. 5, pp. 835–846, Dec. 1997.

[16] Z. Sahinoglu and S. Tekinay, “Self-similar traffic and network performance,” *IEEE Communications Magazine*, vol. 37, pp. 48–52, Jan. 1999.

[17] UUnet, “Latency statistics.” http://www.uu.net/network/latency/index.html.

[18] Sprint, “Sprint network latency statistics.” http://www.sprintbiz.com/internet_solutions/SLAs.html.

[19] V. Paxson, “End-to-end internet packet dynamics,” in *SIGCOMM Symposium on Communications Architectures and Protocols*, (Cannes, France), Sept. 1997.

[20] V. Paxson, *Measurements and Analysis of End-to-End Internet Dynamics*. PhD thesis, University of California at Berkeley, Berkeley, California, May 1997.

[21] International Telecommunication Union (ITU), “Transmission systems and media, general recommendation on the transmission quality for an entire international telephone connection; one-way transmission time,” Recommendation G.114, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Mar. 1993.

[22] International Telecommunication Union, “Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s,” Recommendation G.723.1, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Mar. 1996.

[23] A. Odlyzko, “The Internet and other networks: Utilization rates and their implications,” technical report, AT&T Labs, Florham Park, NJ, Feb. 2000.

[24] SWITCH, “Switchlan traffic statistics.” http://www.switch.ch/lan/stat/.

[25] NORDUnet, “Nordunet network statistics.” http://www.nordu.net/stats/.

[26] ABOVE.net, “Above.net’s real-time network status.” http://stats.sjc.above.net/traffic/.

[27] BBC Internet Services, “Internet link usage.” http://support.bbc.co.uk/support/mrtg/internet/.

[28] A. Odlyzko, “The history of communications and its implications for the Internet,” technical report, AT&T Labs, Florham Park, NJ, June 2000.

[29] B. S. Arnaud, "CANARIE – CA*net 3 – the customer empowered networking revolution," in *Proc. of First Australian Advanced and Innovative Internet Workshop*, (Adelaide, Australia), Mar. 2000.

[30] C. Stefano, "Rural connectivity: Economics," in *Second Advisory Committee Meeting for National Coordination Office for Information Technology Research and Development*, (Arlington, VA), June 1997.

[31] Commission of European Communities, "Commission recommendation on leased lines interconnection pricing in a liberalised telecommunications market," tech. rep., Commission of European Communities, Nov. 1999.

[32] W. B. Norton, "Internet service providers and peering," in *Proc. of NANOG 19*, (Albuquerque, New Mexico), June 2000.

[33] R. Frieden, "Does a hierarchical internet necessitate multilateral intervention?," in *Proc. of 28th Research Conference on Communication, Information and Internet Policy*, (Alexandria, Virginia), Sept. 2000.

[34] B. Zhang, "Assessing the WTO agreements on China's telecommunications regulatory reform and industrial liberalization," in *Proc. of 28th Research Conference on Communication, Information and Internet Policy*, (Alexandria, Virginia), Sept. 2000.

[35] K. G. Coffman and A. M. Odlyzko, "Internet growth: Is there a "moore's law" for data traffic?," tech. rep., AT&T Labs - Research, Florham Park, New Jersey, July 2000.

[36] NLANR, "Assessing average hop count of a wide area internet packet." http://www.nlanr.net/NA/Learn/wingspan.html.

[37] V. Paxson, "End-to-end routing behavior in the internet," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 601–615, Oct. 1997.

[38] NLANR, "Nlanr as path lengths." http://moat.nlanr.net/ASPL/.

[39] BELNET, "Belnet: The belgian research network." http://www.belnet.be/main_uk.html.

[40] B. Cheswick, J. Nonnenmacher, S. Sahinalp, R. Sinha, and K. Varadhan, "Properties of the internet topology," submitted for publication, Bell Labs, July 2000.

[41] Telstra Network, "Telstra internet network performance reports." http://www.telstra.net/ops/.

[42] K. G. Coffman and A. M. Odlyzko, "The size and growth rate of the internet," tech. rep., AT&T Labs - Research, Florham Park, New Jersey, July 1998.

[43] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol label switching architecture," Internet Draft, Internet Engineering Task Force, Aug. 2000. Work in progress.

[44] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP tunnels," Internet Draft, Internet Engineering Task Force, Feb. 2001. Work in progress.

[45] DoIT Network Engineering, "Internet traffic flow size analysis." http://net.doit.wisc.edu/data/flow/size/.

[46] "Napster." http://www.napster.com.

[47] T. cker Chiueh, A. Neogi, and P. Stirpe, “Performance analysis of an RSVP-capable router,” in *Proc. of 4th Real-Time Technology and Applications Symposium*, (Denver, Colorado), June 1998.

[48] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An architecture for differentiated service,” Request for Comments 2475, Internet Engineering Task Force, Dec. 1998.

[49] I. Stoica and H. Zhang, “Providing guaranteed service without per flow management,” *ACM Computer Communication Review*, vol. 29, pp. 81–94, Oct. 1999.

[50] R. Guerin, L. Li, S. Nadas, P. Pan, and V. Peris, “The cost of QoS support in edge devices: An experimental study,” in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (New York), Mar. 1999.

[51] LightReading, “The internet core router test,” in *LightReading Test*, Mar. 2001. http://www.lightreading.com/testing/.

[52] G. Huston, “Next steps for the IP QoS architecture,” Request for Comments 2990, Internet Engineering Task Force, Nov. 2000.

[53] B. Braden, “Isi RSVP host implementation,” software release, ISI, Marina del Rey, California, June 1997.

[54] P. Pan and R. Guerin, “Ibm research RSVP router implementation,” software documentation, IBM T. J. Watson Laboratory, June 1997.

[55] P. Pan and H. Schulzrinne, “Staged refresh timers for RSVP,” in *Proceedings of Global Internet*, (Phoenix, Arizona), Nov. 1997. also IBM Research Technical Report TC20966.

[56] L. Berger, D. Gan, G. Swallow, P. Pan, F. Tommasi, and S. Molendini, “RSVP refresh overhead reduction extensions,” Internet Draft, Internet Engineering Task Force, June 2000. Work in progress.

[57] R. Guerin, S. Herzog, and S. Blake, “Aggregating RSVP-based QoS requests,” Internet Draft, Internet Engineering Task Force, Nov. 1997. Work in progress.

[58] S. Berson and S. Vincent, “Aggregation of internet integrated services state,” in *submitted to IWQOS*, 1998.

[59] F. Baker, “Aggregation of RSVP for IP4 and IP6 reservations,” Internet Draft, Internet Engineering Task Force, June 1999. Work in progress.

[60] P. P. Pan and H. Schulzrinne, “YESSIR: A simple reservation mechanism for the Internet,” in *Proc. International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, (Cambridge, England), pp. 141–151, July 1998. also IBM Research Technical Report TC20967.

[61] P. Pan and H. Schulzrinne, “Lightweight resource reservation signaling: Design, performance and implementation,” Technical Memorandum 10009669-03TM, Bell Labs, Lucent Technologies, Murray Hill, NJ, June 2000.

[62] NANOG, “Isp filter policies.” http://www.nanog.org/filter.html.

[63] Internet Software Consortium, “Internet domain survey.” http://www.isc.org/ds/.

[64] T. Bates, “The cidr report.” http://www.employees.org/ tbates/cidr-report.html.

[65] P. Pan, E. Hahne, and H. Schulzrinne, “The border gateway reservation protocol (BGRP) for tree-based aggregation of inter-domain reservations,” *Journal of Communications and Networks*, June 2000.

[66] T. Li and Y. Rekhter, “A provider architecture for differentiated services and traffic engineering (PASTE),” Request for Comments 2430, Internet Engineering Task Force, Oct. 1998.